Economics of Education Review 86 (2022) 102196

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

With a little help from my friends: Medium-Term effects of a remedial education program targeting Roma minority

Marianna Battaglia [*,a], Lara Lebedinski [b]

[a] Department of Economics (FAE - Fundamentos del Análisis Económico), University of Alicante, Campus de San Vicente, Alicante 03080, Spain
[b] Institute of Economic Sciences, Zmaj Jovina 12, Belgrade 11000, Serbia

ARTICLE INFO

JEL classification:
I21
J15
D04

Keywords:
Remedial education
Roma
Ethnic minority
Standardized test scores

ABSTRACT

A poor-performing student can achieve better results by following the footprints of an older friend. In this paper, we study a remedial education program that takes advantage of this phenomenon. Introduced in Serbian primary schools in 2009, the Roma Teaching Assistant Program targets underachieving students belonging to the Roma minority. It assigns one person, usually Roma, to each school participating to provide support to targeted pupils and create a bridge with their community. We estimate its medium-term effects on educational attainments at the end of primary school by comparing students in schools participating and in schools that applied, but were not selected, before and after the introduction of the program. The impacts on marks and standardized test scores are modest, although positive and bigger in schools with a lower percentage of Roma. Roma students are, however, more likely to choose longer secondary school tracks, a requirement for entering higher education.

1. Introduction

Over the last decades, inequality within countries is rising both in the developed and developing world (Keeley, 2015) and high-quality education for all children independently of their socio-economic background is one important lever to reverse this trend (Abdullah, Doucouliagos, & Manning, 2015; Rodríguez-Pose & Tselios, 2009). The provision of high quality education to all pupils can help to reduce the perpetuation of inequalities from older to younger generations and bridge the achievement gaps between advantaged and disadvantaged students. Attaining more education implies getting higher salaries (Ashenfelter & Krueger, 1994; Harmon, Oosterbeek, & Walker, 2003) and thus education can help disadvantaged students escape the vicious cycle of poverty and get better opportunities and jobs than the generation of their parents. One of the most disadvantaged group in Europe is the Roma ethnic minority. In comparison to the majority of the population in the developed and middle-income countries where they mainly live, most of its members have extremely low educational attainment. A necessary condition to ensure a better life to the young generation of Roma is to provide them with an education that can help rise the economic ladder (Kertesi & Kézdi, 2011).

The Roma Teaching Assistant (RTA) program is one of the main initiatives in Europe targeting this population and aiming at increasing their educational attainment. It assigns one person, usually Roma, to each school participating to provide additional in-class and out-of-class support to students and build a bridge between their community and the school. This paper focuses on the RTA Program in Serbia,[1] where it was first introduced in 2009, and estimates its impacts on affected students in the medium-term. In the short-term, one year after its implementation, the program improved school attendance by 0.12 of a standard deviation, and benefited younger children in terms of educational achievements (Battaglia & Lebedinski, 2015). While on average there was no impact on marks for the whole population of pupils, the program was effective in keeping at school and raising marks of first-year students, by 0.296 of a standard deviation in Language and 0.284 of a standard deviation in Math. The highest impacts were observed in schools with fewer Roma as the program could be more intense in these schools. Medium-term effects, which in our case refer to 5 to 8 years of treatment, that is at the end of the primary school cycle, can substantially differ from short-term effects and it is not clear in which direction they can go: we can expect either stronger positive effects or they can dissolve. On the one hand, first, the teaching assistants gain experience

* Corresponding author.
E-mail addresses: mbattaglia@ua.es (M. Battaglia), lara.lebedinski@ien.bg.ac.rs (L. Lebedinski).

[1] Roma Teaching Assistants exist in some form in most countries where Roma live (Rus, 2006).

https://doi.org/10.1016/j.econedurev.2021.102196
Received 11 May 2021; Received in revised form 6 October 2021; Accepted 1 November 2021
Available online 1 December 2021
0272-7757/

over the years and their learning curve is steep in the first few years.[2] Second, schools might need some time to adapt to changes and find ways to use the newly available resources optimally. Third, in the medium-term children are exposed longer to the program and we can expect that longer exposure has a stronger cumulative effect. On the other hand, the intervention may be less relevant at later ages or assistants may lose their initial dedication and enthusiasm and as a result short-term effects can dissipate.

In order to investigate the RTA effects in the medium-term, we combine three sources of data: (i) the primary school final examination dataset for the years 2008 to 2018, (ii) the list of schools that applied for the RTA Program in 2009 and 2010, and (iii) a dataset containing the characteristics of the assistants in year 2015. All the information is provided by the Serbian Ministry of Education, Science and Technological Development (MoESTD). The final examination dataset includes the whole population of pupils attending the eighth and, therefore, the last grade of primary school in a given year.[3] The final examination is an external standardized test that all students who are finishing primary school take to formally complete it. The school application list contains information on school size and percentage of Roma in school at application, and reports whether the school received the program and at which point in time.

Schools needed to apply to be part of the program. We exploit the fact that, among those applying, some schools were admitted to the program while others were not. We use a difference-in-difference strategy and define as treated the schools which applied either in 2009 or in 2010, got selected and still had an assistant five to six years later.[4] We define as control the schools which applied in any of the two years and never got selected. Since treated and comparison (not admitted) schools may differ in their pretreatment characteristics, we pre-processed the data with the entropy balancing method (Hainmueller, 2012) and obtain a balanced sample in observable characteristics. We use the percentage of Roma in the school at the time of application and all outcomes (test scores, marks and dropouts, enrollment in secondary school and school choice) in the pre-treatment year to balance the data set. In our final sample there are 64 treated schools and 39 control schools. It is important to bear in mind that not all Roma students in treated schools work with the assistant. Since there is only one of them per school, she might decide to help only some students. For this reason, we also exploit the intensity of the program that depends on the number of Roma students in each school. Our results are intention-to-treat effects.

Our outcomes of interest are marks in Language and Math, the probability of not sitting the final examination and therefore complete primary school, and standardized test scores in Language and Math. We also look at secondary school enrollment and secondary school study track. Overall, the RTA does not have statistically significant effects on educational outcomes of students in treated schools, except for those in schools where the percentage of Roma at application is low and where, therefore, the program is more intensively implemented. In these schools, Roma students receive a 0.373 of a standard deviation more in the Language test and 0.350 of a standard deviation more in the Math test than their counterparts in control schools. Moreover, there is evidence that students at the margin of passing a course are less likely to be over-graded and are better prepared in the subjects examined. The program was yet 18.5 percentage points more successful in making Roma students choose longer and more demanding secondary education tracks. The assistants do not work with Non-Roma children and none of these students are directly treated by the RTA. However, in treated schools, there are positive spillovers that are reflected in higher educational attainments of Non-Roma, especially in schools with fewer Roma. We also look at the characteristics of the assistant and if their interaction with students' traits is relevant to our outcomes of interest. Overall, we observe that Roma students are more likely to keep studying in secondary education if they interact with an assistant who is a woman, holds a university degree, and is younger (i.e. below the median age of the assistants in the sample).

This paper speaks to the literature on remedial education programs and the effectiveness of such programs to improve the educational outcomes of a marginalised group.[5] The evidence for different settings suggests that in the short-term remedial education programs are effective in raising educational outcomes. A well-known remedial education program in India called Balsakhi targeted third and fourth graders in primary schools and provided every day two hours of remedial classes during regular school time (Banerjee, Cole, Duflo, & Linden, 2007). The short-term impact on average test scores was substantial, 0.14 standard deviations in the first year and 0.28 in the second year, but the large effects of the program did not persist after pupils left the program. Other remedial education programs, such as the Program for School Guidance in Spain or a remedial math course in Mexico, have found similar effects sizes in different settings (García-Pérez & Hidalgo-Hidalgo, 2017; Gutiérrez & Rodrigo, 2014). The Spanish Program for School Guidance, implemented in both primary and secondary schools, was successful in raising test scores in reading by 0.09 to 0.17 standard deviations (García-Pérez & Hidalgo-Hidalgo, 2017). In Mexico City, a low-cost intervention targeting low-performing students in secondary schools had an impact of 0.21 and 0.26 standard deviations (Gutiérrez & Rodrigo, 2014).

The analysis of medium-term or even long-term effects of educational programs is, however, less common in the literature. One of the most studied programs is the Head Start program. The experience with Head Start suggests that short-term findings might not necessarily translate to long-term effects. For instance, Currie & Thomas (1995) find that the short-term effects on test scores in the case of Head Start are quickly lost for the disadvantaged African-American group of pupils. On the other hand, Lavy, Kott, & Rachkovski (2018) analyze the effect of a high school remedial education program in Israel, almost two decades after its implementation. The program provided low-performing students with additional instruction in order to prepare them for the matriculation exam and succeeded in raising the matriculation rate by 3.3 percentage points in the short-term (Lavy & Schlosser, 2005). The long-term results suggest that the early positive effects persisted and treated students experienced an increase in completed years of college

[2] Similar to teachers who on average improve their teaching skills in the first few years (Hanushek, 2011).

