Cities Drifting Apart:
Heterogeneous Outcomes of Decentralizing Public Education

Zelda Brutti*

European University Institute

This version: January 2016

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. Colombian municipalities were assigned to administer their public education service autonomously solely on the basis of whether they exceeded the 100 thousand inhabitants threshold. Exploiting this discontinuity, I estimate the impact that autonomy has had on student test scores across municipalities, using a regression discontinuity design and fixed-effects regression on a discontinuity sample. I find a test score gap arising between autonomous municipalities in the top quartile and those in the bottom quartile of the development range, in a trend that reinforces over time. From analysis of detailed municipal balance sheet data, I show that the autonomous high-developed municipalities invest in education more than the ad hoc transfers they receive, supplementing these with own financial resources. Indicators of municipal administration quality also show significant differences between the two groups of cities, helping to explain the education outcome patterns.

*E-mail: zelda.brutti@eui.eu. I benefited from suggestions by Steve Pischke, Jérôme Adda, Andrea Ichino, Caroline Hoxby, Alan Manning, Oriana Bandiera, Olmo Silva, Guy Michaels, Economics faculty and students at the London School of Economics, as well as from participants at several conferences. Special thanks to Fabio Sánchez Torres and to the University of Los Andes for important data supply and comments. Sincere gratitude for helpful comments is also extended to Alexander Moore, Matthias Hübener and to Jesús Arias Duarte at IADB, and for data to Germán Cano Torres at DNP. I gratefully acknowledge a scholarship from the Centro Studi Luca d'Agliano in Turin.

1 Introduction

Decentralization of public service provision has been at the top of policy agendas in numerous countries over the last decades, involving services such as education, health, public transport and the supply of energy, water and sewerage systems. In developing and middle-income countries, responsibilities are often handled from a central or regional level down to municipalities[1]. Such reform are expected to yield welfare benefits through better local preference matching, higher governor accountability and increases in the efficiency of service delivery[2]. Welfare losses may instead derive from inadequate management skills of local authorities, increases in administrative and coordination costs, corruption among local bureaucrats or local elites resource capture[3]. These positive and negative repercussions may materialize in different proportions across different regions in the reforming country, giving rise or exacerbating regional inequalities. In this paper I show that entrusting Colombian municipalities with managerial autonomy over local public education has yielded heterogeneous results on local educational outcomes, depending on the level of municipal development at the time of the responsibility takeover.

In this empirical analysis I benefit of an unusually clean decentralization criterion: autonomy over the education service was assigned to cities solely depending on whether they exceeded the 100 thousand inhabitants threshold. This decision rule relieves the analysis from the issues that typically hinder identification of the effects of higher autonomy: non-random selection into autonomy by local authorities, and other nationwide phenomena occurring along with decentralization. In this way, this study introduces innovation to existing literature on the topic. A second valuable aspect of studying the Colombian case is that it yields insight into a context in which decentralization was purely administrative, and not mingled, as it often happens, with changes on the fiscal or political front: managerial authority was transferred, but local taxation and local representation were left unchanged.

Using panel data on standardized student test scores over a period of 10 years after the reform, I show that higher autonomy has proven beneficial for highly developed municipalities, but not so for less developed municipalities. Average test scores in high-developed autonomous municipalities have started to significantly exceed those of their non-autonomous counterparts, in magnitudes that are growing over time. Low-developed autonomous municipalities instead appear to be progressively losing terrain with respect to their non-autonomous counterparts, even though effects are smaller

[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

[2] On informational advantage and heterogeneity in preferences, see seminal work by Musgrave [1959] and Oates [1972]. On accountability, monitoring and elections see Crook and Manor [1998], Manor [1999] and Blair [2000]

[3] Administrative costs are addressed in Breton and Scott [1978] and Panizza [2004]. Corruption and local elites capture are extensively discussed by Bardhan and Mookherjee [2000, 2002, 2005, 2006].

and not always statistically significant. Both test score gains and losses seem to be larger for students with a more advantaged socioeconomic background: they gain more in highly developed cities but lose more in low developed ones, increasing score dispersion in the former group of cities and decreasing it in the latter.

In the second part of the analysis, exploring municipal balance sheet data, I show that high and low developed municipalities implement different spending decisions. Autonomous municipalities of the upper development quartile invest on local education not only the ad-hoc transfers they receive from the government, but also add own financial resources to their per-pupil education budget. Municipalities in the lowest development quartile only spend the education resources they receive from the central government, or somewhat less. The two groups also show significant differences in terms of municipal management and law compliance indicators, in directions that are consistent with the results on student test scores.

The three ways in which this paper adds to existing work in the field are its quasi-experimental estimation setup, the ability to focus on solely administrative power shifts, and the provision of suggestive evidence on channels that drive the heterogeneity in outcomes across local development levels. Implications of the findings may represent relevant references for future public service decentralization reforms to be implemented in low and middle income contexts similar to the Colombian one, especially in presence of significant subnational heterogeneity in development levels and local wealth.

2 Selected Literature

Heterogeneity in the effects of decentralization is modeled by Bardhan and Mookherjee [2000, 2002, 2005, 2006], who show how the combination of strong local elites and weak local institutions implies decentralization to yield under-provision of services to the local poor. Channels for diversity of impacts across places and people are illustrated also in the reviews by Kaiser [2006] and, with a special focus on developing countries, by Juetting et al. [2005]. These reviews and the vast majority of empirical literature fail to establish any clear link between decentralization and poverty reduction, and document higher advantages for the rich with respect to the poor in decentralized contexts. Some studies describe correlations between indicators of local welfare and the spending decisions of local politicians, but do not establish 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] look at Colombian municipalities' balance sheet data and construct original aggregate measures of decentralization, then finding negative association between dependence on central government transfers

and expenditure on education, and positive association with public school enrollment rates. There are studies that aim at isolating causal effects of decentralization processes at different levels of local development, but focusing on fiscal decentralization or at contexts in which administrative decentralization came along with important fiscal and political changes. Hammond and Tosun [2011] apply a spatial error model on 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 non-metropolitan ones. The fixed effects analysis by Zhang [2006] shows that fiscal decentralization in China has promoted regional inequality, mainly due to inequalities in tax bases and thus in fiscal burden, and in the 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 benefited and adapted their expenditure structure to local needs. Closest to this paper in terms of reform context analyzed and in terms of outcomes looked at is Galiani et al. [2008], who 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 non-poor municipalities. This paper differs from the study on Argentina in scale of the reform (transferring to local authorities some additional schools versus the whole education service), in the level of government being looked at (regional versus municipal), and in the availability of a quasi-experimental setup for Colombia but not for Argentina[4].

3 Decentralization in Colombia and the 2001 Reform

Starting in the 1980s, Colombia has been undergoing a progressive decentralization process involving governance and administration, fiscal structure, and the delivery of public services; various authors have looked at the outcomes of these gradual processes, some in a qualitative and some in a quantitative fashion[5]. The reform in 2001 kept the political and fiscal scenarios unchanged and enforced administrative decentralization[6], reallocating local authorities' responsibilities towards the delivery of public services[7].

[4]"The transfer schedule was determined through bilateral negotiations between the federal government and each province" (Galiani et al., 2008, sec.3§3)

[5]Keeping the focus on education outcomes, Borjas and Acosta [2000], Vergara and Simpson [2001] and Caballero [2006] comprehensively illustrate dynamics and descriptive trends of decentralizing the public education system over the nineties, agreeing on generally undistinguished results.

[6]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].

[7]Educational outcomes of the 2001 reform are explored in the descriptive Colombian central bank report by Lonzano et al. [2007], who conclude that the post-reform years have witnessed progress in attendance rates but disappointing results in terms of quality and efficiency. Also Cortés [2010] focuses on the 2001 reform, uses enrollment data up to 2006 and compares municipalities who gained education autonomy to the remaining, finding that the

3.1 Pre-reform context and reform motivations

Colombia is currently structured into local authorities as follows: there are thirty-two departments[8], 1,118 municipalities located within departments and four special districts (see maps in Section 9 in the Appendix). Local authorities enjoy decisional and spending autonomy over a wide range of matters, although the necessary financial resources chiefly consist of central government transfers deriving from national tax revenues[9]. Central government transfers have historically been accounting for around 90% of the total education expenditure (nationwide average), and the remaining 10% is contributed by local authorities, with some local variability in these figures (Borjas and Acosta, 2000, p.6; Iregui B. et al., 2006, p.31;Santa Maria S. et al., 2009, pp.19-20). Up to the 2001 reform, the law had departments and municipalities jointly in charge of public education, entitled to hire personnel and invest in infrastructure and equipment[10]; as a result the division of responsibilities over the management of public education was vague and far from transparent (Borjas and Acosta [2000]). De facto, being the direct recipients of the bulk of education transfers[11], departments were the primary players on the education sector[12]. The elimination of any responsibility overlap 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, eliminating the yearly fluctuations in financial transfers, and updating some obsolete distribution criteria[13].

3.2 Reform content

Regarding the management of public education, the enactment of Law 715/2001 yielded the fundamental change of a clear-cut allocation of responsibility over the service to either municipalities or departments. Municipalities which counted 100 thousand or more inhabitants in the year 2002

former significantly increased enrollments of publicly subsidized pupils into private schools.

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

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

[10]Law 60 / 1993 (distributing competencies across levels of government and assigning resources accordingly), Law 115 / 1994 (the 'comprehensive education act'), and respective follow-up decrees.

[11]See Table 1

[12]For example, departmental payrolls included 85-90% of all public school teachers [Corte Constitucional, 1997, par.16], and departments decided on their allocation across municipalities. Municipalities were then responsible for allocating teachers across schools within their territory, and hired the remaining 10-15% that 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). Also see the DDTS [2004] report, p.6.

[13]For the official document motivating the reform, see: "Exposición de motivos 715 de 2001 Nivel Nacional", Congreso de Colombia, Gaceta del Congreso 294 de 2000. For further discussions of this matter see among others Sarmiento and Vargas, 1997; Alesina et al. [2000]; Borjas and Acosta [2000]; Vergara and Simpson [2001] and the technical report by DNP [2002].

became "certified in education" (certified municipalities, from now onwards), meaning responsible for the public education service on their territories. The education transfers from the central government, which are assigned on a per-pupil base, started to flow into their treasuries. Municipalities with fewer than 100 thousand inhabitants were not certified, and their public education is run by the departments they belong to. The next subsection further clarifies the concept of autonomy and discusses the shift in responsibilities. The forty municipalities certified in 2001 account for around one third of Colombia's population and pupil share; their size ranges from 105 thousand to over 2 million inhabitants[14][15].

The 2001 reform affected not only the education service, but also the provision of healthcare, water and sewerage and other smaller public services. Nonetheless, it was only for the education sector that this reform separated municipalities into autonomous and not, and used the 100 thousand inhabitants rule. Another task performed by the 2001 reform was updating the formulas used by the central government to compute financial transfers financing local public services; section 3.2.1 below provides further details and discussion on this aspect.

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

Table 1 summarizes competencies of local authorities before and after the 2001 reform, and indicates percentages of education transfers flowing to their treasuries.

As illustrated in the table, the reform left the role of the central government unchanged but polarized both financial transfers and managerial responsibilities of local authorities. From receiving a narrow share of transfers and being subject to departmental supervision, certified municipalities transitioned into a situation of full managerial and financial autonomy, while non-certified ones gave up their already limited powers to the respective departments[16]. For the rest of the analysis I will consider certified municipalities as "treated" by the decentralization reform and the non-certified counterparts as "untreated", since both figures and anecdotal evidence indicate that a truly substantial change in regime has happened for the former group but not for the latter. How a violation of this premise would affect the interpretation of empirical results is discussed in Section 5.3.

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

[14] See their locations in panel c), Section 9 Appendix.

[15] The reform provided for a transition period of two years (2002 and 2003), during which certified local authorities took over the school infrastructure, started the effective management of the service, and had the opportunity to reorganize staffing plans on their territories. During these two years temporary transfer amounts were set, and from 2004 onwards the new transfer system became fully operational.

[16] With only a 3% of total funds still flowing to non-certified municipalities, with pre-set destination. These funds need to be spent entirely on school infrastructure and school material, according to departments' directions [MEN, 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 onwards (Law 715/2001)**	
		Certified 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 extra education programs	Departments and municipalities, under departments' supervision	Teacher hiring, training and placement; School infrastructure and materials; School transport and any extra education programmes	Municipality only
		Non-Certified Municipalities	
		Transfers:	97% to department 3% to municipality
		Teacher hiring, training and placement; School infrastructure and materials; School transport and any extra education programmes	Department only (maintenance duties for municipality)

Author's illustration, based on Laws 60/1993, 115/1994 and 715/2001 (República de Colombia); Borjas and Acosta [2000]; DNP[2002]. Percentages are author's derivation: pre-reform is based on 2001 data in DNP[2002], p.16; post-reform is based on 2004 data in DNP[2004a] and DNP[2004b]. Percentages for departments include the four special districts.

number and seniority of teachers employed, with some adjustment based on number of inhabitants, local poverty and administrative efficiency. From 2002 onwards the allocation criteria were tilted towards a student headcount base, but with number of teachers still playing a key role, and again with some adjustment for local poverty and population density; these changes applied to transfers to all local authorities, certified and not[17]. Both before and after the reform, transfers are meant to be exclusively used for the service to which they are dedicated, administered in separate accounts and thus not fungible with respect to the remaining revenues and expenses of the local authority [MEN, 2003].

3.3 Further relevant aspects

3.3.1 Population and population cutoff

The population figures that were used for the 2001 reform were issued by the National Statistics Office (DANE). The counts were not prepared ad hoc for the reform but issued on the occasion of the 1993 general census, as forward projections. Certification was assigned to those municipalities that according to the projections for the year 2002 were exceeding 100 thousand inhabitants. The cutoff was sharply implemented, and no exceptions were made in either direction; the way population figures arose allows us to set aside any potential suspicion of targeted count manipulation.

Beyond its use in the 2001 reform, the 100 thousand inhabitant cutoff does not play any significant role in Colombia's legislation and it is never used in other matters involving municipal public service provision. The cutoff appears in the municipal classification that is performed every fiscal year by the central government. Current inhabitant count and current revenues are jointly used for the classification, and, given appropriate current revenues, 100 thousand inhabitants may represent the lower bound for a 'first' category city[18]. This categorization is updated every year and is used for setting limits to salaries of the mayor, of council members and administrative staff and 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.

Figure 7 in the Appendix shows smoothness of various municipal characteristics around 100 thousand inhabitants. Most notably, student test scores just before the reform (in 2001) do not exhibit discontinuities at the 2002 treatment cutoff. If we believe test scores to reflect a range of underlying municipal characteristics, especially those affecting education outcomes, the lack of discontinuities

[17]The transition to a transfer system giving more weight to student head-counts should have, if anything, favored municipalities characterized by low levels of local development„ as such areas have historically been disadvantaged in terms of teacher provision [Corte Constitucional, 1997, par.19]. Evidence on central government transfers having become more redistributive over time is presented in Table 15 in the Appendix.

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

in pre-reform scores injects further confidence on the absence of any relevant transitions occurring at 100 thousand inhabitants. Further falsification and robustness tests on these aspects are performed along the empirical analysis.

3.3.2 Districts and special municipalities

Districts are local authorities whose nature is mixed between departments and municipalities. Already before 2002, districts were drawing the totality of their financial entitlements for education directly into their treasuries and managing them autonomously. The four Colombian districts are Bogotá, Barranquilla, Cartagena and Santa Marta, and they are excluded from the analysis.

There are two municipalities[19] whose freedoms on local education policy had been formally enhanced in 1999-2000, even though the substantial implications of the procedure remained unclear. I exclude these two cities from the analysis.

4 Data

4.1 Test scores

Colombia has a long running tradition of standardized testing in public schools; ICFES is the government agency in charge of conducting and assessing the tests across the whole country. The most complete and frequent test score data refers to the Saber11 examination, which is administered to all students completing high school[20], and which is widely accepted as the reference examination to evaluate the quality of Colombian secondary education. Saber11 evaluates a range of school subjects; test scores range from 0 to 100 in each subject and are standardized by subject at the national level, to a mean of 50 and a standard deviation of 10. This is, each student's score is informative about his/her position relative to the national average in that subject. Individual-level Saber11 test scores are made available by ICFES for the years 2000 to 2012, with information about the school and municipality to which each student belongs, and some information student background.

4.2 Municipal Development Measures

The development level of Colombian municipalities is being evaluated periodically by government agencies: relevant data is collected by the National Statistics Office (DANE) and the summative indicators are calculated by the National Planning Department (DNP). Up to the year 2013, the

[19] The municipalities of Armenia (department of Quindio) and San Juan de Pasto (department of Nariño).

[20] This is, students completing 11 years of schooling. The first 9 years are compulsory, the last 2 are optional.

most informative and widely used indicator on local development was the Municipal Development Index (hereafter MDI[21]). 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 energy, water and sewerage systems, literacy rates and poverty ratios, and 'financial status' variables such as per capita tax revenue and public spending, and dependency on central government transfers; the higher index value, the better local development. I use the 2001 MDI index to measure the local development of municipalities at the time of the reform. As can be seen in Figures 1, 5 and 6, size and local development level are overall positively correlated but with high variation at all size ranges. Municipalities which obtained certification in 2002 had MDI values ranging from 28 to 70; the empirical analysis will use the distribution of development of certified cities to determine development quartiles.

5 Empirical framework

The aim is to identify the impact of municipal autonomy over the management of local education on student test scores, especially looking out for heterogeneous patterns that the effect might display across different levels of local development. The next subsections first introduce and then discuss the two identification strategies adopted.

5.1 Sharp Regression Discontinuity Design

The fact that in 2001 certification was assigned solely on the basis of the 100 thousand municipal population cutoff sets the conditions for a sharp regression discontinuity (RD) design. The subsequent methodological summary draws on the excellent outlines by Imbens and Lemieux [2008] and Pettersson-Lidbom [2008], to which I refer for a more detailed discussion of the RD methodology. Consider the equation (I) $Y_i = \alpha + \ \tau \ C_i + \epsilon_i$, where Y_i represents average test scores in municipality i and C_i is a dummy signaling whether municipality i was certified in education ($C_i = 1$) or not ($C_i = 0$). The consistent estimation of the treatment effect τ is hindered by the fact that most likely certification C_i is correlated with other municipal characteristics enclosed in ϵ_i. In our setup though, we know that the sole assignment rule for C_i was population count P_i, and specifically whether P_i exceeded $c = 100\,000$ or not, such that $C_i = \mathbb{1}\{P_i > c\}$ where $\mathbb{1}\{.\}$ is the indicator function. In this case we have that conditioning on population P_i will remove any correlation between C_i and ϵ_i, so that treatment C_i is as good as randomly assigned conditional on P_i[22]. Thus the ideal specification

[21] Translation from the original *Índice de Desarrollo Municipal (IDM)*. Data on the index is 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 IDM (2013 onwards).

[22] Other ways to express this is saying that the 'conditional mean independence' or 'selection on observables' or 'unconfoundedness' assumption holds, $E[\epsilon_i | C_i, P_i] = E[\epsilon_i | P_i]$.

of a 'control function' $h(P_i)$ is such that its insertion into our equation (I) will completely purge it from any dependence between C_i and ϵ_i [Heckman and Robb, 1985]. In practice it is difficult to guess the ideal functional form of $h(P_i)$ and thus a common approach is to use flexible functions such as high order polynomials of P_i[23]. In order to ensure that τ is capturing only the effect of the treatment, we also need all other characteristics of the observed units not to change discontinuously at the treatment cutoff. In our case, we need municipal characteristics other than certification in education not to exhibit 'jumps' at 100 thousand inhabitants, otherwise we would not know which part of the estimated effect on test scores is due to certification and which part is due to the other changes happening at 100 thousand inhabitants. Smoothness of municipal characteristics was discussed in section 3.3.1, and further evidence on pre-reform smoothness of test scores is given through the falsification tests in section 12.

I estimate the model

$$Y_i = \alpha + \quad \tau^{RD} \quad C_i + f(P_i) + \epsilon_i \tag{1}$$

where $f(P_i)$ is approximated by a third oder polynomial in P_i, and interpret τ^{RD} as the average treatment effect of certification in education. I then introduce the municipal development variable D_i, expressing the MDI indicator illustrated in section 4.2. Heterogeneity across levels of local development can be explored either by applying 1 to different subsamples, or by introducing an interaction term between certification and development, obtaining

$$Y_i = \alpha + \tau_0^{RD} C_i + \tau_1^{RD} C_i \cdot D_i + \beta D_i + f(P_i) + \epsilon_i \tag{2}$$

where $\tau_0^{RD} + \tau_1^{RD} D_i$ can be interpreted as the average treatment effect of certification at development level D_i. Section 6 shows the estimation results of 1 on the full sample of municipalities and on four subsamples split by development of certified cities (upper half and lower half, highest quartile and lowest quartile), and the estimation results of 2.

Regression discontinuity designs are notably data demanding and rarely free of obstacles (more on this in the discussion section 5.3). The fact that only forty cities obtained certification in 2002 and that their population sizes are not all clustered at 100 thousand inhabitants poses difficulties in terms of available sample size and precision. I am able to reduce sampling variance by using ten post-reform years of data (2002-2012); including year fixed effects is not necessary in this setup as the test score outcomes are standardized each year according to the yearly national performance. Further discussion of the RD estimation is available in Section 5.3.

[23] This is sometimes referred to as 'global polynomial series' estimator. Other estimators of the treatment effect in a RD setting are kernel estimators, and estimators based on trimming data close to the boundary such as local linear regression or other nonparametric methods. In a recent working paper, Gelman and Imbens [2014] recommend to prefer the latter group to the global polynomial method I employ here. These methods are data thirsty and require high numbers of observations close to the boundary, a luxury that is unavailable in this setting.

5.2 Fixed Effect Regression on a Discontinuity Sample

An alternative method for estimating the average treatment treatment effect of certification in education on student test scores is applying the fixed effects (FE) concept, which exploits over-time variation in the performance of municipalities that acquire certification. The basic FE model reads as follows:

$$Y_{it} = \alpha + \tau^{FE} C_{it} + \boldsymbol{M_i} + \mathbf{T_t} + \epsilon_{it} \tag{3}$$

where test scores in municipality i and year t, Y_{it}, are regressed on certification status C_{it}, which takes value 1 in years from 2002 onwards for municipalities who obtained certification, and is 0 otherwise. Municipality fixed effects are $\boldsymbol{M_i}$ and time fixed effects are $\boldsymbol{T_t}$; the effect of certification is captured by τ^{FE}.

I limit the sample to municipalities with a number of inhabitants close to the certification cutoff, both from the left and from the right, in order to avoid confounders such as municipal characteristics varying with population size to threaten identification. Following Angrist and Lavy [1999], I refer to this as our 'discontinuity sample'. We are thus estimating 3 on a sample of cities that are similar to each other in size, out of which some acquired education autonomy ($C = 1$) in 2002 and some did not.

For the main specification in the empirical analysis I use municipalities between 80 thousand and 130 thousand inhabitants - which results in a sample of thirty cities, eleven of which acquired certification in 2002 and nineteen did not[24], and whose population counts and development indices are illustrated in Figure 1 with dark bars and light bars respectively[25]. Table 2 shows some relevant summary statistics separately for certified and non-certified municipalities, and highlights the similarity of the two groups in terms of pre-reform characteristics - including pre-reform test score levels. The thirty cities in the discontinuity sample account for about 8.35% of the student population enrolled in primary and secondary school in 2012[26].

In order to pursue our goal of identifying heterogeneity in the effects of autonomy by levels of local development, model 3 is augmented with an interaction term between certification status C_{it} and development measure D_i, obtaining

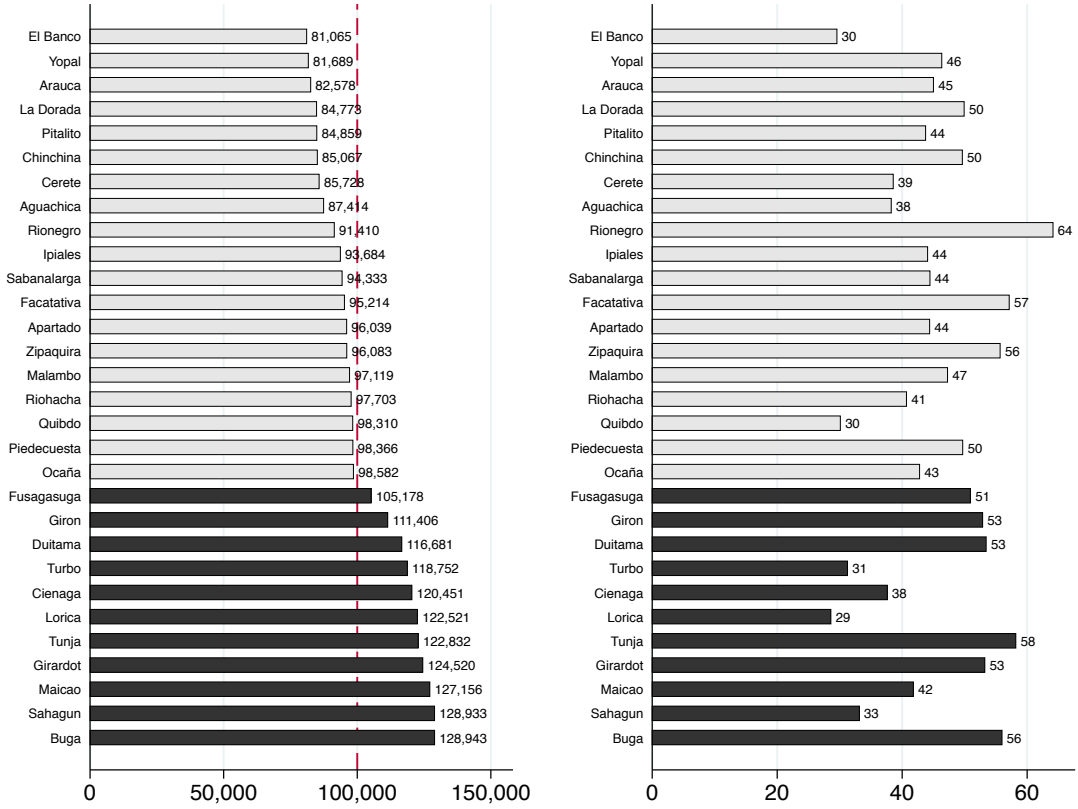
$$Y_{it} = \alpha + \tau_0^{FE} C_{it} + \tau_1^{FE} C_{it} \cdot D_i + \boldsymbol{\gamma M_i} + \boldsymbol{\delta} \mathbf{T_t} + \epsilon_{it} \tag{4}$$

[24]Results are robust to extending or restricting the sample; regressions on different samples, including the same sample used for the RD estimation, are presented in Table in the Appendix.

[25]Figure 6 in the Appendix shows the two distributions for a wider range of municipalities.

[26]Author's calculation, based on enrollment data provided by MEN (Ministry of Education).

Figure 1: Population and MDI distribution of the 30 municipalities around the inhabitant cutoff



where the effect of certification at development level D_i will be given by the estimates of $\tau_0^{FE}+\tau_1^{FE}D_i$. Section 6 shows results of this fixed effects approach.