[3] In Serbia, primary education lasts eight years. In the first four grades, pupils get one teacher who teaches all compulsory subjects, except English, and in the upper four years one teacher per subject. School is compulsory until age 15 and children first enroll when they are aged 6.5 at the beginning of the scholastic year, in September.

[4] We consider five to six years later, that is 2015, because this is the only year for which we have information on assistants. This is not problematic for two reasons. First, we can assume that the assistant in 2015 is the same person that the schools hired in either 2009 or 2010. Second, even if we cannot know whether schools had assistants in the period after 2015, all pupils in our sample had assistants in the lower four grades throughout their education, that is in the grades in which the pupils are most exposed to the program. More details on the RTA in Sections 2 and 3.

[5] There are other programs targeting disadvantaged pupils, which cannot be categorized as remedial education programs, and have been analyzed extensively in the literature. These studies include merit pay for principals, teachers, and students (Fryer, 2010; Podgursky & Springer, 2007), professional development for teachers (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009), getting parents to be more involved (Domina, 2005), placing disadvantaged students in better schools through desegregation busing (Angrist & Lang, 2004) or altering the neighborhoods in which they live (Jacob, 2004; Sanbonmatsu, Kling, Duncan, & Brooks-Gunn, 2006). Their findings suggest that there is no panacea for improving educational outcomes of disadvantaged pupils: some programs work and others do not in a specific context.

schooling, in annual earnings, in months employed, and in intergenerational income mobility, especially if coming from below median income families. In our paper we contribute to this strand of the literature providing new evidence on the medium-term effects of a remedial education program, 5 to 8 years after its implementation. Our outcomes are not limited to standardized tests, but we also examine the impact on completion of primary school and secondary school enrollment.

Furthermore, we contribute to the literature on discrepancies between teacher assessment and standardized test scores by providing evidence on how teachers raise expectations and start grading more strictly as a response to the intervention. This literature examines predominantly differences along two dimensions: first, between natives and immigrants and second, between girls and boys. Burgess & Greaves (2013) show evidence of differential grading among immigrants in England, over- and under-assessment by teachers, and they argue that such behavior is motivated by stereotypes. Conditional on standardized test scores, teachers over-grade ethnic groups that scored higher in the previous years, and under-grade ethnic groups that scored lower. Conversely, Diamond & Persson (2016) show, with Swedish data, that teachers may inflate grades in high stakes exams for students who had a "bad day", but do not discriminate on immigrant status or gender. Such discretion has long-term consequences in terms of level of education and earnings. Alongside, Calsamiglia & Loviglio (2019) provide evidence that a student in a classroom with better peers receives lower grades from the teacher than an identical student with worse peers, that is, teachers grade on a curve leading to a negative source of distortion. With respect to the differences between genders, the literature suggests that in some settings boys are under-graded while girls can be over-graded (Cornwell, Mustard, & Van Parys, 2013; Lavy, 2008; Lindahl, 2007). Lavy (2008) provides evidence that the differential grading is correlated with the teacher's characteristics, suggesting that the bias is driven by teachers' discrimination.

Finally, studies on same-race teachers are relevant for our study. While the RTA program is unique in targeting the largest ethnic minority in Europe, and there is no study on the importance of having a co-ethnic as a Roma assistant in the medium- or long-term, we conjecture that the fact that Roma are helped by same-race teachers can make an important difference. Gershenson, Hart, Hyman, Lindsay, & Papageorge (2018) show that African-American students with a same-race teacher in early grades of primary school have higher graduation rates in high school and are more likely to enroll in college. They provide evidence that this positive effect is explained by the presence in the classroom of a role model of the same background. Battaglia & Lebedinski (2017) show in the same context as this paper that when pupils are taught by someone from their community, parents' expectations on their returns to education and on secondary school achievement increase. A poor-performing student can improve her educational attainments by following the footprints of an older friend or neighbor, especially in an extremely socially and economically deprived environment.

The paper proceeds as follows. In Section 2, we describe the RTA Program, Section 3 describes the dataset, Section 4 explains the empirical strategy and Section 5 reports our results. Section 6 discusses the mechanisms at play and Section 7 concludes.

2. Roma Teaching Aassistant program in Serbia

The Roma minority arrived in Europe from India between 6th and 11th century (Kenrick & Taylor, 1998), but they have preserved their traditions and they have rarely assimilated to the majority population. In most of the countries where they reside, Roma households are poorer than Non-Roma households, making decent quality housing less affordable to them (Perić, 2012). One of the reasons for the poverty-stricken housing conditions of Roma, is their situation in the labor market: Roma are severely disadvantaged, with women facing even more difficulties than men (O'Higgins, 2012), and this is mainly caused by their low educational background. Due to their modest education, Roma often work in the informal labour market where job quality is low (Lebedinski, 2019; O'Higgins, 2012). Policies aiming at improving their educational outcomes are important to lift them out of poverty. However, disparities between Roma and Non-Roma arise already at an early age and increase over time. In Serbia, for instance, Baucal (2006) finds that in the third grade of primary education, Roma pupils lag behind their Non-Roma peers already 2.2-2.6 of school years in terms of cognitive skills. For Hungary, Kertesi & Kézdi (2011) report a raw gap of one standard deviation for reading and mathematics in the eighth grade between Roma and Non-Roma. The gap decreases when accounting for health, family background, and school and class fixed effects, indicating that a large part of it can be explained by environmental factors. However, Bhabha et al. (2018) also documents that discrimination in education can be an additional obstacle that Roma children face.

The RTA Program was first introduced in Serbian schools in 2009 and as of 2015, has covered 156 schools in the entire country. The program consists of assigning one person, the so-called Roma teaching assistant, to each participating school. The duties and responsibilities of the assistants have been included and defined in the main law regulating education in Serbia since 2009.[6] According to the law, assistants have three broad duties. First, their main responsibility is to provide educational support to children, especially in the lower four grades. If the assistant cannot reach all students, the youngest and, among them, the most disadvantaged Roma are those more likely to be helped. After-school classes are the most common mode of educational support (Milivojević, 2015). We estimate that, depending on the time of the year, assistants spend 60–80% of their time providing after-class educational support to their pupils. Second, they provide support to teachers, educators and other school employees in order to improve their work with Roma children. Third, they cooperate with parents (or guardians), the community, institutions and other relevant organizations and by doing so, they build a bridge between the school and the Roma community. The specific duties of an assistant can vary depending on the needs of the school, and there is not a compulsory amount of hours per week devoted to after-class activities nor a target amount of treated students per school.[7] We only know that a treated pupil attends at least two hours of after-school classes per week with the assistant. In the first few months on the job, some assistants work primarily in the field with parents to get to know the community and the circumstances of the children and turn after the introductory phase to standard tasks of learning support. Work with the community remains an important part of the duties of the assistants, especially because dropouts and numerous absences are challenging obstacles to Roma pupils' education.

Since the program targets a specific minority, one can be worried that it could potentially stigmatize Roma children and therefore have a negative effect on them. We take into account this possibility while discussing the results, but it is reasonable to assume that all schools which applied to the program are aware and arguably acknowledge the difficulties that Roma children face at school and are expected to promote diversity and reduce the potential stigma. This is also why the assistants are referred in the school to as *pedagogical assistants*, with no reference to the target group.

In 2009, the MoESTD developed a rulebook which defines a one-year study program that each assistant needs to attend to get a license as RTA. The assistants enroll in the program only after having been selected and

[6] Law on Fundamentals of Education System.

[7] Since 2019, so after our period of study, the number of pupils per assistants has been regulated and since then a full-time employed assistant is expected to work with 35 pupils. This number corresponds approximately to the average number of Roma students in schools with a low percentage of Roma at application. More precisely, schools below the median have, on average, 49 Roma students, while schools above the median have on average 118 Roma students. In schools in the first quartile of Roma at application, the average number of Roma students is 42.

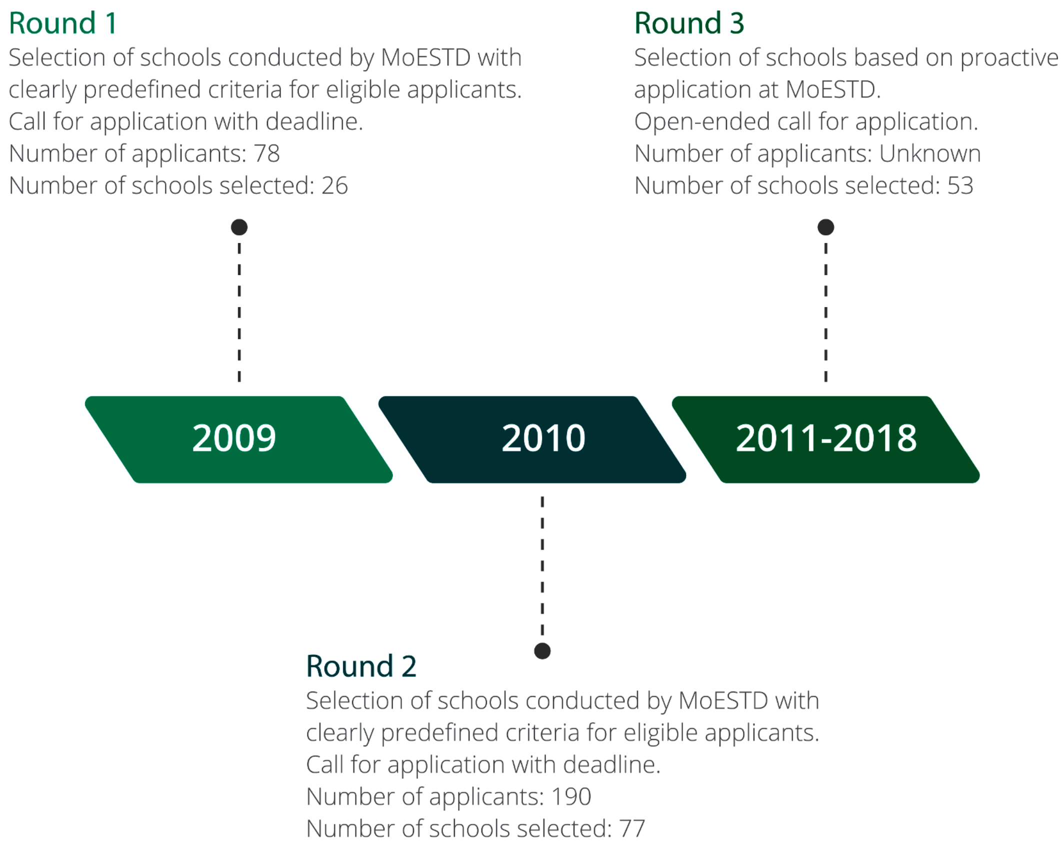

Fig. 1. Timeline.

hired by the school. The study program for RTAs aims to train pedagogical assistants for the various aspects of their work, such as preparation and realization of teaching activities, development and safety of children, work with family, cooperation with school pedagogues and school psychologists, organizations and institutions. Since the official beginning of the program, all RTAs are hired on an annual basis and their contracts are extended each year. Their salaries are fully funded by the MoESTD.

In order to get a Roma teaching assistant, schools had to apply. There were initially two rounds of applications: the first round in 2009 and the second round in 2010. In 2009, 78 schools applied for an assistant and, out of these, 26 schools were selected. The program was expanded in 2010 to another 77 schools out of 190 applicants. The main criterion for getting admitted to the program was having a share of Roma between 5% and 40%. Schools applying in the first round were also required to offer a preschool program. This condition was relaxed in the second round because in 2010 Roma teaching assistants were also assigned to preschool institutions. Starting from 2011, once a school has selected a candidate, it needs to apply for funding from MoESTD through its regional office, and it is not clear how MoESTD approves or rejects applications from schools. Presumably, the regional office makes a recommendation whether a school needs an assistant and the MoESTD approves this decision. Since then, an additional 53 schools joined the program. Similar to the schools, Roma teaching assistants had to apply and the following requirements were defined for the candidates: knowledge of the Romani language - the mother tongue of the Roma people, secondary school diploma and experience in working with children. Although it was not required for the teaching assistant to be Roma, the fact that they were expected to speak Romani implies that almost all of them belong to the Roma minority. This adds a relevant feature to the program. The teaching assistants can act as an important reference point for the community and can be seen as a role model for their students: in order to be assistants, they need to have invested in education in the first place and, thanks to such investment, they have obtained a good full-time job in the formal sector. By sharing their successful experience with students, they can motivate them to believe that they can achieve analogous results.

The number of applicants and schools selected in each round is reported in the timeline in Fig. 1. This study uses only applicants from the first two rounds, in 2009 and 2010, because in these rounds the selection criteria were universal and clearly pre-defined. The round 3 applicants and schools are not considered in the evaluation since the selection of schools is based on proactive applications and we do not know the criteria which apply.

Unfortunately, we only have information on the percentage of Roma in the entire school at application, and since we have data available only for the last grade of primary school, we cannot know the proportion of Roma enrolled at school in each year, or in each grade. This implies that formally we cannot test whether there is some selection of children into schools due to the program. Parents who decide to get their child enrolled in participating schools may differ from those who enrolled their child in this same school before the program, and from parents who send their child to not participating schools. Once the intervention is known in the community, Roma parents who care more about education may choose participating schools instead of another school in the neighborhood. Non-Roma parents may be either happy that a teacher assistant takes care of Roma children so that their child will benefit from a better learning environment and thus send her to a participating school, or may dislike the pro-Roma program, fear that many Roma children will get enrolled as a response, and decide to get their child enrolled in a different school. Nonetheless, concerns about changes in school composition are mitigated in this context. By law in Serbia, children are expected to go to the closest primary school to their home. In fact, each primary school has a catchment area and children from the catchment area are given preference when enrolling. While it is possible to enrol a child in a school in a catchment area where one does not belong, it happens only in rare cases and only if there are places available in the school after all children from the catchment area enrolled. The median aerial distance between two primary schools in Serbia is 2.37 km, suggesting that enrolling a child in a school outside of the own catchment area is also costly in terms of time. Furthermore, additional data for the initial two years of the RTA (2009 and 2010) suggest that the share of Roma enrolled in the first grade remained unchanged once the program was introduced. For this period there is no evidence of selection into treated schools.[8]

3. Dataset and descriptive statistics

3.1. Description of the dataset

We use three sources of data: (i) the MoESTD primary school final examination dataset, (ii) the MoESTD list of schools that applied for the RTA Program in 2009 and 2010, and (iii) a dataset containing the characteristics of the assistants in year 2015.

The MoESTD final examination dataset includes the whole population of pupils attending the eighth and, therefore, the last grade of primary school in a given year. The final examination is a standardized test that all pupils who are finishing eight years of primary school take at the end of the last year of school. The pupils can formally finish primary school only if they sit the test. The test consists of three parts: one in Language (Serbian), one in Math, and one in a mix of different subjects (geography, chemistry, physics, history and biology).[9] The test is nationwide and is scheduled on the same three days in all primary schools in Serbia. On each of the three days, pupils get examined in one of the three parts (Language, Math and Combined). Pupils get assigned to a secondary school depending on their total score at the final examination, their average marks from the 6th to the 8th grade and their expressed preferences regarding which school they want to enroll. Students reveal their preferences only after being informed about their final examination test score and they can list up to 20 schools. In 2011 there was a policy change as to who is required to sit the final exam. Until 2011, only pupils who wanted to enrol in a four-year secondary school track (either technical or general) were required to sit the final exam. Conversely, pupils who wanted to enroll in a three-year technical track were not asked to take the final exam. The policy change introduced in 2011 meant that one had to sit the final exam in order to receive a primary school diploma. There is no requirement in terms of achievement at the test: everyone who is present at the test receives the primary school diploma but the presence is required and everyone's test scores are recorded. As of 2011, pupils who did not want to continue education, but wanted a primary school diploma and those who continued in three-year tracks had to sit the final exam.

Our dataset includes the test results for the years 2008 to 2018 and it contains demographic information on pupils, their marks in all subjects from 6th to 8th grade, their test scores for each of the three tests and information on their secondary school enrollments. It also contains information on whether a pupil belongs to the Roma minority.[10] The MoESTD does not have digitized datasets of final examinations before 2008.

The final examination dataset is complemented with a list of schools that applied for the RTA Program in 2009 and 2010. The school application list contains information on school size and percentage of Roma in school at application and reports whether the school received the program and at which point in time. From the original list of schools that applied and were selected in 2009 and 2010, the following schools were excluded from our sample: (1) schools with less than the minimum percentage of Roma at the time of application required to participate (5%) or without the percentage recorded in our dataset,[11] (2) schools that received an assistant initially but did not have an assistant anymore in 2015 and later,[12] (3) primary schools for functional education, (4) schools without any Roma in 8th grade in the pre-treatment or treatment years,[13] and (5) schools which were assigned a zero weight in the balancing procedure.[14] As a result, schools are defined as treated if they applied either in 2009 or in 2010, got selected and still had an assistant at the time of follow-up and in the following years, and if they had a share of Roma of at least 5% at the time of application and at least one Roma in the eight grade in both pretreatment and treatments years. Schools are defined as control if they applied in any of the two years and never got selected, and if they had a share of Roma of at least 5% at the time of application and at least one Roma in the eight grade in both pretreatment and treatments years. In our final sample, there are 64 treated schools and 39 control schools, located all around Serbia. We complement the treated schools dataset with socio-economic characteristics of the Roma teaching assistants.