5.3 Discussion of identification

As anticipated in section 5.1, the probably most salient difficulty of applying the RD methodology to the context of this paper is the limited sample size available, which exposes the analysis to both low power and potentially excessive small sample variation. In fact the confidence levels at which the null is being rejected in the result section are not particularly impressive[27]. Moreover, also cities distant from the treatment cutoff are used for estimation of the RD model[28], which relies on the assumption that the population polynomial $f(P_i)$ is able to 'control' for those municipal characteristics that vary with size and may confound the effect of autonomy on test scores. The less

[27]The limited amount of available data points also prescribes parsimony in the number of model parameters to estimate. The parsimonious regression model 1 is therefore chosen for the main specification instead of the RD approach that leaves two different sets of parameters on the two sides of the treatment cutoff (see Imbens and Lemieux [2008] for a discussion). Table 17 in the Appendix shows result of the latter method too.

[28]All cities between 10 and 500 thousand inhabitants.

Table 2: Municipalities in the discontinuity sample (80,000 - 130,000 inhabitants)

	Certified (N=11)	Non certified (N=19)	Difference	
Population in 1992	99,998	73,182	26,816	***(1,820)
Population in 2002	120,670	91,043	29,627	***(692)
Population in 2012	127,756	112,305	15,451	***(3,705)
Municipal Development Index (MDI) 2001	40.99	40.59	0.40	(1.92)
Unsatisfied Basic Needs indicator (UBN) 1993	45.18	45.34	-0.16	(0.96)
Saber 11 Math score 2001	40.38	40.45	-0.07	(0.39)
Saber 11 Language score 2001	45.27	45.33	-0.06	(0.62)
Public primary school gross enrollment rates 2001	0.67	0.67	-0.00	(0.06)
Public secondary school gross enrollment rates 2001	0.61	0.61	0.00	(0.05)

Standard error of mean difference in parentheses; *p<0.10 **p<0.05 ***p<0.01

one needs to move away from the cutoff (the larger the sample close to the cutoff), the more likely it is that such confounders are properly eliminated. Despite these drawbacks, results display a fairly stable path across subsamples and are robust to different specifications (see section 17.1).

As anticipated in Section 3.2.1, one may hold that the 2001 reform implied not only an increase in autonomy for the cities which obtained certification, but also some loss of autonomy for the cities which did not, leading to a certain extent of 'inverse treatment' in the control group[29]. In the case the reform had induced changes in the education trends of 'untreated' municipalities too, both the regression discontinuity and the fixed effects analysis would be affected through alterations in their control groups. One would then need reject the interpretation of estimated results as the effects of an 'increase in autonomy' (or effects of decentralization), and rather look at results as the effects of 'an autonomy gap' (or effects of authority polarization). These interpretational issues are however confined to the background as one keeps in mind the primary objective of capturing heterogeneity in the effect of autonomy across different levels of the local development spectrum.

[29]For the reasons and context explained in Sections 3.2 and 3.2.1, the author's assessment is that this scenario ought not to be excessively worried about.

6 Results

6.1 Regression Discontinuity results

6.1.1 Baseline results

Table 3 shows estimation results for the regression discontinuity models (1) and (2), on Mathematics and Spanish student test scores on the 2002-2012 period. In all cases I have excluded from the analysis municipalities of special sizes, namely those below 10 thousand and above 500 thousand inhabitants (municipal categories "Special" and "Sixth" - see footnote 18). Column (1) of each panels report the outcome of model (1). The estimate of the average effect of certification in education on municipal test scores is close to zero. Columns (2) to (5) of each panel explore heterogeneities in the effect, considering different subsamples of the 2001 MDI distribution of certified cities: model (1) is applied to the lower and upper 50% development range and to the bottom and top 25% of the range. As anticipated in Section 4.2, development quartiles are constructed referring to the distribution of local development of *certified* cities. Figure 5 in the Appendix illustrates how in correspondence of lower levels of this distribution we find a larger number of smaller, non-certified cities. A larger 'control group' at low levels of development is what causes low-development sample sizes to be larger than the high-development ones throughout the analysis.

Looking at the result pattern, a test score gap appears to be opening between the most and the least developed autonomous cities. More precisely, high developed cities who become autonomous do better and low-developed cities do worse than their non-certified counterparts[30]. The effect on the high developed group is larger and more precisely estimated. Table 12 in the Appendix shows that in the pre-reform period (years 2000 and 2001 data) this pattern is not visible, and neither there is any evidence for differential score growth in the different development subsamples. In Section 14 in the Appendix I perform formal tests on the difference between pre-reform and post-reform RD coefficients.

The magnitudes of the effects are sizable: negative 1.5 points (0.15 student standard deviations) for certified cities in the least developed quartile and positive 2 points (0.2 student standard deviations) for certified cities in the most developed quartile. The three panels in the first column of Figure 2 depicts these estimation results graphically.

Columns (6) in Table 3 show the estimation of model (2), where certification status is linearly interacted with the development percentile to which each municipality belongs, as an alternative way to capture heterogeneity in the effect. This second estimation approach confirms the pattern

[30]Keeping in mind that test scores are nationally standardized, the post-reform bifurcation could arise because certified municipalities change their performance and non-certified ones remain static, or the other way round, or a mix of both effects. Refer to the discussion in Section 5.3.

previously emerged: the effect of certification is increasing in MDI values, starting negative for low MDI values and becoming positive at higher ones.

6.1.2 Consolidation over time

After looking at the average effect on scores over the whole post-reform period 2002-2012, I will now concentrate on years further away from the reform date. Cohorts of high school students taking the Saber 11 exam in later years have been exposed to the reform for longer[31]; moreover one should allow for certified municipalities to gradually implement their medium and long-horizon education plans. Table 4 shows how the estimated effect on test scores grows in absolute values as model (1) is run on periods further and further away from the reform date (time periods starting in 2004, 2007 and 2010). Figure 3 uses point estimates and confidence intervals from Table 4 to illustrate how certified cities in the top 25% development range progressively increase their score gap with respect to their non-certified counterparts, while the opposite happens for cities in the bottom 25% development range. Looking at the last period (2010-2012), the point estimates have reached about a third of a student standard deviation into both directions. In Section 13 of the Appendix I show analogous over-time changes in the effect estimations obtained using mutually exclusive year bins instead of progressively later time periods.

6.2 Fixed effects regression results

The results of the fixed effects identification strategy described in Section 5.2 are shown in Table 5, for both the basic model (3) and the specification that allows for effect heterogeneity in development (4). The sample is composed of the 30 municipalities closest to the inhabitant cutoff: 19 non-certified ones with more than 80 thousand inhabitants and 11 certified ones with less than 130 thousand inhabitants. The outcome variables are municipal test score averages in years 2000 to 2012. The first two columns of each panel refer to the basic model, showing OLS and municipality fixed effects estimations. The third and fourth column show OLS and fixed effects estimations of the main specification, using the MDI 2001 as a proxy for municipal development. The average effect of certification is estimated through the basic model as close to zero and statistically not significant. The model allowing for heterogeneity across development levels unveils the same pattern that was detected through the regression discontinuity identification: the point-estimated effect of autonomy

[31] In the spirit of the exercise performed by Galiani et al. [2008] in their paper on Argentinian school decentralization.

Table 3: Effect of certification on Saber11 test scores - Main RD estimation

(a) Mathematics

	(1) All	(2) Bottom 25%	(3) Bottom 50%	(4) Top 50%	(5) Top 25%	(6) Interaction
Certified	0.17	-1.58	0.40	0.73	2.20**	-2.37***
	(0.63)	(0.99)	(0.94)	(0.87)	(0.86)	(0.84)
Certif. * MDI '01						0.06***
						(0.02)
MDI '01						0.07***
						(0.01)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	7,572	6,536	7,100	472	275	7,561
R-sq.	0.013	0.003	0.003	0.011	0.050	0.084

Standard errors clustered by municipality in parentheses

* p<.10 ** p<.05 *** p<.01

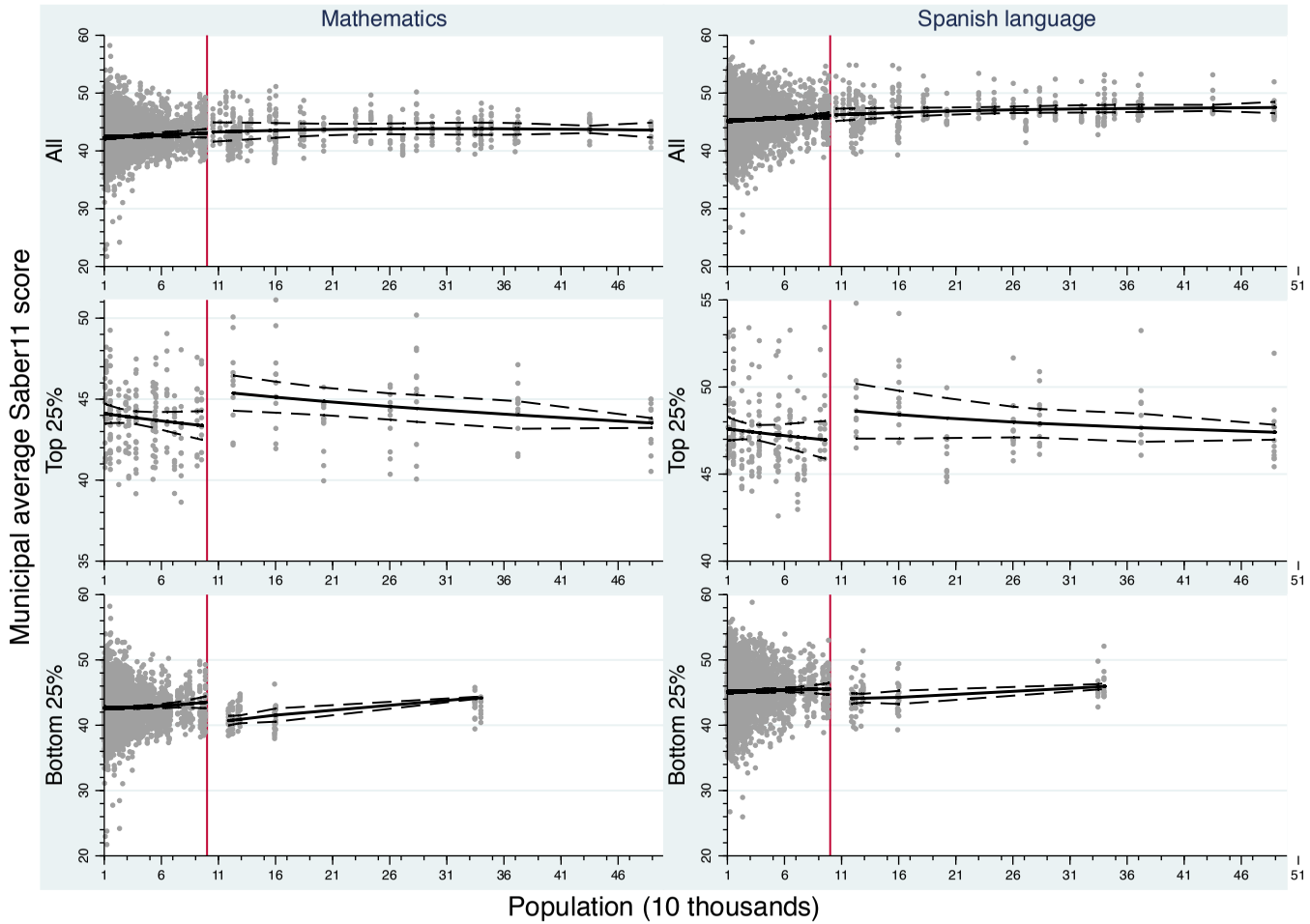
(b) Spanish Language

	(1) All	(2) Bottom 25%	(3) Bottom 50%	(4) Top 50%	(5) Top 25%	(6) Interaction
Certified	0.07	-1.55	0.32	0.52	1.81	-2.11**
	(0.66)	(1.00)	(0.94)	(0.90)	(1.14)	(0.96)
Certif. * MDI '01						0.05**
						(0.02)
MDI '01						0.09***
						(0.01)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	7,572	6,536	7,100	472	275	7,561
R-sq.	0.018	0.003	0.004	0.011	0.035	0.123

Standard errors clustered by municipality in parentheses

* p<.10 ** p<.05 *** p<.01

Figure 2: Certification on Saber 11 scores - Graphical results



Municipal averages of Saber 11 scores against population. Left column for Mathematics, right column for Spanish language. Solid lines are predicted scores, dashed lines are 95% confidence intervals on the prediction.

Table 4: Certification on Saber 11 test scores - progress over time

[Regression Discontinuity Estimation]

(a) Top 25% MDI '01

	Mathematics			Spanish Language		
	(1)	(2)	(3)	(4)	(5)	(6)
	Post 2004	Post 2007	Post 2010	Post 2004	Post 2007	Post 2010
Certified	2.37**	3.00***	3.80**	1.64	1.36	1.92
	(0.92)	(1.06)	(1.52)	(1.09)	(1.04)	(1.29)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	225	150	75	225	150	75
R-sq.	0.069	0.137	0.199	0.038	0.126	0.137

(b) Bottom 25% MDI '01

	Mathematics			Spanish Language		
	(1)	(2)	(3)	(4)	(5)	(6)
	Post 2004	Post 2007	Post 2010	Post 2004	Post 2007	Post 2010
Certified	-1.80	-2.23*	-3.17**	-1.47	-1.61	-2.03*
	(1.12)	(1.29)	(1.60)	(1.02)	(1.06)	(1.12)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	5,344	3,609	1,809	5,344	3,609	1,809
R-sq.	0.003	0.005	0.007	0.003	0.004	0.007

Standard errors clustered by municipality in parentheses
* p<.10 ** p<.05 *** p<.01

goes from being between one and two points negative at low levels of development, crosses the zero threshold at a MDI level of around 45 and grows to reach a positive value of around two points at MDI levels of 70, as illustrated graphically in Figure 4. These magnitudes are very similar to the ones estimated with the RD technique for the lowest and highest development quartile[32].