3.2. Descriptive statistics and balancing weights

The empirical strategy exploits the fact that, while all schools applied, some schools were admitted to the program while others were not. Not admitted schools provide the so-called comparison group. Since all schools in our sample applied to be part of the RTA, we can expect they are all equally motivated to participate and that concerns about differences in unobservables can be mitigated. Nonetheless, we have not been informed by the MoESTD why certain schools are selected and others not. At first sight, the differences in observable characteristics do not suggest obvious criteria: for instance, the difference in the percentage of Roma at application between selected and not selected schools is statistically indistinguishable, and they are equally distributed in all country districts (see more below in Fig. 2 of Section 4).[15] However, the treated and comparison schools could still differ in their pretreatment characteristics and outcomes. For this reason, we first show their differences for both Roma and Non-Roma in terms of percentage of Roma at application and outcome variables in the year prior to the introduction of the program and then, to alleviate the problem, we preprocess the data with the entropy balancing method (Hainmueller, 2012).

Our outcomes of interest are marks in Language and Math, the probability of not sitting the final examination, and standardized test

[8] In the first year of the program, that is in 2009, the number of Roma pupils enrolling at school for the first time remained the same as in the pretreatment year in both participating and not participating schools. With data from Battaglia & Lebedinski, 2015, we can show that in the pretreatment year, first grade Roma pupils in schools joining the program in 2009 corresponded to 29% of all Roma enrolled in these schools. In schools joining later, they were 26%. In the first year of the program, these percentages were 29% for participating schools and 28% for not participating schools.

[9] The last part was only introduced in 2014. This is why we cannot use it in our analysis.

[10] Since 2009, this information is used to determine who is eligible for an affirmative action policy for secondary school enrollment. The policy allows a student to get additional points for admission to secondary school based on her ethnicity.

[11] There are no schools with more than the maximum percentage of students required to be selected, 40%.

[12] Recall that we have information on assistants only for 2015. We excluded 15 schools. They are not different in observable characteristics from those in our sample.

[13] Schools without Roma in the 8th grade in the pre-treatment or treatment years are omitted because the difference-in-difference methodology requires that each school has Roma both in the pre-treatment and treatment period.

[14] We lose in total 9 schools that would have been considered as treated and 11 schools that would have been considered as control, as these schools get assigned a 0 wt.

[15] The district corresponds to an area smaller than a region and bigger than a municipality. In Serbia (excluding Kosovo), there are 4 regions, 25 districts and 145 municipalities.

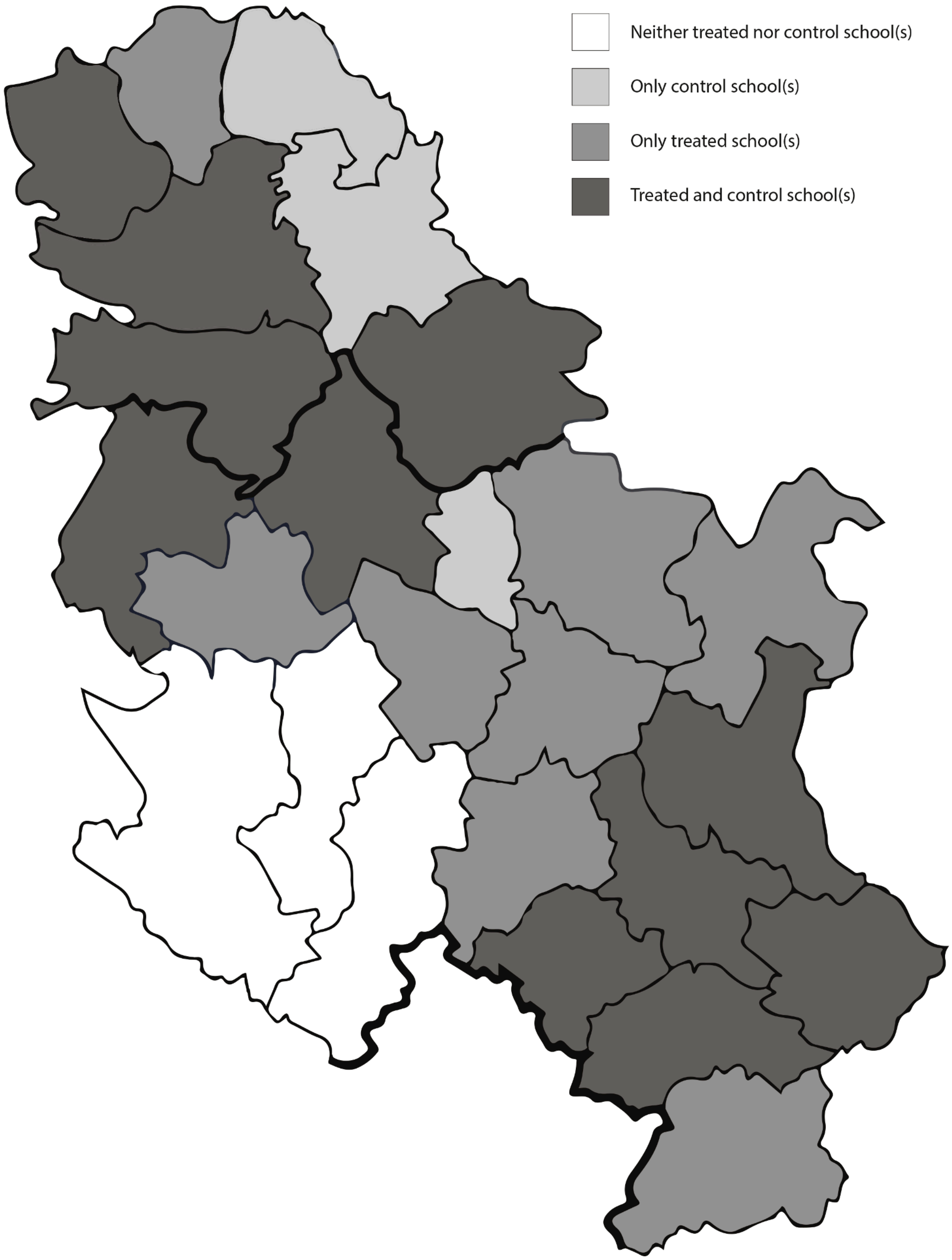

Fig. 2. Geographical distribution of treated and control schools (at the district level).

scores in Language and Math. Marks range from 1 (worst) to 5 (best). Mark 1 in a given subject is considered being insufficient, that is a non-passing grade, while mark 2 or higher are passing grades.[16] We also consider secondary school enrollment and school track. In Serbia, only primary school is compulsory and we examine whether a child continues studying after finishing the compulsory part of the education. Secondary school enrollment is a dummy equal to 1 if the student enrolls in secondary education, conditional on having completed primary education and therefore having done the test. School track choice is a dummy equal to 1 if she enrolls in a four-year track (either technical or general) and 0 if in a less demanding three-year technical track. Graduating in a four-year secondary school track is a requirement for entering higher education.

Panels A and C of Table 1 show that, overall, the means of school characteristics and outcome variables in treated and control schools before the program are fairly balanced for Roma pupils. The only statistically significant difference is observed for the enrollment in secondary education that is higher in control schools. Among Non-Roma, we find more differences in the pretreatment year, namely in terms of

[16] Students at the eighth grade who receive an insufficient mark in a subject at the end of the academic year (May) are expected to take a make-up exam for that subject in June. For those in previous grades, the make-up exams are only in August. Standardized tests take place at the end of June. This implies that in our sample we observe few students who received 1 in either Language or Math but could sit the final tests. These students passed the make-up exams. The maximum number of points at tests in Language and Math can vary by year, but we have standardized test scores to alleviate this problem and make them comparable between years.

Table 1
Balancing of school characteristics of treated and control schools at application in 2008 - Roma and Non-Roma.