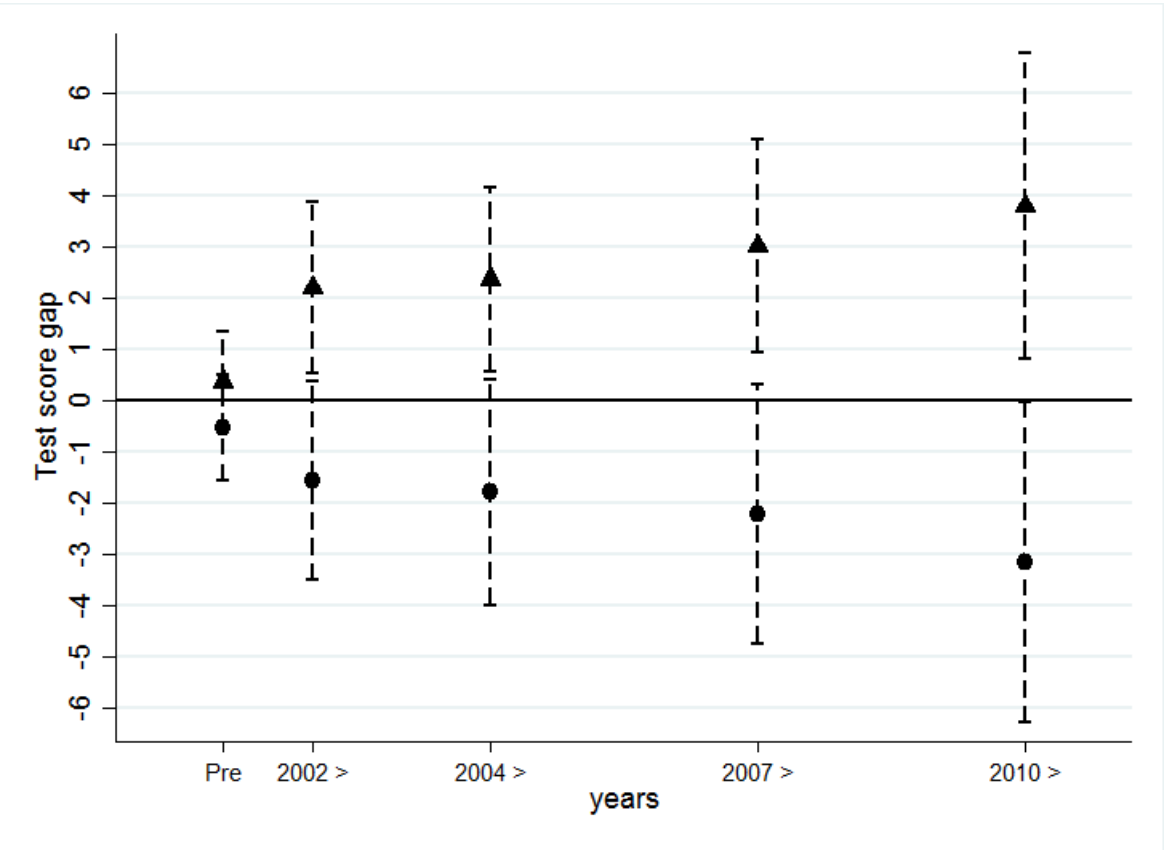
6.3 Discussion of main results

Two different strategies have been employed to isolate the effect of local autonomy in the education service on quality of education, as measured by student test scores, at different levels of municipal development. Both techniques yielded the conclusion that cities at the upper end of the development range benefitted from the acquired autonomy while those at the lower end did not, and rather took

[32]Even though they preserve the qualitative pattern, results for Language are again less imposing, as in the RD strategy.

Figure 3: Effect of certification over time

[Regression Discontinuity Estimation]



RD estimations of the effect of certification on average Math test scores, for certified municipalities in the top 25% and the bottom 25% development range (triangles and circles series respectively). Capped spikes indicate 95% confidence intervals on point estimates.

loss on it, giving rise to a test score gap between the two groups that appears to be widening over time.

This section concludes with the reflection that the two identification strategies adopted rely on different assumptions but are shown to yield effect estimations which are qualitatively alike and quantitatively very similar, which represents a rewarding result on its own. Moreover, the regression discontinuity design estimates the effect of autonomy *for city sizes around 100 thousand*, while the fixe d effects methodology estimates the *average* effect of higher autonomy *across all certified cities* included in the sample[33]. By comparing the two sets of results, we thus learn that the effect magnitudes seem to be fairly stable across city sizes ranging from around 100 thousand to around

[33]In the main results, cities between 80 and 130 thousand inhabitants; in the robustness checks, cities between 50 and 250 thousand and between 10 and 500 thousand inhabitants.

Table 5: Effect of certification on Saber 11 test scores - Main FE estimation

	Mathematics				Spanish language			
	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE	(7) OLS	(8) FE
Certified	0.06	0.02	-3.02*	-3.67***	-0.05	-0.01	-1.93	-0.92
	(0.76)	(0.53)	(1.64)	(1.26)	(0.80)	(0.26)	(1.64)	(0.55)
Certif * MDI '01			0.07*	0.08**			0.04	0.02*
			(0.04)	(0.03)			(0.04)	(0.01)
MDI '01			0.09***				0.13***	
			(0.02)				(0.02)	
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	389	389	389	389	389	389	389	389
N groups		30		30		30		30
R-sq.	0.39	0.67	0.58	0.68	0.38	0.77	0.63	0.77

Standard errors clustered by municipality in parentheses; * p<.10 ** p<.05 *** p<.01

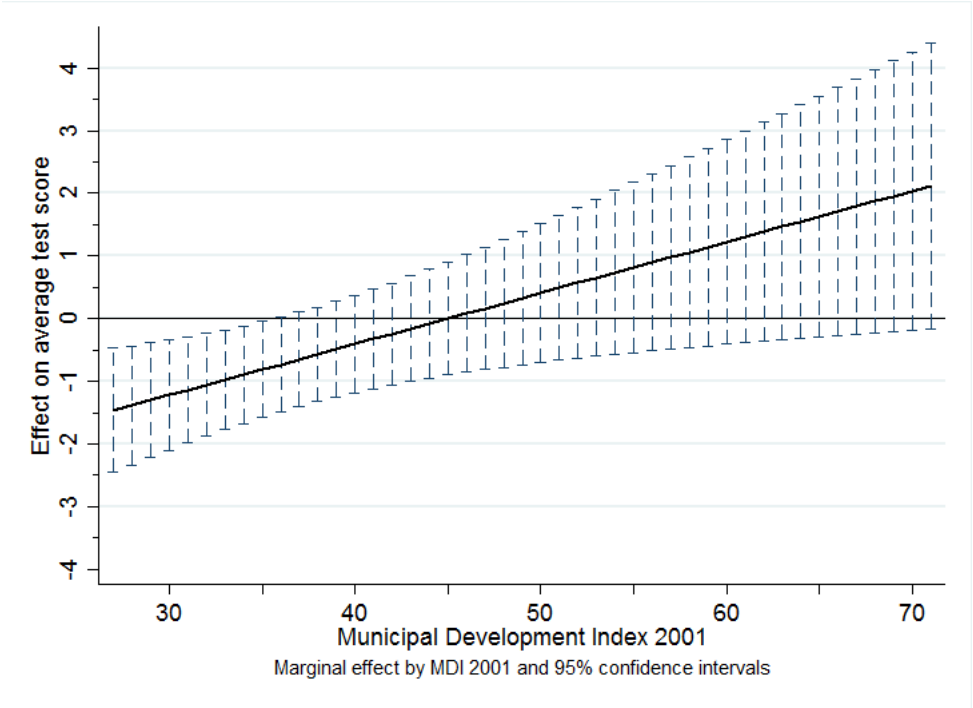


Figure 4: Effects of certification at different development levels (FE estimation)

500 thousand inhabitants[34].

[34]Considering also the exercises of variation of the sample range that are performed in the robustness checks (Table 19).

In the remaining sections of the paper the focus will lie upon cities in the highest and lowest quartiles of the certified development range, as these are the two groups on which significant reform effects have been identified. The goal will be exploring these effects in further detail and providing explanations for the test score dynamics found.

6.4 Compositional effects, migration and public-private education

This section addresses the question of whether the over-time changes in test scores purely reflect changes in student performance (an 'intensive margin' result), or whether the pool of test takers has also been changing as a result of the reform ('compositional effect' or 'extensive margin'). The pool of test takers may change if we observed responses to the reform such as selective migration (into or away from the newly autonomous municipalities), switching of students between public and private schools, or changes in student school dropout patterns. Tables 6 and 7 show regression discontinuity and fixed effects estimations respectively, of test taker characteristics on the municipal autonomy indicator (certification status). Characteristics being looked at are number of test takers, share of female students, share of students whose mother is low educated or high educated, and the share of students who work while studying. In the RD results, the only statistically significant pattern we are able to spot is an apparent shift of high-educated families away from low-developed municipalities and into high-developed ones, which might suggest some degree of selective migration of the better educated families. Looking at the FE specifications though, at development ranges which are relevant to certified municipalities (MDI = 29 to 70 approx.), the magnitudes of these shifts are estimated close to zero. Patterns on low-educated mothers are never statistically significant and the share of working students does not exhibit changes. Overall it seems prudent to conclude that with the available data I am not able to pin down any clear and robust compositional effects on Saber 11 test candidates, as a consequence of the decentralization reform.

The absence of significant compositional changes in the pool of test candidates is in line with what one would expect after considering two basic facts about Colombia's school population[35]. The first is the significant and persistent gap between the quality of private and public education. Private schools score substantially and consistently better on standardized tests such as the national ICFES or international PISA results (Cerquera et al. [2000]; Barrera-Osorio et al., 2012; Gamboa and Waltenberg, 2011), have smaller class sizes[36] and are four times more likely to offer full-day school programs with respect to public schools, which instead see double the frequency of morning-only or late hours programs ([Bonilla Mejía, 2011]. In sum, it is fair to say that private education in Colombia, as well as in the rest of Latin America, is still a privilege restricted to well-off families (also

[35] In addition to the fact that the publicity of this regime change on mass media has been very limited.

[36] Approximately 35 students per teacher in public schools and 25 students per teacher in private schools (averaged over the period 1998-2008). Author's own calculations using national statistics office education data (DANE C-600).

Table 6: Municipal certification and test taker characteristics (RD)

		(1) N. takers	(2) Share female	(3) Low M.Ed.	(4) High M.Ed.	(5) Working
a) All		–127.91	–0.00	0.04	0.00	0.01
		(99.84)	(0.01)	(0.03)	(0.01)	(0.01)
b) Bottom 25%		–262.10	–0.00	0.10	–0.04**	0.01
		(163.87)	(0.00)	0.06	(0.03)	(0.03)
c) Top 25%		81.04	0.01	–0.04	0.10**	0.03
		(460.65)	(0.02)	(0.06)	(0.04)	(0.03)
d) All, w. interaction		185.83	0.02	0.03	–0.05*	0.06
		(185.86)	(0.03)	(0.06)	(0.02)	(0.04)
Certif×MDI'01		–6.71	–0.00	0.00	0.00*	–0.00
		(4.79)	(0.00)	(0.00)	(0.00)	(0.00)
Mean Y (All)		321◇	0.54	0.60	0.06	0.12
N [n. municip.]	a)	7,523 [698]	7,523 [698]	4,821 [697]	4,821 [697]	4,829 [698]
	b)	6,488 [603]	6,488 [603]	4,158 [602]	4,158 [602]	4,166 [603]
	c)	275 [25]	275 [25]	175 [25]	175 [25]	175 [25]
	d)	7,512 [697]	7,512 [697]	4,814 [696]	4,814 [696]	4,821 [697]

RD regressions of different outcome variables (in columns) on certification status and a third degree population polynomial. Cells show coefficient and standard errors on the certification regressor. The rows refer respectively to: a) All municipalities; b) municipalities in the bottom 25% devlopment range; c) municipalities in the top 25% development range; d) All municipalities, with main effect and development interaction term. SEs clustered by municipalities in parentheses. * p<0.10, ** p<0.05, *** p<0.01. ◇ The mean of certified municipalities is $1,947$ test takers.

see Gamboa and Waltenberg [2011]for a discussion). The implementation of a reform that shifts responsibility over public schools from the regional to the municipality level would not be expected to close the gap between public and private education, or to make public institutions significantly more attractive to well-off families. The second fact to keep in mind relates to the first: the family of the typical public-school student in Colombia is less likely to be informed about a decentralization reform occurring, to form strong predictions about its effects on educational quality, or to have the means and opportunity to migrate to a different municipality.

6.5 Heterogeneity across people

Along with the heterogeneity in effects across municipalities, heterogeneity across people within municipalities is a dimension that ought to be looked at. In this section I look at how autonomy has impacted the dispersion of test scores in cities of the high and low developed group. Moreover,

Table 7: Municipal certification and test taker characteristics (FE)

		(1) N. takers	(2) Share female	(3) Low M.Ed.	(4) High M.Ed.	(5) Working
a) All		31.50	–0.00	–0.02	0.00	0.00
		(68.72)	(0.01)	(0.02)	(0.01)	(0.02)
b) All, w. interaction		128.70	–0.00	–0.04	–0.03*	–0.09
		(270.37)	(0.05)	(0.04)	(0.02)	(0.05)
Certif×MDI'01		–2.15	–0.00	0.00	0.00**	0.00
		(5.03)	(0.00)	(0.00)	(0.00)	(0.00)
Mean Y		971	0.55	0.49	0.08	0.11
N [n. municip.]	a)	389 [30]	389 [30]	270 [30]	270 [30]	270 [30]
	b)	389 [30]	389 [30]	270 [30]	270 [30]	270 [30]

Municipal FE regressions of different outcome variables (in columns) on certification status. Cells show coefficient and standard errors on the certification regressor. The rows refer respectively to: a) All municipalities; b) All municipalities, with main effect and development interaction term. SEs clustered by municipalities in parentheses. * p<0.10, ** p<0.05, *** p<0.01

I am interested in investigating whether students from different socioeconomic backgrounds have been differently affected by local autonomy on public education.