	Treated		Control		Diff.	P-value
	(1) Mean	(2) Std. dev	(3) Mean	(4) Std. dev.	(5) [(1)-(3)]	(6)
Panel A: Before balancing - Roma						
Percentage of Roma in school at appl. (2008 or 2009)	0.169	0.136	0.175	0.215	-0.006	0.921
Mark in Language in the last grade	2.887	1.078	2.740	1.017	0.147	0.210
Mark in Math in the last grade	2.511	0.918	2.423	0.865	0.087	0.356
Dropout in the last grade (=1)	0.372	0.484	0.459	0.500	-0.087	0.223
Standardised testscore Language	-0.844	0.922	-0.874	0.915	0.030	0.843
Standardised testscore Math	-0.761	0.944	-0.846	0.807	0.085	0.451
Enrolled in secondary school (=1)	0.963	0.189	0.991	0.097	-0.027*	0.063
Enrolled in four-year track (=1)	0.680	0.467	0.705	0.458	-0.025	0.658
Observations	734		196			
Number of schools	73		48			
Panel B: After balancing - Roma						
Percentage of Roma in school at appl. (2008 or 2009)	0.167	0.133	0.125	0.094	0.042	0.145
Mark in Language in the last grade	2.914	1.078	2.818	1.105	0.096	0.485
Mark in Math in the last grade	2.530	0.922	2.505	0.951	0.025	0.859
Dropout in the last grade (=1)	0.353	0.478	0.345	0.477	0.009	0.882
Standardised testscore Language	-0.844	0.922	-0.747	0.904	-0.097	0.643
Standardised testscore Math	-0.761	0.944	-0.778	0.754	0.017	0.883
Enrolled in secondary school (=1)	0.963	0.189	0.988	0.109	-0.025	0.131
Enrolled in four-year track (=1)	0.680	0.467	0.744	0.438	-0.064	0.250
Observations	713		164			
Number of schools	64		39			
Panel C: Before balancing - Non-Roma						
Percentage of Roma in school at appl. (2008 or 2009)	0.130	0.083	0.112	0.071	0.018	0.134
Mark in Language in the last grade	3.723	1.176	3.674	1.173	0.049	0.436
Mark in Math in the last grade	3.377	1.219	3.345	1.226	0.032	0.620
Dropout in the last grade (=1)	0.125	0.331	0.178	0.382	-0.052**	0.034
Standardised testscore Language	-0.093	0.996	-0.156	1.002	0.063	0.474
Standardised testscore Math	-0.033	1.015	-0.154	0.975	0.120	0.241
Enrolled in secondary school (=1)	0.984	0.126	0.990	0.100	-0.006*	0.070
Enrolled in four-year track (=1)	0.872	0.334	0.870	0.336	0.002	0.894
Observations	6345		2506			
Number of schools	73		48			
Panel D: After balancing - Non-Roma						
Percentage of Roma in school at appl. (2008 or 2009)	0.128	0.082	0.109	0.057	0.019	0.148
Mark in Language in the last grade	3.738	1.177	3.672	1.174	0.066	0.412
Mark in Math in the last grade	3.379	1.221	3.403	1.225	-0.024	0.789
Dropout in the last grade (=1)	0.121	0.327	0.153	0.360	-0.032	0.221
Standardised testscore Language	-0.092	0.998	-0.115	0.999	0.023	0.829
Standardised testscore Math	-0.041	1.014	-0.090	0.969	0.049	0.687
Enrolled in secondary school (=1)	0.984	0.127	0.991	0.096	-0.007*	0.061
Enrolled in four-year track (=1)	0.874	0.331	0.844	0.363	0.031	0.198
Observations	5935		2037			
Number of schools	64		39			

Robust standard errors corrected for clustering at the school level are: * significant at 10%, ** significant at 5%, *** significant at 1%. Marks range from 1 (worst) to 5 (best). They are categorical.

the probability of not sitting the final examination and enrollment in secondary education. Both are lower in treated schools: there are fewer students who did not complete primary education, but, among those completing, there are fewer who continue studying. As mentioned before, to reduce the problem of differences in pre-treatment characteristics between treated and control groups, we pre-process the data with the entropy balancing method (Hainmueller, 2012). The entropy balancing method is grounded in the idea to reweigh each observation from the comparison group so that the reweighed data satisfy a set of specified moment conditions. In our case, we specify that the first two moments (mean and variance) of pretreatment characteristics and outcome variables of the treatment group should match these two moments of the comparison group.[17] Since in our case treatment is assigned at the school level, also balancing weights are estimated at the school level. After pre-processing the data with entropy balancing, there are no statistically significant differences between Roma in treatment and control schools in the pretreatment year (Table 1, panel B). For Non-Roma, even after balancing, treated schools have marginally fewer students enrolled in secondary education, but overall the differences in outcomes are not statistically significant (Table 1, panel D).

Once we establish that the dataset is balanced in the pretreatment year, we report the predetermined characteristics and outcomes in the pretreatment and treatment years for both Roma and Non-Roma (Tables 2 and 3, respectively). In addition to the characteristics discussed before, the table reports the following covariates we use in the analysis: gender, whether the pupil was born in a different district than the one where the school is located, age at test, and class size.

Table 2 shows that among Roma, there are no differences in neither pretreatment characteristics nor outcomes between treated and comparison schools, except for class size that is significantly higher in treated schools. On average, children in the eighth grade are 15 years old and roughly 17% of them were born in a different district than the one of the school. Their marks are relatively low, less than 3 on a scale 1 to 5, and approximately 35% of them are enrolled in the last year of primary school but do not sit the final exam (*dropout*). Once they take the exam,

[17] All variables used in the balancing are reported in Table 1.

Table 2
Roma - Means of covariates and outcomes in pre-treatment and treatment years.

	Pre-treatment year (2008)			Treatment years (2014–2018)		
	Treated Schools	Comparison Schools	Difference	Treated Schools	Comparison Schools	Difference
	(1)	(2)	(3)[(1)-(2)]	(4)	(5)	(6)[(4)-(5)]
Characteristics						
Female (=1)	0.452	0.504	-0.053	0.481	0.514	-0.033
			(0.052)			(0.036)
Pupil born in different district (=1)	0.163	0.183	-0.021	0.112	0.147	-0.034*
			(0.067)			(0.020)
Age at test	15.164	15.145	0.019	15.159	15.088	0.071
			(0.052)			(0.059)
Class size	25.749	19.146	6.603***	20.433	17.095	3.338**
			(1.621)			(1.579)
Outcomes						
Mark in Language in the last grade	2.914	2.818	0.096	2.450	2.530	-0.079
			(0.137)			(0.121)
Mark in Math in the last grade	2.530	2.505	0.025	2.209	2.240	-0.030
			(0.139)			(0.062)
Dropout in the last grade (=1)	0.353	0.345	0.009	0.118	0.102	0.015
			(0.059)			(0.034)
Standardised test score in Language	-0.844	-0.747	-0.097	-1.186	-1.280	0.094
			(0.209)			(0.117)
Standardised test score in Math	-0.761	-0.778	0.017	-1.099	-1.211	0.112
			(0.116)			(0.069)
Enrolled in secondary school (=1)	0.963	0.988	-0.024	0.852	0.861	-0.009
			(0.016)			(0.028)
Enrolled in four-year track (=1)	0.680	0.744	-0.064	0.524	0.490	0.033
			(0.056)			(0.063)
Observations	713	164		2593	665	
Number of schools	64	39		64	39	

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. Marks range from 1 (worst) to 5 (best). They are categorical.

Table 3
Non-Roma - Means of covariates and outcomes in pre-treatment and treatment years.

	Pre-treatment year (2008)			Treatment years (2014–2018)		
	Treated Schools	Comparison Schools	Difference	Treated Schools	Comparison Schools	Difference
	(1)	(2)	(3)[(1)-(2)]	(4)	(5)	(6)[(4)-(5)]
Characteristics						
Female (=1)	0.487	0.483	0.004	0.487	0.479	0.008
			(0.021)			(0.008)
Pupil born in different district (=1)	0.131	0.148	-0.017	0.114	0.127	-0.012
			(0.025)			(0.022)
Age at test	15.017	15.031	-0.013	14.900	14.892	.008
			(0.012)			(0.010)
Class size	26.201	19.839	6.362***	22.409	19.308	3.101***
			(1.361)			(0.944)
Outcomes						
Mark in Language in the last grade	3.738	3.672	0.066	3.776	3.714	0.061
			(0.080)			(0.059)
Mark in Math in the last grade	3.379	3.403	-0.024	3.383	3.351	0.032
			(0.088)			(0.066)
Dropout in the last grade (=1)	0.121	0.153	-0.032	0.009	0.009	0.000
			(0.026)			(0.002)
Standardised test score Language	-0.092	-0.115	0.023	-0.024	-0.201	0.177**
			(0.107)			(0.081)
Standardised test score Math	-0.041	-0.090	0.049	-0.053	-0.244	0.192**
			(0.122)			(0.068)
Enrolled in secondary school (=1)	0.984	0.991	-0.007	0.945	0.945	0.000
			(0.004)			(0.008)
Enrolled in four-year track (=1)	0.874	0.844	0.031	0.861	0.810	0.051***
			(0.024)			(0.016)
Observations	5935	2037		24692	8161	
Number of schools	64	39		64	39	

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%. Marks range from 1 (worst) to 5 (best). They are categorical.

Table 4
Roma children.