Using the self-reported background information on Saber11 test takers, I divide students by social status as proxied by their mother's education (ME): low mother education for compulsory education (9 school years) or less, high mother education is education beyond compulsory. Information on mother education is available for all years excluding 2005, 2006 and 2007[37]. Table 8 shows estimates obtained applying our main regression discontinuity model (1) on test score standard deviations (SD), and then on test scores of students belonging to the two social background categories.

The picture that emerges from these results is that students with higher social background seem to be more susceptible to changes in local autonomy: they gain more in high-developed cities and they lose more in low-developed ones, with respect to lower social background students.

7 Channels

7.1 Expenditure on education

In the pursuit of the reasons behind heterogeneous educational outcomes between high-developed and low-developed autonomous cities, the perhaps most straightforward starting point is expen-

[37]Reason for the smaller sample sizes on the mother education (ME) specifications, with respect to the standard deviation (SD) specifications.

Table 8: Effect of local autonomy on score dispersion and by social background

(a) Mathematics

	Bottom 25% cities			Top 25% cities		
	(1)	(2)	(3)	(4)	(5)	(6)
	score SD	Low ME	High ME	score SD	Low ME	High ME
Certified	-0.45*	-1.56	-3.05*	0.43*	2.28**	2.76***
	(0.24)	(1.26)	(1.80)	(0.23)	(0.97)	(0.91)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	6,524	4,128	3,932	275	175	175
R-sq.	0.002	0.002	0.002	0.007	0.046	0.029

(b) Spanish Language

	Bottom 25% cities			Top 25% cities		
	(1)	(2)	(3)	(4)	(5)	(6)
	score SD	Low ME	High ME	score SD	Low ME	High ME
Certified	-0.13	-1.34	-2.85**	0.02	1.47	1.66*
	(0.15)	(1.06)	(1.40)	(0.23)	(1.19)	(0.88)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	6,524	4,128	3,932	275	175	175
R-sq.	0.004	0.002	0.004	0.007	0.033	0.035

diture choices. Using detailed balance sheet data[38] of municipalities in the highest and lowest development quartiles, in Table 9 I perform t-tests on the mean expenditures of the two groups over the post-reform period 2002-2012. Central government transfers that municipalities receive to finance education services (*SGP Educación*) are also compared, as well as the resulting difference between spending and transfers. Recall that spending on education exceeding government transfers is covered by municipalities using their own resources, which are represented mainly by local tax and fees collection, and by capital gains[39].

What emerges from the comparisons performed in Table 9 is that the average per-pupil expendi-

[38]"Ejecuciones municipales, formato largo". Reported yearly by municipalities to the government agency DNP (Departamento Nacional de Planeación). Source: Universidad de Los Andes, Bogotá, 2015.

[39]Examples of local tax and fees are the housing and land ownership tax, tax on gasoline consumption, traffic fines. Examples of capital gains are interests on municipal accounts and rents from municipal-owned infrastructure and land.

ture by municipalities in the highest development quartile is almost 23% higher than the one of municipalities in the lowest development quartile. Within education expenditure, the difference on personnel salaries is around 13%, up to 30% on school infrastructure and material and as much as 63% higher on other education expenses and programs[40]. The asymmetry in spending is not matched by any asymmetry in central government resources received. In fact, while the low-developed group appears to be spending on education barely as much as it receives in education transfers, the high-developed group is integrating transfers with own resources, for around 12% of their total education spending[41]. The differences in spending behavior uncovered through analysis of municipal balance data provides at least suggestive evidence towards the explanation of student test score dynamics previously identified.

Table 9: Per-pupil expenditure and transfers received

	All certified	Low 25%	High 25%	Δ H - L	Δ%
A) Education spending	1160.58	1021.83	1282.54	260.71***	22.66%
	(402.76)	(406.27)	(438.28)	(81.28)	
- Salaries	930.44	836.58	954.25	117.67**	13.14%
	(302.62)	(319.52)	(280.06)	(58.12)	
- Infrastr. and material	98.46	103.95	140.94	37.00*	30.21%
	(90.20)	(93.11)	(132.88)	(21.90)	
- Others	82.65	52.65	100.82	48.17***	62.77%
	(93.85)	(73.70)	(97.78)	(16.56)	
B) Education transfers	1153.41	1114.91	1134.53	19.61	1.74%
	(415.96)	(383.34)	(319.94)	(57.13)	
A) - B)	7.17	–93.08	148.01**		
	(34.23)	(66.73)	(54.26)		
N.obs (expenditure)	240	56	50	106	
N.obs (transfers)	345	88	70	158	
N. municipalities	35	9	7	16	

Table of mean annual per-pupil expenditures and central government transfers received (2002-2012, in thousands of Colombian pesos) and t-tests on the mean differences. Standard deviations in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

[40]Examples of the most frequent balance sheet items in this category are school transport, teacher formation, planning and development of school information systems, investments in efficiency of the municipal education management authority, contracts with private institutions for additional education services.

[41]Table 15 in the Appendix shows that municipalities in the higher developed group have been enjoying higher availability of own resources both before and after the reform. The resource gap between the two groups has been significantly narrowed in post-reform years, mainly through compensatory transfers from the central government.

7.2 Administration indicators

Table 10 summarizes data on municipal evaluation processes that are carried out by the government on a yearly basis[42]. The 'Legal compliance rate' (*Índice de Cumplimiento de Requisitos Legales*) indicates the extent to which the municipality administration is found to adhere to the national norms in its use of government transfers, in all sectors of activity. The frequency of detection of accounting irregularities and illicit use of funds determine the rating received through this index [DNP, 2014]. The 'Managerial capabilities index' (*Índice de Capacidad Administrativa*) measures the extent to which the municipal administration appears suitable and prepared to perform its tasks thoroughly and to promote local development. The elements factoring into the index are the stability of managerial employees, the level of competency of clerks, the availability of suitable IT equipment and the level of automation of administrative processes [DNP, 2011]. The takeaway from Table 10 is that there are significant differences in these administration quality indicators between high-developed and low-developed municipalities, in the direction one would expect. This holds true for both municipalities that were certified (panel a)) and for smaller municipalities (panel b)). Adding to the evidence on expenditure behavior, the striking differences in these quality indicators provide further suggestive evidence towards the channels through which test score dynamics may have come about. Municipalities in the high-development range, characterized by higher availability of financial resources and a more pro-education spending policy, along with higher quality administration capabilities, were able to improve education quality on their territories with respect to the centralized management. The opposite has been true for cities in the low-development range, with fewer local resources and worse management skills.

7.3 Oaxaca decomposition of test score differences

Having documented the existence of significant differences in education expenditure and administration indices between highly developed and low developed municipalities, it is desirable to gauge the extent to which such differences are able to explain the gap in student performance between the two groups of cities. Table 11 shows the results of the decomposition technique proposed by Oaxaca [1973] and Blinder [1973], which splits test score gaps into explained and unexplained components[43]. The results show that around one third of the gap is attributable to raw differences in

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

[43]This is sometimes known as the 'twofold' decomposition approach. See the excellent illustration by Jann [2008] for reference. The baseline coefficient vector is obtained by regressing student test scores on per-pupil expenditure, the two administration quality indices and the third-degree population polynomial on the sample of all certified cities. Baseline coefficients are then used to analyze the score differential betwen cities in the highest and the lowest development quartile. The differential is decomposed into a part that is explained by group *differences in predictors* ("Explained" component, or "quantity effect"), and a part that is attributed to *different returns to predictors* across the two groups ("Unexplained" component).

Table 10: Municipal administration indices

a) Certified municipalities	All certified	Low 25%	High 25%	Δ H - L
Managerial capabilities	67.99	46.01	79.12	33.11***
	(25.08)	(26.84)	(16.77)	(4.38)
Compliance rate	77.87	69.74	79.18	9.43*
	(22.95)	(29.45)	(21.50)	(5.01)
N.obs	245	63	49	112
N. municipalities		9	7	16
b) All municipalities	All	Low 25%	High 25%	Δ H - L
Managerial capabilities	63.67	61.98	80.39	18.41***
	(25.25)	(25.56)	(17.39)	(1.99)
Compliance rate	74.65	73.73	83.35	9.6***
	(22.23)	(22.66)	(16.52)	(1.77)
N.obs	4841	4187	168	4355
N. municipalities				

Table of evaluation indices (mean 2005-2012, scales 1-100) and t-tests on the mean differences. Standard deviations in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

expenditure and administration quality, with the latter playing the by far larger role. According to this decomposition, if highly developed cities recorded the same per-pupil education expenditure and the same administration quality indices as we find among low developed cities, their average test score performances would have been one point (or 10% of a student standard deviation) lower.

Differences in expenditure quantities and administration indices do not directly account for the remaining two thirds of the test score gap: it is *returns* to expenditure and especially to administration capacity that appear to differ significantly between highly- and low developed cities. These differences in returns are likely to capture differences between the two groups of cities which are not currently being accounted for.

8 Conclusion

In this paper I have taken advantage of an unusually favorable reform setting to show that cities characterized by different levels of local development have reacted differently to higher autonomy over the public education service. Levels of municipal development embody the wealth of local

Table 11: Oaxaca decomposition of test score differences

	Mathematics			Spanish Language		
	2005	2007	2010	2005	2007	2010
Differential						
Prediction Low dev.	42.067***	41.959***	41.665***	43.484***	43.659***	43.057***
	(0.42)	(0.46)	(0.64)	(0.37)	(0.36)	(0.45)
Prediction High dev.	45.600***	45.813***	46.621***	47.166***	47.250***	47.499***
	(0.46)	(0.49)	(0.63)	(0.40)	(0.41)	(0.46)
Difference	-3.533***	-3.854***	-4.956***	-3.682***	-3.591***	-4.442***
	(0.62)	(0.68)	(0.90)	(0.55)	(0.54)	(0.65)
Explained						
Expenditure	-0.174	-0.181	-0.180	-0.069	-0.045	-0.093
	(0.15)	(0.17)	(0.24)	(0.11)	(0.13)	(0.19)
Admin. capacity	-0.827***	-0.922***	-1.377**	-0.553***	-0.558**	-1.040**
	(0.28)	(0.33)	(0.62)	(0.21)	(0.23)	(0.42)
Legal requirements	-0.037	-0.032	-0.008	-0.057	-0.041	-0.001
	(0.06)	(0.06)	(0.04)	(0.05)	(0.05)	(0.02)
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
Total	-1.010**	-1.094**	-1.697**	-0.997**	-0.966**	-1.564***
	(0.42)	(0.49)	(0.77)	(0.39)	(0.38)	(0.56)
Unexplained						
Expenditure	-0.950	-0.205	2.139	1.520**	1.920	3.300*
	(0.84)	(0.79)	(1.65)	(0.72)	(1.34)	(1.85)
Admin. capacity	-2.564**	-2.491**	-4.728***	-2.907***	-2.803***	-4.120***
	(1.08)	(1.17)	(1.48)	(0.91)	(1.04)	(1.35)
Legal requirements	-0.401	-0.370	0.562	-0.284	-0.100	0.709
	(0.68)	(0.83)	(1.49)	(0.54)	(0.50)	(1.04)
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
Total	-2.522***	-2.760***	-3.259***	-2.685***	-2.625***	-2.878***
	(0.62)	(0.69)	(0.98)	(0.57)	(0.56)	(0.73)
N	110	94	47	110	94	47

Oaxaca-Blinder decomposition of test score gaps between highly and low developed municipalities, into explained and unexplained components. Columns indicate time periods from 2005, 2007 and 2010 onwards respectively. All models include population controls (third degree polynomial). Standard errors clustered by municipality in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

population and the amounts of own financial resources available. In the ten years following the handover of responsibilities, cities in the highest development quartile have significantly improved their student's test score performance with respect to non-autonomous counterparts; cities in the lowest development quartile instead display the opposite test score trend. The test score gaps are growing stronger over time. The high-developed group of cities invests in education more than the ad

hoc financial transfers it receives from the central government, while cities in the lowest development quartile barely invest their financial allocation. The largest differences in investment shares between the two groups are on "other education programs", which include teacher formation, school transport, planning and maintenance of school information systems and investments into the efficiency of the local education authority itself. Spending differences are also found on the infrastructure and school material investments, and on school staff. Moreover, high developed municipalities perform significantly better on different administration quality indicators with respect to low developed cities, which suggests additional explanations for why, once given autonomy on the delivery of the public service, their results have started drifting apart.

The findings of this study sound a note of caution in the design of decentralization reforms in contexts in which subnational heterogeneity in wealth and development is an issue. When handing responsibilities in public service delivery to the local level, less advantaged localities may need additional training and support in order to avoid regional inequality to grow, and decentralization to backfire for segments of the population.

References

A. Alesina, A. Carrasquilla, and J.J. Echavarría. Decentralization in Colombia. *Fedesarrollo Working Paper Series*, 15, Aug 2000.

J. Angrist and V. Lavy. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–75, 1999.

P. Bardhan and D. Mookherjee. Capture and governance at local and national level. *American Economic Review [PaP]*, 90(2):135–39, 2000.

P. Bardhan and D. Mookherjee. Decentralization of governance and development. *Journal of Economic Perspectives*, 16(4):185–205, 2002.

P. Bardhan and D. Mookherjee. Decentralizing antipoverty program delivery in developing countries. *Journal of Public Economics*, 89(4):675–704, Apr 2005.