	Language	Math	Dropout	Standardized Test Scores: Language	Standardized Test Scores: Math	Secondary School: Enrollment	Secondary School: Four-year track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main results							
treatedpost	-0.114	0.010	-0.005	0.347	0.135	0.025	0.185***
	(0.171)	(0.154)	(0.055)	(0.238)	(0.151)	(0.030)	(0.065)
sharpened q-value	[0.974]	[0.999]	[0.999]	[0.974]	[0.789]	[0.974]	[0.044]
post	0.293*	-0.300*	-0.280***	-0.513**	-0.441***	-0.126***	-0.384***
	(0.173)	(0.160)	(0.046)	(0.244)	(0.138)	(0.027)	(0.058)
female	0.368***	0.139***	0.016	0.257***	-0.022	-0.021	0.119***
	(0.043)	(0.037)	(0.017)	(0.042)	(0.026)	(0.015)	(0.023)
Observations	4135	4135	4135	3443	3443	3443	2995
Mean in control	2.551	2.297	0.162	-1.189	-1.092	0.866	0.545
Panel B: Placebo 2009 to 2012 versus 2008							
treatedpost	-0.250	-0.062	-0.006	0.131	0.074	0.013	0.121
	(0.176)	(0.159)	(0.046)	(0.274)	(0.187)	(0.036)	(0.094)
Observations	3188	3188	3190	2075	2075	2075	1505
Mean in control	2.564	2.299	0.357	-1.166	-1.086	0.932	0.624

Notes: The table presents the treatment effects on education outcomes for Roma children. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In square brackets we report sharpened (FDR) q-values.

Table 5
Roma children: Percentage of Roma at application.

	Language	Math	Dropout	Standardized Test Scores: Test Language	Standardized Test Scores: Test Math	Secondary School: Enrollment	Secondary School: Four-year track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Lower than median percentage of Roma at application							
treatedpost	-0.144	-0.066	0.193**	0.373*	0.350**	0.038	0.240**
	(0.218)	(0.190)	(0.075)	(0.198)	(0.152)	(0.042)	(0.098)
sharpened q-value	[0.345]	[0.455]	[0.062]	[0.062]	[0.07]	[0.287]	[0.062]
post	-0.399**	-0.349**	-0.337***	-0.415**	-0.486***	-0.107**	-0.337***
	(0.189)	(0.173)	(0.062)	(0.184)	(0.140)	(0.042)	(0.085)
Observations	1402	1402	1402	1211	1211	1211	1062
Mean in control	2.571	2.284	0.134	-1.164	-1.037	0.860	0.528
Panel B: Higher than median percentage of Roma at application							
treatedpost	-0.187	0.023	-0.137*	0.333	-0.097	0.006	0.131
	(0.273)	(0.262)	(0.077)	(0.439)	(0.287)	(0.041)	(0.081)
sharpened q-value	[0.999]	[0.999]	[0.655]	[0.999]	[0.999]	[0.999]	[0.655]
post	-0.186	-0.244	-0.229***	-0.577	-0.330	-0.128***	-0.402***
	(0.279)	(0.274)	(0.069)	(0.439)	(0.262)	(0.032)	(0.072)
Observations	2733	2733	2733	2232	2232	2232	1933
Mean in control	2.527	2.313	0.195	-1.222	-1.164	0.874	0.567
Panel C: First quartile of Roma at application							
treatedpost	-0.185	-0.302*	0.194**	0.391**	0.448**	0.043	0.270**
	(0.221)	(0.181)	(0.089)	(0.190)	(0.208)	(0.042)	(0.118)
sharpened q-value	[0.156]	[0.076]	[0.076]	[0.076]	[0.076]	[0.156]	[0.076]
Observations	682	682	682	593	593	593	540
Mean in control	2.660	2.266	0.149	-1.247	-1.029	0.906	0.552

Notes: The table presents the treatment effects on education outcomes for Roma children. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In square brackets we report sharpened (FDR) q-values.

they continue with secondary education with a high certainty. As for the Non-Roma pupils, Table 3 shows that on average in both treated and comparison schools, children in the eighth grade are slightly younger than Roma and more likely to be born in the same district of the school. On average, they also have better marks than Roma and a lower probability of not sitting the final examination (around 13%). Among those completing primary education, almost everyone enrolls in secondary school, mainly in more demanding tracks. While in the treatment years there are no differences in outcomes for Roma between treated and comparison schools, we do find that Non-Roma pupils in treated schools score higher on standardised test scores in Language and Math and are more likely to enroll in 4-year tracks than pupils from comparison schools.[18]

Higher participation in the final examination together with a consequently lower enrollment in secondary school from pretreatment

[18] In Figs. A.1 and A.2 of the Appendix, we provide a graphical representation of the raw trends of the outcome variables for Roma and Non-Roma. Overall, the graphs confirm what discussed here. For Roma, the average outcomes generally follow a similar trend, although there are some outlier years. For Non-Roma we observe some divergence for standardised test scores and enrollment in 4-year secondary school tracks in the treatment years.

Table 6
Roma children: Marks and standardized test scores.

			Lower than median		First quartile	
	Language	Math	Language	Math	Language	Math
	(1)	(2)	(3)	(4)	(5)	(6)
Std. testscore Serbian	0.524***		0.543***		0.556***	
	(0.023)		(0.034)		(0.039)	
treatedpost	-0.298*	-0.027	-0.175	-0.085	-0.366	-0.553***
	(0.167)	(0.158)	(0.254)	(0.265)	(0.224)	(0.180)
post	-0.267*	-0.279**	-0.407*	-0.337	-0.366*	-0.175
	(0.149)	(0.140)	(0.220)	(0.229)	(0.212)	(0.165)
Std. testscore Math		0.475***		0.473***		0.523***
		(0.026)		(0.038)		(0.051)
Observations	3443	3443	1211	1211	593	593
Mean in control	2.689	2.407	2.645	2.340	2.725	2.313

Notes: The table presents the treatment effects on education outcomes for Roma children. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7
Non-Roma children.

				Stand. Test Scores		Secondary School	
	Language	Math	Dropout	Language	Math	Enrollment	Four-year track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main results							
treatedpost	-0.026	0.034	0.029	0.100	0.100	0.008	0.017
	(0.084)	(0.063)	(0.026)	(0.077)	(0.112)	(0.010)	(0.020)
post	-0.015	-0.085	-0.136***	-0.108*	-0.147*	-0.065***	-0.070***
	(0.090)	(0.052)	(0.023)	(0.061)	(0.088)	(0.009)	(0.018)
Observations	40,824	40,824	40,825	39,453	39,453	39,453	37,557
Mean in control	3.712	3.368	0.042	-0.169	-0.178	0.951	0.826
Panel B: Placebo 2009 to 2012 versus 2008							
treatedpost	-0.063	0.013	0.023	0.071	0.093	0.004	-0.007
	(0.092)	(0.065)	(0.018)	(0.070)	(0.112)	(0.005)	(0.014)
Observations	39,577	39,577	39,580	36,338	36,338	36,338	28,172
Mean in control	3.716	3.377	0.103	-0.151	-0.161	0.978	0.868
Panel C: Lower than median Roma at application							
treatedpost	0.079	0.010	0.016	0.138	0.216	0.002	0.048**
	(0.072)	(0.084)	(0.034)	(0.095)	(0.138)	(0.015)	(0.021)
post	-0.102*	-0.041	-0.116***	-0.075	-0.162	-0.064***	-0.089***
	(0.055)	(0.055)	(0.032)	(0.063)	(0.108)	(0.013)	(0.017)
Observations	21,822	21,822	21,823	21,149	21,149	21,149	20,044
Mean in control	3.714	3.342	0.037	-0.126	-0.119	0.949	0.833
Panel D: Higher than median Roma at application							
treatedpost	-0.161	0.093	0.057	0.074	-0.018	0.015	-0.025
	(0.170)	(0.104)	(0.036)	(0.133)	(0.188)	(0.015)	(0.036)
post	0.111	-0.157	-0.168***	-0.157	-0.122	-0.067***	-0.039
	(0.182)	(0.099)	(0.030)	(0.123)	(0.156)	(0.011)	(0.035)
Observations	19,002	19,002	19,002	18,304	18,304	18,304	17,513
Mean in control	3.704	3.444	0.058	-0.296	-0.356	0.959	0.805
Panel E: First quartile of Roma at application							
treatedpost	0.164***	0.010	0.029*	0.251***	0.361***	-0.013	0.075***
	(0.060)	(0.064)	(0.017)	(0.055)	(0.054)	(0.009)	(0.020)
Observations	11,988	11,988	11,989	11,580	11,580	11,580	10,991
Mean in control	3.712	3.242	0.043	-0.197	-0.124	0.953	0.818

Notes: The table presents the treatment effects on education outcomes for Non-Roma children. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

year to treatment years are explained by a regulatory change introduced in 2011. Note that this pattern is observed for both Roma and Non-Roma, in treated and control schools. Prior to 2011, pupils with passing grades who wanted to enroll in three-year tracks in secondary schools were not required to sit the final examination and would get a primary school diploma independently of attending the final exam. Since 2011, only pupils who sit the final exam, regardless of the number of points at the exam, receive a primary school diploma. As a result, also pupils who wanted a primary school diploma without further continuing education or those wanting to enroll in three-year tracks were required to sit the final exam. Due to this change in requirements, the number of pupils who sit the final exam increased largely in treatment years (evident from the fall in the outcome variable *Dropout (=1)*), but the share of pupils enrolled in secondary school, conditional on sitting the final exam, decreased (evident from the outcome variable *Enrolled in secondary school (=1)*). The reform incentivized more pupils to sit the exam in order to get a primary school diploma, but now, conditional on sitting the exam, a larger share of them do not enroll in secondary education. The unconditional values show that there are no differences in the percentage of students in the eight grade of compulsory education that keep studying, before and after the regulatory change. Among eight grade Roma, there were 76% of students before and 74% after who went

Table 8
Characteristics of the assistant.