P. Bardhan and D. Mookherjee. Decentralisation and accountability in infrastructure delivery in developing countries. *The Economic Journal*, 116, Jan 2006.

Felipe Barrera-Osorio, Darío Maldonado, and Catherine Rodríguez. Calidad de la educación básica y media en Colombia: Diagnóstico y propuestas. *Serie Documentos de Trabajo Universidad del Rosario*, 126, October 2012.

T. Besley and R. Burgess. The political economy of government responsiveness: Evidence from India. *The Quarterly Journal of Economics*, 117(4):1415–1451, Nov 2002.

H. Blair. Participation and accountability at the periphery: Democratic local governance in six countries. *World Development*, 28(1):21–39, Jan 2000.

Alan S. Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, pages 436–455, 1973.

Leonardo Bonilla Mejía. Doble jornada escolar y calidad de la educación en colombia. *Documentos de trabajo sobre Economía Regional - Banco de la República*, 143, April 2011.

G.J. Borjas and O.L. Acosta. Education reform in Colombia. *Fedesarrollo Working Paper Series*, 19, Aug 2000.

A. Breton and A. Scott. *The Economic Constitution of Federal States.* University of Toronto Press, Toronto, 1978.

P. Caballero, editor. *Foro "Descentralización y certificación en educación, experiencias y desafíos, el caso colombiano"*, Bogotá, D.C., Mar 2006. PREAL GDyÁ - Conversemos sobre educación.

Daniel Cerquera, Paula Jaramillo, and Natalia Salazar. La educación en Colombia: evolución y diagnostico. Boletines de divulgación económica, Departamento Nacional de Planeación, Bogotá, D.C., 2000.

C. Corte Constitucional. Sentencia su.559/97. Technical report, Corte Constitucional de Colombia, 1997.

D. Cortés. Do more decentralized local governments do better? An evaluation of the 2001 decentralization reform in Colombia. *Serie documentos de trabajo - Universidad del Rosario*, 84, Jun 2010.

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

DNP DDTS. Evaluacion del sistema general del participaciones 2003. Technical report, Departamento Nacional de Planeación - Dirección de Desarrollo Territorial, Bogotá, D.C., Dec 2004.

DNP. Documento conpes social n.57. Technical report, Departamento Nacional de Planeación, Bogotá, D.C., Jan 2002.

DNP. Documento conpes social n.89. Technical report, Departamento Nacional de Planeación, Bogotá, D.C., Dec 2004a.

DNP. Documento conpes social n.77. Technical report, Departamento Nacional de Planeación, Bogotá, D.C., Jan 2004b.

DNP. Evaluación del desempeño integral de los municipios. Technical report, Departamento Nacional de Planeación - Dirección de Desarrollo Territorial, Bogotá, D.C., 2011.

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

J.P. Faguet. Does decentralization increase responsiveness to local needs? Evidence from Bolivia. *Journal of Public Economics*, 88:867–94, 2004.

J.P. Faguet and F. Sanchez. Decentralization's effects on educational outcomes in Bolivia and Colombia. *World Development*, 36(7):1294–1316, 2008.

J.P. Faguet and F. Sanchez. Decentralization and access to social services in Colombia. *Public Choice*, 160(1-2): 227–249, Jul 2014.

C. Ferraz and F. Finan. Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311, 2011.

S. Galiani, P. Gertler, and E. Schargrodsky. School decentralization: Helping the good get better, but leaving the poor behind. *Journal of Public Economics*, 92:2106–20, 2008.

Luis Fernando Gamboa and Fábio D. Waltenberg. Inequality of opportunity in educational achievement in latin america: Evidence from pisa 2006-2009. *ECINEQ - Society for the Study of Economic Inequality*, August 2011.

Andrew Gelman and Guido Imbens. Why high-order polynomials should not be used in regression discontinuity designs. *NBER Working Paper No. 20405*, Aug 2014.

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

G.W. Hammond and M.S. Tosun. The impact of local decentralization on economic growth: Evidence from U.S. counties. *Journal of Regional Science*, 51(1):47–64, 2011.

J.J. Heckman and Richard Robb. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1):239–267, 1985.

G. Imbens and G. Lemieux. Regression Discontinuity Designs: A guide to practice. *Journal of Econometrics*, 142(2): 615–635, Feb 2008.

A.M. Iregui B., L. Melo B., and J. Ramos. Evaluación y análisis de eficencia de la educación en colombia. Technical report, Banco de la Republica, Colombia, Bogotá, D.C., Feb 2006.

Ben Jann. The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4):453–479, 2008.

J. Juetting, E. Corsi, C. Kauffmann, I. McDonnell, H. Osterreider, N. Pinaud, and L. Wegner. What makes decentralization in developing countries pro-poor? *The European Journal of Development Research*, 17(4):626–48, 2005.

K. Kaiser. Decentralization reforms. In A. Coudouel and S. Paternostro, editors, *Analyzing the Distributional Impact of Reforms*, volume 2. World Bank, Washington, 2006.

I. Lonzano, J. Ramos, and H. Rincón. Implicaciones fiscales y sectoriales de la reforma a las transferencias territoriales en Colombia. *Borradores de Economía - Banco de la Republica, Colombia*, 437, Apr 2007.

J. Manor. *The Political Economy of Democratic Decentralization*. World Bank, Washington, Mar 1999.

MEN. Directiva ministerial n.04/2003. Technical report, Ministero de Educación Nacional, Bogotá, D.C., Mar 2003.

R. Musgrave. *The Theory of Public Finance*. McGraw-Hill, New York, 1959.

W. Oates. *Fiscal Federalism*. Harcourt Brace Jovanovich, New York, 1972.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, pages 693–709, 1973.

U. Panizza. What drives fiscal decentralisation? *CESifo DICE Report 1/2004*, 2004.

Per Pettersson-Lidbom. Do parties matter for economic outcomes? A regression-discontinuity approach. *Journal of the European Economic Association*, 6(5):1037–1056, Sep 2008.

R. Reinikka and J. Svensson. Local capture: Evidence from a central government transfer program in Uganda. *The Quarterly Journal of Economics*, 119(2):678–704, May 2004.

Dennis A. Rondinelli, John R. Nellis, and G. Shabbir Cheema. Decentralization in developing countries: A review of recent experience. *World Bank Staff Working Papers - Management and Development Series*, 581(8), Jul 1983.

M. Santa Maria S., N. Millan U., J. Moreno B., and C.F. Reyes. La descentralización y el financiamento de la salud y la educación en los departamentos: Â¿Cuáles son las alternativas? Informe final, Federación nacional de departamentos y Fedesarrollo, Dec 2009.

A. Sarmiento and J.E. Vargas. Descentralización de los servicios de educación y salud en colombia. *Conyuntura Social - Fedesarrollo / Instituto SER de Investigación*, 16:91–136, May 1997.

R. John Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.

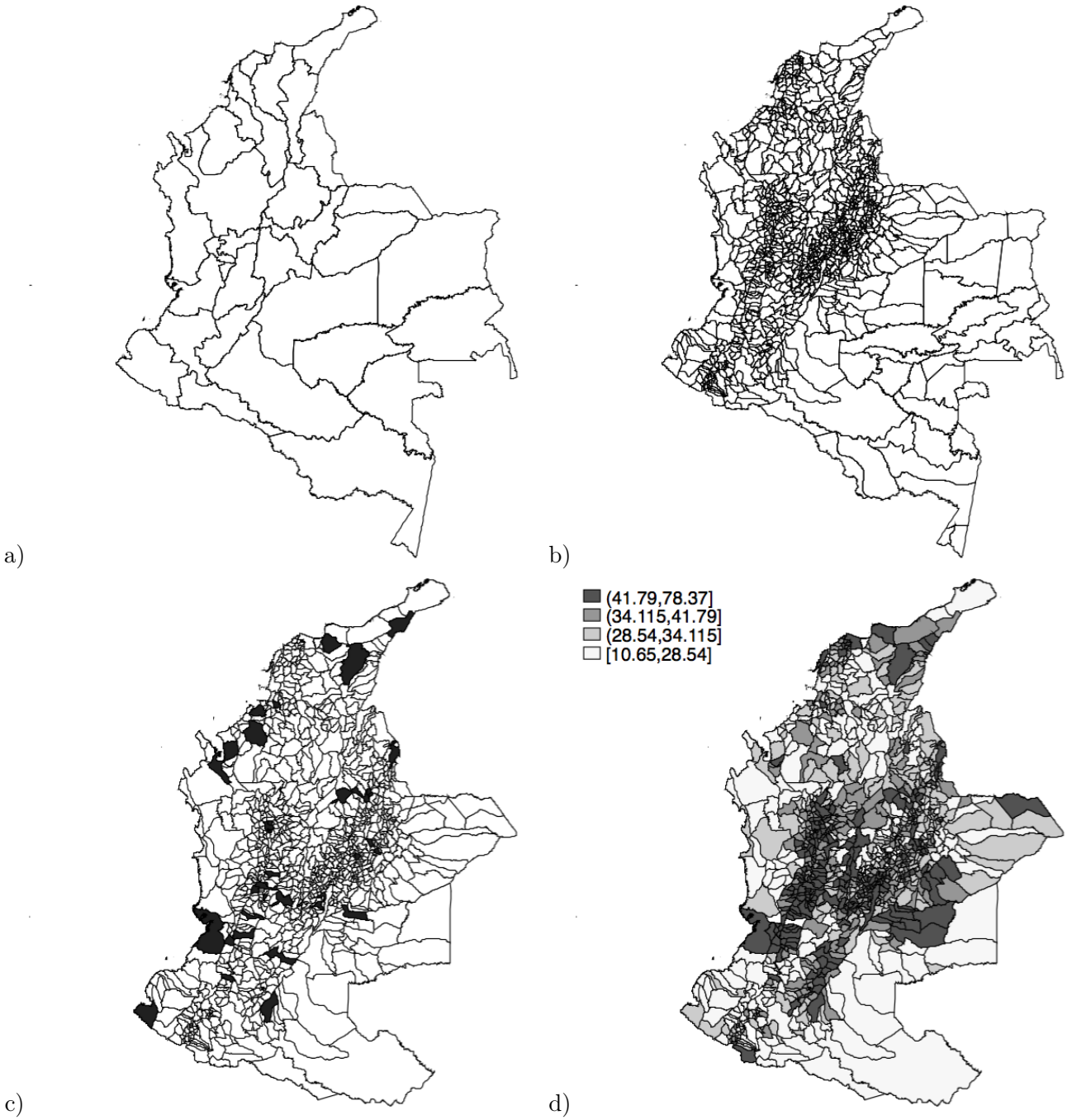
G. Toro. Trabajamos por la descentralizacion, la gobernabilidad y la autonomia local. Presentacion, Federación Colombiana de Municipios (FCM), Jun 2006.

C.H. Vergara and M. Simpson. 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, Dec 2001.

X. Zhang. Fiscal decentralization and political centralization in China: Implications for growth and inequality. *Journal of Comparative Economics*, 34(4):713–726, Dec 2006.

Appendix

9 Maps



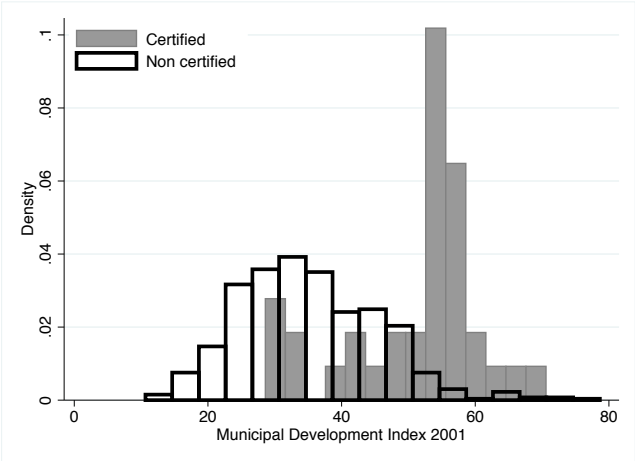
a) Colombia's departments; b) Colombia's municipalities; c) Municipalities which were certified in education in 2002 (in black); d) Distribution of Municipal Development Index in 2001. In maps c) and d) the rural south-east is omitted to allow larger zoom on the densely populated area.

10 Population and Municipal Development Index distributions

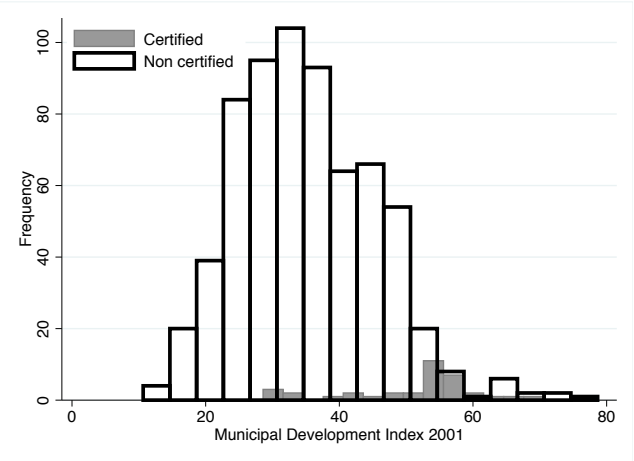
Figure 5 shows the distributions of the municipal development measures by certification status. Figure 6 extends Figure 1, illustrating population and MDI distributions for a wider range of municipalities.