	Secondary School															
	Enrollment								Four-year Track							
	Assistant		University		Tenure		Age		Assistant		University		Tenure		Age	
	Male	Female	Without	With	Longer	Shorter	Higher	Lower	Male	Female	Without	With	Longer	Shorter	Higher	Lower
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
treatedpost	0.019	0.038	0.019	0.053	0.021	0.041	0.055	0.018	0.152*	0.216***	0.171**	0.223***	0.140*	0.226***	0.092	0.235***
	(0.042)	(0.028)	(0.035)	(0.033)	(0.036)	(0.033)	(0.042)	(0.030)	(0.077)	(0.068)	(0.071)	(0.073)	(0.077)	(0.069)	(0.077)	(0.068)
post	-0.110***	-0.145***	-0.125***	-0.132***	-0.120***	-0.134***	-0.122***	-0.130***	-0.386***	-0.383***	-0.386***	-0.384***	-0.400***	-0.372***	-0.415***	-0.362***
	(0.029)	(0.031)	(0.028)	(0.034)	(0.030)	(0.029)	(0.030)	(0.029)	(0.060)	(0.060)	(0.059)	(0.061)	(0.059)	(0.059)	(0.059)	(0.059)
Observations	2248	2039	2892	1350	2192	2095	1770	2517	1925	1807	2494	1196	1895	1837	1538	2194
Mean in control	0.866	0.866	0.866	0.866	0.866	0.866	0.866	0.866	0.545	0.545	0.545	0.545	0.545	0.545	0.545	0.545
Placebo 2009 to 2012 versus 2008																
treatedpost	0.005	0.032	0.006	0.047	-0.006	0.039	0.009	0.031	0.136	0.094	0.114	0.124	0.113	0.123	0.057	0.157
	(0.042)	(0.040)	(0.038)	(0.046)	(0.041)	(0.041)	(0.045)	(0.037)	(0.112)	(0.104)	(0.104)	(0.106)	(0.101)	(0.114)	(0.107)	(0.106)
Observations	1347	1202	1732	784	1346	1203	1017	1532	964	891	1240	590	959	896	739	1116
Mean in control	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.624	0.624	0.624	0.624	0.624	0.624	0.624	0.624

Notes: The table presents the treatment effects on education outcomes for Roma children depending on the characteristics of the assistant. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to secondary education, of which 52% chose a four-year track. Among Non-Roma, about 94% of eight grade students went to secondary education both before and after, choosing mainly more demanding tracks (78% before and 84% after).

The last relevant piece of information we have are the characteristics of the assistant in the 64 treated schools. We know their gender, age, whether they hold a university degree, and their work experience in the treated school, in the year 2015. Slightly less than half of the assistants are female (49%), and they are on average 36 years old - with the youngest being 21 and the oldest 51. They have, on average, 4.3 years of experience in the school where they are currently working, that is, almost all of them were hired when the program was first implemented. Unusual for a minority with really low education levels, 23% of them received a university degree. Among them, 70% are women.

4. Empirical strategy

We estimate the impact of the RTA Program using a difference-in-difference methodology with school fixed effects:

$$\begin{aligned} outcome_{ist} &= \beta_0 + \beta_1 treated_{st} {*} post_{st} + \beta_2 post_{st} + \\ &+ \rho_1 X_{ist} + \mu_s + \gamma_t + \epsilon_{ist} \end{aligned} \tag{1}$$

where $outcome_{ist}$ stands for the outcomes of pupil i in school s at time t, namely last grade marks in Language (Serbian) and Math, whether the child does not sit the final examination and thus does not finish primary school (*Dropouts*) and standardized test scores in Language (Serbian) and Math. We also investigate secondary school choices. More precisely, we are interested in the effect of the program on (i) the probability of enrolling in the secondary school and on (ii) choosing a four-year secondary school track (either general or technical track) versus a three-year technical track, conditional on enrolling. Our coefficient of interest for the overall impact of the program is β_1. We consider it as the medium-term impact: when we evaluate its effects, treated pupils were first exposed to the program 5 to 8 years before.[19] $post_{st}$ is equal to 1 for the years 2014 to 2018 and 0 for the year 2008. Year 2008 is used as a baseline and we examine the impact for 2014 and the subsequent years because the generation finishing school in 2014 was the first to be treated. Assistants mainly work with pupils in the first four grades. Consequently, students enrolled in the fourth grade when the RTA was first implemented, in 2009, should be in the eighth grade in 2014. This is the first treated generation. We run the main analysis for Roma and Non-Roma students separately. Since the program is targeting only the former, it is plausible to expect effects mainly on this subgroup of the entire population.[20] The set of exogenous individual characteristics is expressed by X_{ist}. The control variables, reported in Tables 2 and 3, include: a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. μ_s corresponds to school fixed effects and γ_t to survey-year fixed effects. Robust standard errors are clustered at the school level.

[19] We know that at the beginning of the program, assistants worked predominantly with lower four grades. While we cannot exclude that they followed some pupils also in the upper four grades of primary school, it is very unlikely that assistants can provide much support to the pupils in the upper grades where the subjects become more complex (physics, chemistry, biology, etc.).

[20] We also investigate the impacts for the full sample of students and whether there are differential effects between treated Roma and Non-Roma. To capture the condition of being a pupil of Roma ethnicity and being in a treated school s, we estimate the effects as follows: $outcome_{ist} = \beta_0 + \beta_1 treated_{st}{*}post_{st} + \beta_2 post_{st} + \delta_1 roma_{is} + \delta_2 treated_{st}{*}roma_{is} + \delta_3 post_{st}{*}roma_{is} + \delta_4 treated_{st}{*}post_{st}{*}roma_{is} + \rho_1 X_{ist} + \mu_s + \gamma_t + \epsilon_{ist}$. δ_4 is our main coefficient of interest. X_{ist} is the set of exogenous individual characteristics, μ_s are school fixed effects and γ_t are survey-year fixed effects. The results are reported in Section A.3 in the Appendix.

To capture the heterogeneity due to the intensity of the program, we run the previous specification for schools with less or more Roma than the median percentage of Roma at school at application and by quartiles. Recall that there is only one assistant per school. If there are fewer Roma students in a school, they have a higher probability of being treated. At the same time, the more Roma students in a school, the higher can be the spillover effects.

Our data face two important limitations: first, we observe pupils only at the end of primary school, and we cannot follow them over the years. In the period that we study, schools kept their records on educational attainments of pupils exclusively on paper and this data is not available in digital format. Consequently, the MoESTD does not have individual level records of primary school pupils by grade. Second, we cannot formally test the parallel trend assumption because the MoESTD does not have digitized datasets of final examinations before 2008.

As a consequence of the first limitation, in the absence of panel data, we are not able to control for differential attrition between control and treatment schools. We cannot exclude the possibility that more fragile Roma are, as a result of the program, more likely to stay in treatment schools than in control schools until the end of compulsory schooling. However, if this were the case, then we would underestimate the impacts of the program and our estimates should be interpreted as lower bounds.

Second, the difference-in-difference approach relies on the parallel trends assumption: in the absence of the program, treatment and control schools would have had a parallel trend in the outcomes of interest. Since there is available information only for one pre-treatment year, that is 2008, it is difficult to formally test the assumption. We propose two complementary exercises to provide confidence in the hypothesis. As shown in the previous section, we first use the entropy balancing method. This allows to reduce possible differences in baseline levels, and not just in trends, of the treated and control groups. Ryan, Kontopantelis, Linden, & Burgess (2018) illustrate, via a Monte Carlo simulation, that matched difference-in-difference does well at dealing with non-parallel trends in a context of health policy interventions. Moreover, in the pre-treatment period, there were no other policies that affected differently treated and control schools and that might be a threat to the parallel trends assumption. Second, we run a placebo test by comparing treated and control schools in the first three treated years versus treated and control schools before the introduction of the program. We include in the placebo test years from 2009 to 2012, that is the years immediately after the introduction of the program, because when the program started, assistants were working with lower grades and we can assume that students in grades higher than four were not treated. Students who are in the eighth grade from 2009 to 2012 were unlikely targeted by the assistants: when the program started they were already in the fifth, sixth, seventh or eighth grade of primary education. Significant coefficients in placebo regressions would invalidate this estimation strategy and would question the adequacy of our comparison group.