Figure 5: Distribution of MDI by certification

(a) Densities



(b) Frequencies



(c) Scatterplot

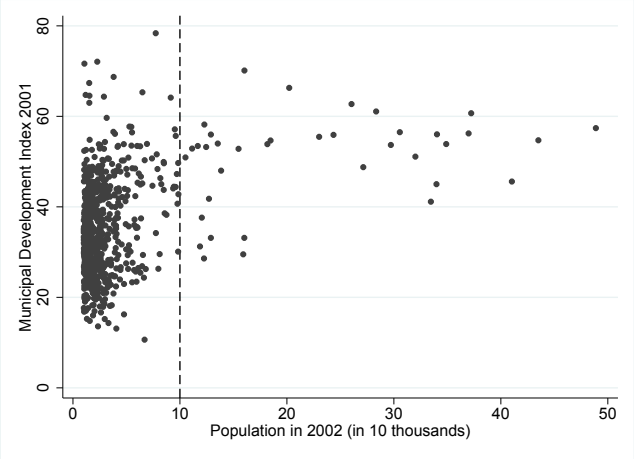
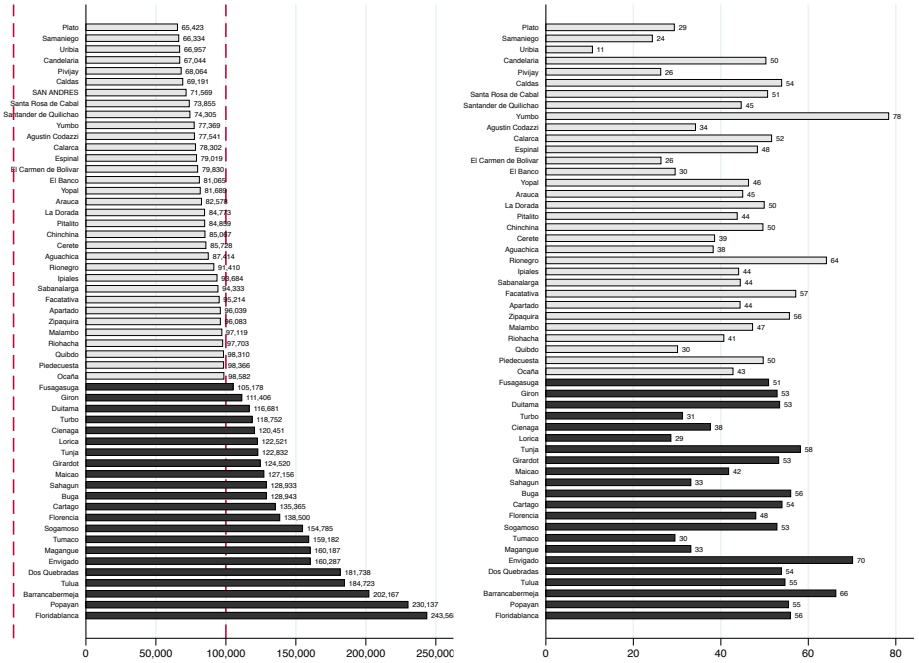
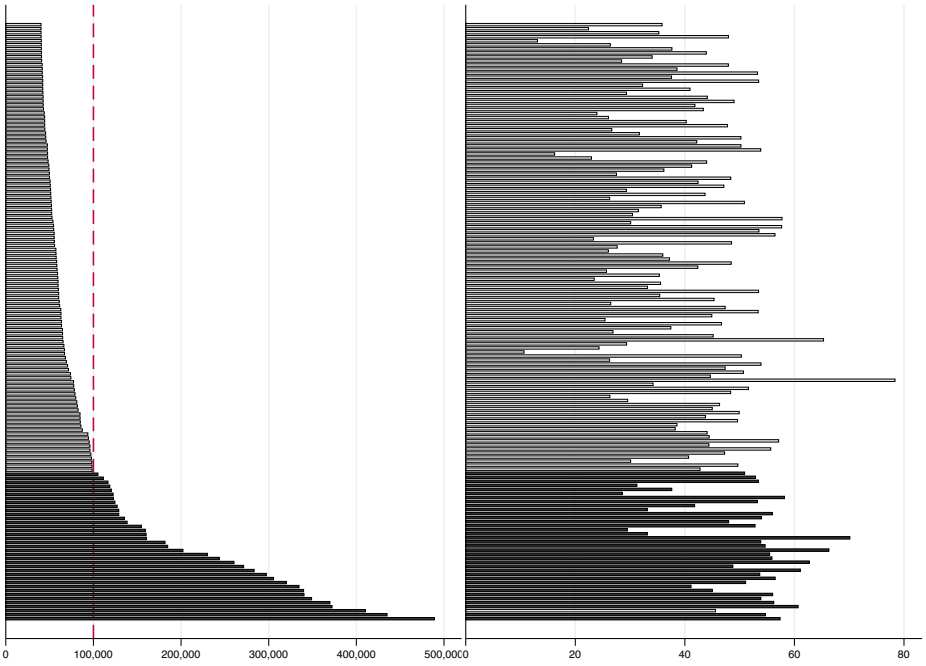


Figure 6: Population and MDI distributions

(a) 65 to 250 thousand inhabitants

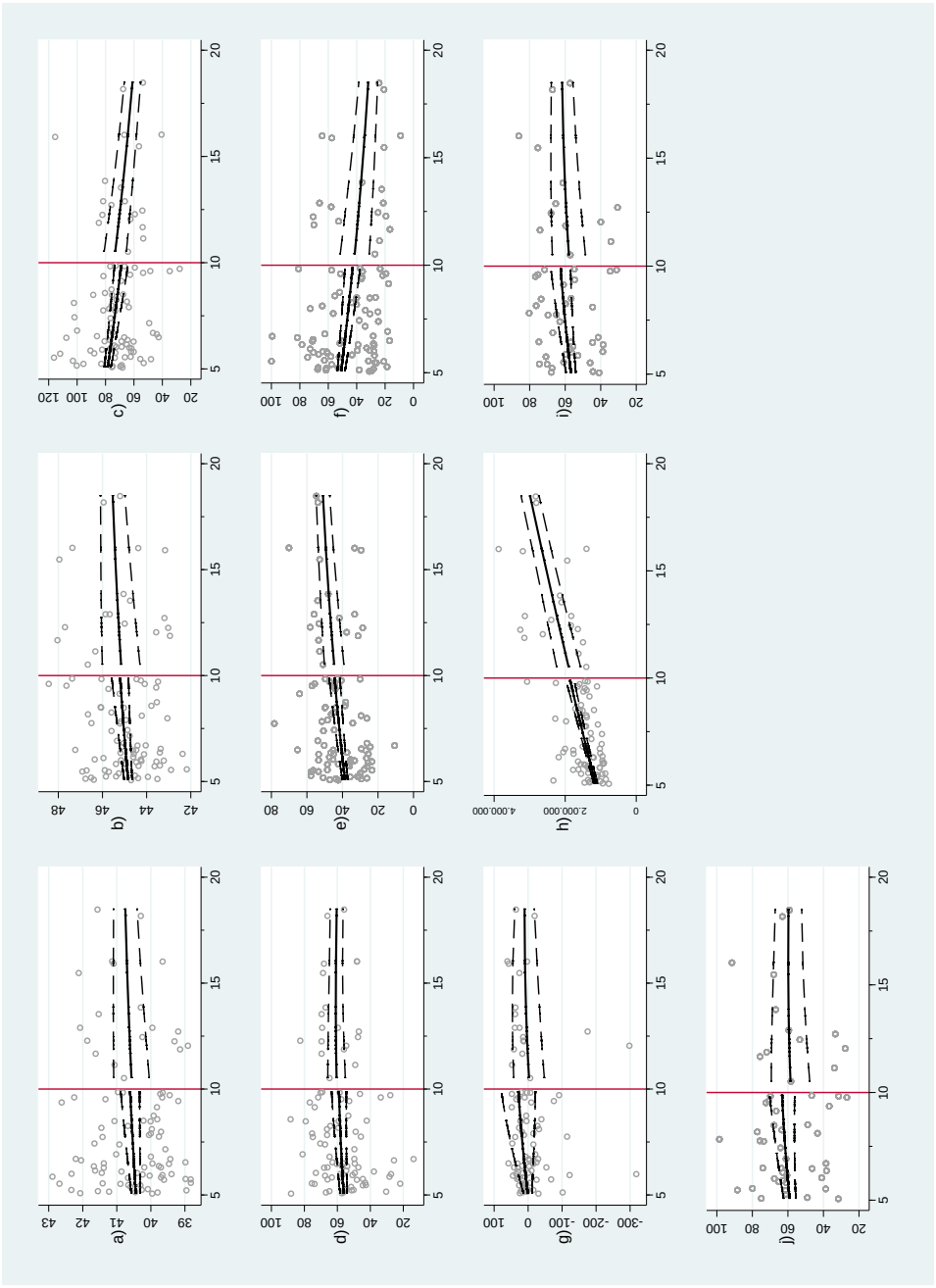


(b) 40 to 500 thousand inhabitants



11 Smoothness checks

Figure 7: Smoothness of municipal characteristics across the discontinuity



a) Municipal Maths score average 2001 (value); b) Municipal Spanish Language score average 2001 (value); c) Gross primary school coverage 2001 (percent); d) Gross secondary school coverage 2001 (percent); e) Municipal Development Index 2001 (value); f) Unsatisfied Basic Needs indicator 1993 (value); g) Share of saved municipal current revenues (percent); h) Central Govt. transfers for education 2001 (1,000s of Pesos); i) Transparency index 2005 (value); j) Visibility and accountability index 2005 (value).

12 Common pre-reform trend and falsification test

In order to dissolve residual doubts about whether municipalities on the two sides of the certification cutoff may have been evolving in different ways over time also in absence of the reform, I perform a test on the pre-reform trend. Lamentably the available pre-reform years of test score data are only two (2000 and 2001), thus the test will look at the changes between those two years only: the variable $\Delta Y = Y_{2001} - Y_{2000}$ is the outcome variable in panel (a) of Table 12. No discontinuities nor patterns are discovered in the results, neither across the four development subsamples nor through the interaction term specification, confirming our belief that cities above and below the treatment cutoff are not intrinsically different from each other. Panel (b) of the same table shows the results of the RD estimation on pre-reform data, in order to verify that the test score patterns and discontinuities identified in the main results were not already existing before the treatment. Again the subsample analysis does not reveal any particular relationship between scores and development before autonomy, while the interaction term specification does show a pattern qualitatively similar to the post-reform scenario but significantly weaker in magnitude. The overall conclusion I draw from the two panels of Table 12 is that before the 2001 decentralization reform, among municipalities sized around 100thousand inhabitants, there was no evident relationship between development measures and student test scores levels or growth rates. That relationship emerged only once cities were endowed with decisional and financial autonomy over the public education service.

13 Progress over time using time bins

Expanding on Section 6.1.2, here I provide a different approach to the analysis of over-time behavior of the certification effect. Instead of looking at progressively later time periods as done in Table 4, Table 13 and Figure 8 show results of RD estimations performed on successive and mutually exclusive 3-year bins. Standard errors are larger with respect to the previous approach, as fewer data points enter each bin; the result patterns remain the same and average effects can be observed growing over time.

Table 12: Common-trend test and falsification test - RD estimation

(a) Pre-reform trend in scores (outcome is $\Delta Y = Y_{2001} - Y_{2000}$)

	Mathematics					Spanish Language				
	(1) B 25%	(2) B 50%	(3) T 50%	(4) T 25%	(5) Int.	(6) B 25%	(7) B 50%	(8) T 50%	(9) T 25%	(10) Int.
Certified	-0.461	-0.272	-0.436	0.468	-0.075	0.178	-0.589	0.393	-0.309	0.364
	(0.42)	(0.27)	(0.56)	(0.59)	(0.52)	(0.59)	(0.40)	(0.57)	(0.60)	(0.79)
Certif.*MDI '01					-0.002					-0.011
					(0.01)					(0.01)
MDI '01					0.011**					-0.010*
					(0.00)					(0.01)
.										
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	581	632	43	25	674	581	632	43	25	674
R-sq.	0.00	0.00	0.06	0.10	0.01	0.00	0.00	0.05	0.04	0.01

(b) Pre-reform scores (2000 and 2001)

	Mathematics					Spanish Language				
	(1) B 25%	(2) B 50%	(3) T 50%	(4) T 25%	(5) Int.	(6) B 25%	(7) B 50%	(8) T 50%	(9) T 25%	(10) Int.
Certified	-0.534	-0.166	0.731	0.362	-1.813***	-1.294*	0.544	-0.159	1.760	-1.531*
	(0.53)	(0.40)	(0.43)	(0.50)	(0.62)	(0.76)	(0.82)	(1.21)	(1.25)	(0.86)
Certif.*MDI'01					0.040***					0.032*
					(0.01)					(0.02)
MDI '01					0.010**					0.075***
					(0.00)					(0.01)
.										
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,177	1,280	86	50	1,364	1,177	1,280	86	50	1,364
R-sq.	0.00	0.00	0.08	0.06	0.01	0.01	0.01	0.03	0.08	0.22

Standard errors clustered by municipality in parentheses. * p<.10 ** p<.05 *** p<.01

Table 13: Certification on Saber 11 test scores - progress over time

[Regression Discontinuity Estimation]

(a) Top 25% MDI '01

	Mathematics				Spanish Language			
	(1) 2002-2004	(2) 2005-2007	(3) 2008-2010	(4) 2011-2012	(5) 2002-2004	(6) 2005-2007	(7) 2008-2010	(8) 2011-2012
Certified	1.591**	1.456*	2.209**	4.222**	2.642	1.331	1.134	2.294
	(0.77)	(0.83)	(0.86)	(1.69)	(1.57)	(1.15)	(0.85)	(1.41)
.								
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	75	75	75	50	75	75	75	50
R-sq.	0.072	0.119	0.134	0.260	0.054	0.075	0.143	0.164

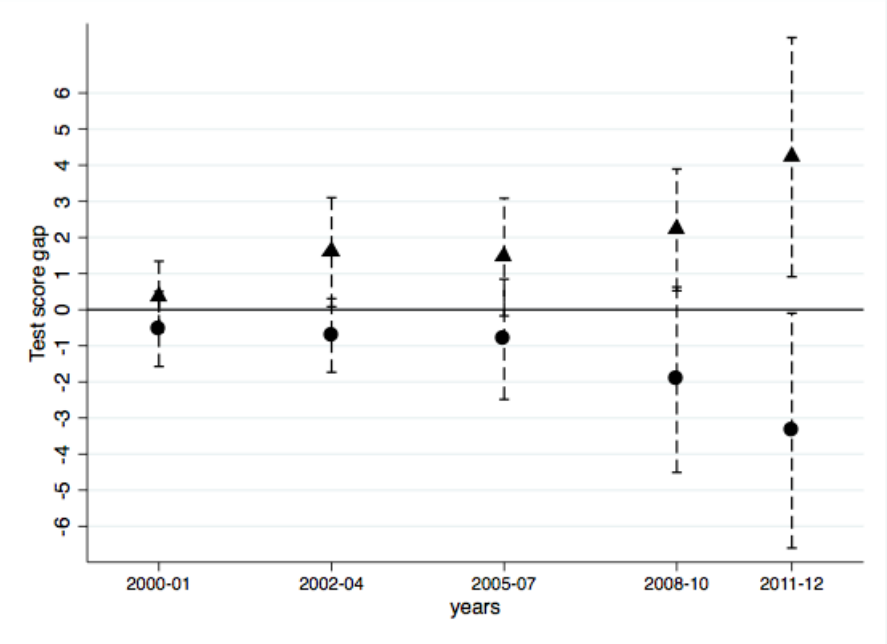
(b) Bottom 25% MDI '01

	Mathematics				Spanish Language			
	(1) 2002-2004	(2) 2005-2007	(3) 2008-2010	(4) 2011-2012	(5) 2002-2004	(6) 2005-2007	(7) 2008-2010	(8) 2011-2012
Certified	-0.714	-0.820	-1.940	-3.350**	-1.523	-1.404	-1.412	-2.195*
	(0.52)	(0.85)	(1.31)	(1.66)	(1.04)	(1.13)	(0.89)	(1.33)
.								
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,731	1,796	1,803	1,206	1,731	1,796	1,803	1,206
R-sq.	0.004	0.003	0.005	0.007	0.003	0.004	0.002	0.008

Standard errors clustered by municipality in parentheses
* p<.10 ** p<.05 *** p<.01

Figure 8: Effect of certification over time

[Regression Discontinuity Estimation]



RD estimations of the effect of certification on average Math test scores, for certified municipalities in the top 25% and the bottom 25% development range (triangles and circles series respectively). Capped spikes indicate 95% confidence intervals on point estimates.

14 Testing the difference between pre- and post-reform coefficients

In this section I formally test the difference between the RD coefficient estimates obtained using post-reform data (Section 6.1) and the 'falsification' coefficients obtained using pre-reform data (Section 12), for both the high and the low development groups. Table 14 shows difference estimates and the associated standard errors. Focusing on mathematics test scores, for high-developed cities the differences between pre-and post-reform coefficients are statistically significant on every time span considered. For the low-developed group instead, statistical difference is reached only once we look further away from the reform date, at the years 2010-2012, suggesting a slower emergence of the reform effects on this group. Coefficients on language scores do not reach statistically significant differences between pre- and post-reform years.

15 Financial resources of municipalities

In Table 15 I show how the level of local development embodies significant differences in the amount of financial resources available to the municipal administration of autonomous cities. Central government transfers do implement some redistribution, but differences in local tax collection and capital

Table 14: Differences between pre-and post-reform coefficients

	Mathematics				Spanish Language			
	Post 02	Post 04	Post 07	Post 10	Post 02	Post 04	Post 07	Post 10
a) Top 25%	1.72*	1.89*	2.52**	3.32**	1.53	1.36	1.08	1.64
	(0.87)	(0.93)	(1.08)	(1.51)	(0.94)	(0.88)	(0.84)	(1.12)
b) Bottom 25%	–1.05	–1.27	–1.70	–2.64*	–0.95	–0.86	–1.00	–1.42
	(0.98)	(1.11)	(1.29)	(1.59)	(0.99)	(1.00)	(1.05)	(1.11)
N a)	325	275	200	125	325	275	200	125
N.mun. a)	25	25	25	25	25	25	25	25
N b)	7,713	6,521	4,786	2,986	7,713	6,521	4,786	2,986
N.mun. b)	603	602	603	603	603	602	603	603

Table of differences between pre-reform and post-reform coefficients of certification on student test scores. Standard errors on differences in parentheses. Pre-reform coefficients are obtained estimating RD model (1) on 2000-2001 student test scores. Post reform coefficients are obtained estimating RD model (1) respectively on 2000-2012, 2004-2012, 2007-2012 and 2010-2012 student test scores. * p<0.10, ** p<0.05, *** p<0.01.

gains sustain the advantage of high-developed cities with respect to low-developed ones.

Table 15: Percapita resources of certified municipalities in bottom and top development quartile

	Pre-reform (1998-2001)			Post-reform (2002-2012)		
	Low 25%	High 25%	Δ%	Low 25%	High 25%	Δ%
Total	154.63	343.71	75.88%***	638.61	716.24	11.46%*
	(41.68)	(122.75)		(280.38)	(250.99)	
Transfers	113.72	82.08	–32.32%**	480.20	260.36	–59.37%***
	(40.33)	(30.98)		(208.34)	(88.40)	
Tax collection	24.38	158.77	146.77%***	54.33	254.57	129.65%***
	(13.72)	(73.18)		(28.34)	(117.02)	
Capital gains	10.23	29.76	97.67%***	83.11	138.39	49.91%***
	(12.48)	(19.35)		(91.60)	(87.31)	
N.obs	23	14	37	90	64	154
N. municipalities	8	7	15	9	7	16

Table of mean annual percapita resources reported by municipalities in pre- and post-reform years (in thousands of Colombian pesos), and percentage differences between the two groups. Standard deviations in parentheses. Significance stars refer to t-tests on the mean differences, * p<0.10, ** p<0.05, *** p<0.01

16 Descriptive evolution of test scores

Table 16 illustrates test score differences between highly developed (first quartile) and low developed (last quartile) municipalities, at several points in time, for both Mathematics and Spanish language. Differences are indicated separately for the group of cities that obtained certification in 2002 and for the group that did not. In both groups, highly developed cities always see higher test scores than low developed ones, even before the decentralization reform. Nevertheless we can observe that for cities that obtained autonomy in 2002, the test score gap between high and low developed members increases in the post-reform period significantly more than what it does in the never-autonomous group of cities. These are descriptive patterns that do not represent estimations of the causal effects of the 2002 reform (refer to the main results for such estimations): instead they inform us about the bare over-time evolution of student performance in one group of cities relative to the other.

Table 16: Over time evolution in test score differences

	Pre 2002	Post 2002	Post 2004	Post 2007	Post 2010
A) Mathematics					
A1. Cities certified in 2002	1.09***	2.83***	3.16***	3.79***	4.81***
	(0.28)	(.62)	(.68)	(.82)	(1.11)
A2. Cities not certified in 2002	0.01	1.40***	1.52***	2.02***	2.74***
	(.12)	(.23)	(.25)	(.29)	(.37)
A1. - A2.	1.08***	1.43**	1.63**	1.77**	2.07*
	(.31)	(.65)	(.71)	(.87)	(1.16)
B) Language					
B1. Cities certified in 2002	2.50***	3.32***	3.33***	3.45***	4.29***
	(.62)	(3.32)	(.70)	(.68)	(.84)
B2. Cities not certified in 2002	1.61***	2.05***	2.00***	1.96***	2.64***
	(.30)	(.25)	(.24)	(.24)	(.30)
B1. - B2.	0.89	1.28*	1.32*	1.49**	1.65*
	(.69)	(.75)	(.74)	(.72)	(.88)
N.obs	1,227	6,811	5,569	3,759	1,884
N. municipalities	621	628	628	628	628

Table of differences in test scores between cities in the highest and in the lowest development quartile, all conditional on population (third-degree polynomial). Differences are indicated at progressive points in time, and separately for cities certified in 2002 (rows A1. and B1.) and for cities not certified in 2002 (rows A2. and B2.). Standard errors clustered by municipality in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

17 Robustness checks

17.1 Regression Discontinuity Estimation

17.1.1 Different polynomials at each side of the cutoff

Table 17 replicates the results shown in Table 3, now allowing for a different polynomial on each side of the certification cutoff. In econometric terms, this table shows the results of fitting the models $Y_i = \alpha + \tau^{RD} C_i + \beta D_i + f(P_i) + f(P_i) \times C_i + \epsilon_i$ and $Y_i = \alpha + \tau_0^{RD} C_i + \tau_1^{RD} C_i * D_i + \beta D_i + f(P_i) + f(P_i) \times C_i + \epsilon_i$, where $f(P_i)$ is a third-order polynomial of population P_i. The results from the main section are robust to these alternative model specifications.

17.1.2 Correction of statistical significance for multiple testing

If one considers columns (2) to (5) in Table 3 to be testing four separate hypotheses that are related to each other in unknown ways, the α levels used as benchmarks should be appropriately adjusted. Table 18 reports the results for municipalities in the highest and lowest development quartile, with significance stars corrected for multiple testing according to the method illustrated by Simes [1986]. Let $p_{(1)}, p_{(2)}, p_{(3)}, p_{(4)}$ be the p-values, ordered from smallest to largest, for testing hypotheses $H_0 = \{H_1, H_2, H_3, H_4\}$, each corresponding to one of the four development categories used. Then each $H_{(j)}$ is rejected if $p_{(j)} \leq j\alpha/n$ for any $j = 1..4$, and H_0 is rejected if all $H_{(j)}$ are rejected.

17.2 Difference-in-Differences Estimation

17.2.1 Different cutoffs for the discontinuity sample

Table 19 shows the results of Table 5 employing different choices of the discontinuity-sample. Columns (4) to (8) correspond to the same sample as used for the RD estimation.

Table 17: Certification on Saber 11 test scores - by Municipal Development Index '01 (2 polynomials)

(a) Mathematics

	(1) All	(2) Bottom 25%	(3) Bottom 50%	(4) Top 50%	(5) Top 25%	(6) Interaction
Certified	0.176	-2.259*	1.211	-0.307	3.065***	-2.267**
	(1.12)	(1.37)	(1.60)	(1.27)	(0.81)	(0.88)
Certif.*MDI'01 perc.						0.038***
						(0.01)
MDI'01 percentile						0.023***
						(0.00)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	7,572	6,536	7,100	472	275	7,561
R-sq.	0.013	0.003	0.004	0.028	0.073	0.084

(b) Spanish Language

	(1) All	(2) Bottom 25%	(3) Bottom 50%	(4) Top 50%	(5) Top 25%	(6) Interaction
Certified	0.072	-3.602***	1.274	-0.432	2.379**	-2.043**
	(1.14)	(1.31)	(1.60)	(1.18)	(1.10)	(0.96)
Certif.*MDI'01 perc.						0.035***
						(0.01)
MDI'01 percentile						0.031***
						(0.00)
.						
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes
N	7,572	6,536	7,100	472	275	7,561
R-sq.	0.018	0.003	0.004	0.029	0.073	0.124

Standard errors clustered by municipality in parentheses. * p<.10 ** p<.05 *** p<.01

Table 18: Certification on Saber 11 test scores - Significance corrected for multiple testing

(a) Top 25% MDI '01

	Mathematics				Spanish Language			
	(1) Post '02	(2) Post '04	(3) Post '07	(4) Post '10	(5) Post '02	(6) Post '04	(7) Post '07	(8) Post '10
Certified	2.20*	2.37*	3.00**	3.80*	1.81	1.64	1.36	1.92
	(0.86)	(0.92)	(1.06)	(1.52)	(1.14)	(1.09)	(1.04)	(1.29)
.								
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	275	225	150	75	275	225	150	75
R-sq.	0.050	0.069	0.137	0.199	0.035	0.038	0.126	0.137

(b) Bottom 25% MDI '01

	Mathematics				Spanish Language			
	(1) Post '02	(2) Post '04	(3) Post '07	(4) Post '10	(5) Post '02	(6) Post '04	(7) Post '07	(8) Post '10
Certified	–1.58	–1.80	–2.23	–3.17	–1.55	–1.47	–1.61	–2.03
	(0.99)	(1.12)	(1.29)	(1.60)	(1.00)	(1.02)	(1.06)	(1.12)
.								
F(Population)	Yes	Yes	Yes	Yes	Yes	Yes		
N	6,536	5,344	3,609	1,809	6,536	5,344	3,609	1,809
R-sq.	0.003	0.003	0.005	0.007	0.003	0.003	0.004	0.007

Effect of certification on student test scores. with significance stars corrected for multiple testing according to the Simes (1986) method. * p(j)<j0.10/4, ** p(j)<j0.05/4, *** p(j)<j0.01/4, for any j=1,2,3,4 of the ordered p-values.

Table 19: Different cutoffs for the discontinutiy sample (FE)

		50,000 - 250,000				10,000 - 500,000		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mate	Mate	Lang	Lang	Mate	Mate	Lang	Lang
Certified	0.526	–3.210***	0.240	–1.346***	0.775***	–2.778***	0.264**	–1.232**
	(0.35)	(0.91)	(0.18)	(0.50)	(0.22)	(0.80)	(0.13)	(0.48)
Certified * MDI '01		0.077***		0.032***		0.070***		0.029***
		(0.02)		(0.01)		(0.02)		(0.01)
.								
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,227	1,214	1,227	1,214	8,938	8,925	8,938	8,925
N groups	95	94	95	94	698	697	698	697
R-sq.	0.60	0.61	0.71	0.71	0.42	0.42	0.62	0.62

Standard errors clustered by municipality in parentheses. * p<0.10, ** p<0.05, *** p<0.01.