We now turn to a discussion of the Serbian context and whether other educational reforms can be relevant for the robustness of the difference-in-difference methodology. In the period considered in this study, compulsory preschool education was extended to 9 months. This change happened at the same time in all schools and there is no reason to believe that treated and control school were affected differentially.[21] We are therefore confident that this policy could not invalidate our methodology. Additionally, there was a regulatory change in the period of this study. From 2011 onwards, sitting the final exam became a requirement for attaining a primary school diploma. Prior to 2011, the final exam was only attended by pupils who wanted to enroll in four-year secondary education. This regulatory change could affect the composition of the pupils in our sample if it induces a different selection into test taking between treatment and control schools. In our sample, after balancing, the pre-treatment characteristics of the treated and control schools do not show significant differences (Table 1), supporting our argument that these two groups of schools are comparable.[22] Moreover, both policy changes were implemented in the whole country in a specific year and we trust that treated and control schools were affected by these two reforms and any other policy change in the same way for two main reasons. First, Serbia is a highly centralized country and educational policy is governed and commonly implemented at once in the whole country by the MoESTD. Second, as shown in Fig. 2, treated and control schools are equally distributed in the different 25 districts and they are not clustered in certain areas. It is not possible that some schools were affected by some reforms, while others were not. It is therefore highly unlikely that there is any policy unrelated to the RTA program, which targeted only treated schools and could lead us to confound the results.

We are finally interested in the characteristics of the assistants and whether they affect differently their pupils. We look at gender, age, whether they hold a university degree or not, and their working experience (longer or shorter than the median years of tenure). We concentrate on Roma children.

5. Results

5.1. Impact of RTA on Roma children

Since the program was targeting Roma children, we first analyze the RTA impacts on them. Columns (1) and (2) of Table 4 report the program effects on marks in Language and Math, column (3) on the probability of not sitting the final examination, and columns (4) and (5) report the standardized test scores in Language and Math. Bear in mind that not all students enrolled in the last academic year take the exam to finish primary school. Almost 17% of Roma children attending the eighth grade do not sit the final examination and have no test scores, as the reduction in the number of observations from columns (3) to (4) suggests. Columns (6) and (7) report the probability of enrolling in secondary school and, conditional on continuing the studies, of choosing a longer track. In Panel B, we show the coefficients of the placebo regressions. All regressions control for students' characteristics and year and school fixed effects.

Overall, the RTA program does not seem to be effective in improving neither Roma pupils' schooling marks, nor their standardized test scores. Roma in treated schools are also no more likely to keep studying than their counterparts in control schools due to the program. Nonetheless, if they go to secondary school, they are 18.5 percentage points more likely to choose a four-year secondary school track (either general or technical) versus a three-year technical track. In control schools, 54.5% of Roma students enrolled in secondary education choose a four-year secondary school track. In treated schools, thanks to the RTA, 73% chose a longer track.[23] [24] The coefficients of the placebo regressions

[21] A six months program of compulsory preschool education was introduced in 2006/2007 and the program was extended to 9 months in 2009/2010. The generation entering primary school in 2007 was the first to attend 6 months of preschool, while the generation starting primary school in 2010 was the first to attend 9 months of a preschool program in 2009/2010.

[22] We also need to assume that the selection effect is linearly additive, as specified in standard selection models (see for instance Heckman (1979)).

[23] The values unconditional of having sat the final examination suggest the same pattern. They are not reported, but they are available upon request. Roma who reach the eighth grade of primary education in treated schools are not more likely to keep studying but they are almost 20 percentage points more likely to choose a more demanding track. In control schools, 76% of Roma students enrolled in secondary school, and, conditional on enrolling, 52% choose a four-year track. In treated schools, this share is almost 72%.

[24] All the results are confirmed if we test for multiple hypothesis. We use sharpened False Discovery Rate (FDR) q-values that provide the expected proportion of rejections that are type I errors (false rejections).

reported in Panel B are all not statistically significant, therefore validating the comparability of our control group.[25] [26]

5.1.1. Intensity of the program

The percentage of Roma in school at application is a crucial predetermined characteristic to understand the program impact. A previous study of the short-term effects has shown that children in schools with fewer Roma benefit more from the program (Battaglia & Lebedinski, 2015). There is only one assistant per school and the higher is the number of Roma pupils she needs to interact with, the lower will be the intensity of the program. If the assistant cannot reach all Roma students, she is instructed to focus on the first four grades, and among them, the most disadvantaged Roma are those more likely to be helped. In Table 5 we report the impact of the RTA on our outcomes of interest by percentage of Roma at application: Panel A documents the effects for the schools with a percentage of Roma at application lower than the median (11%), Panel B for schools with a percentage of Roma higher than the median, and Panel C for schools in the lower quartile of the percentage of Roma at application (lower than 7.1% of Roma students).

Among schools with a below median percentage of Roma at application, we find higher test scores in both Language and Math for treated students. In treated schools, Roma students receive 0.373 of a standard deviation more in Language and 0.350 of a standard deviation more in Math than their counterparts in control schools. Surprisingly, we also observe an increase in students enrolled in the last grade who do not sit the final exam. Although we cannot exclude alternative explanations because of the lack of data, it is possible that more fragile Roma are, because of the program, more likely to stay in treatment schools. They enroll in the last grade of compulsory education, but they do not have any interest in continuing studying and, therefore, in obtaining the diploma. The RTA keeps the most fragile Roma students in school, but it is not enough to nudge them into enrolling in secondary school. Finally, the last column of Table 5 suggests that the overall impact on the four-year track observed in Table 4 is driven by Roma students in treated schools with less Roma.[27] In schools with a higher percentage of Roma at application, effects on standardized test scores smooth away: their coefficients are not anymore statistically significant. The higher is the number of Roma, the lower is the intensity of the program, since there is only one assistant per school, and the fewer its impacts. We acknowledge that schools with few or many Roma pupils could differ in other ways that can also interact with the program and impact its success. These differences, however, are not invalidating our findings that the program is more successful in schools with a smaller percentage of Roma.

Among schools with fewer Roma at application, we next focus on the first quartile. The recent regulation of 2019 expects the assistant to work with 35 students. In schools in the first quartile of Roma at application, the average number of Roma students is 42. This subset of schools is therefore the one where the RTA could be expected to be more effective.

Panel C shows that indeed these are the schools where most of the effects are observed. Although the remedial program does not seem to improve Roma pupils' standardized test scores on average, it clearly does it among schools with a percentage of Roma at application below the median, but especially in the first quartile. In this subset of schools, we find higher test scores for treated students both in Language and Math: in treated schools, Roma students receive 0.391 of a standard deviation more in the Language test and 0.448 of a standard deviation more in the Math test than their counterparts in control schools. Roma children are also 27 percentage points more likely to choose a four-year track. Nonetheless, higher test scores are combined with lower marks in the corresponding subjects: in treated schools, Roma students receive 0.302 less of a point in Math.[28]

This puzzling evidence could suggest that in treated schools, especially with fewer Roma at application, teachers mark more rigorously Roma pupils due to the program. In other words, we conjecture that teachers now have higher expectations of Roma pupils and they are less likely to over-grade them.[29] We test our hypothesis by estimating the following regression:

$$\begin{aligned} subjective_{ist} &= \beta_0 + \beta_1 treated_{st} * post_{st} + \beta_2 post_{st} + \\ &\quad + \delta_1 objective_{ist} + \rho_1 X_{ist} + \mu_s + \gamma_t + \epsilon_{ist} \end{aligned} \tag{2}$$

where *subjective*$_{ist}$ stands for the marks in Language and Math of pupil i in school s at time t and *objective*$_{ist}$ for the corresponding test scores. We consider marks to be subjective evaluations because they possibly incorporate teachers' discriminative behavior as they are non-blind assessments. On the contrary, standardized tests are assessed by external teachers and are not expected to suffer from any bias. The remaining variables are as in specification (1). If the coefficient β_1 for the treatment effects is still significant once we control for the standardized test score in the same subject (*objective*$_{ist}$), then there is likely some discriminating behavior at play. A positive (negative) coefficient of β_1 would suggest that teachers in treated schools over-grade (down-grade) Roma students due to the program. Put differently, we compare pupils' marks in treated and control schools, conditional on having the same test score: if teachers' evaluations were to be fully objective, there should not be differences in the way students are graded in the two types of schools. There are two main reasons why teachers' assessment could differ from objective test scores. First, teachers' assessments take into account non-cognitive skills such as behavior, which is not accounted for by standardized test scores (Cornwell et al., 2013). Second, in certain contexts, teachers grade differently ethnic minorities: depending on their expectations, they can either over-grade or under-grade them (Burgess & Greaves, 2013). There is much anecdotal evidence suggesting that teachers over-grade Roma pupils due to lower expectations and that they see keeping them at school as the main challenge and goal. The presence of differential grading in Serbia is empirically confirmed by a study which shows that in the third grade of primary school Roma pupils take a lower test score than Non-Roma pupils conditional on teacher marks (Baucal, 2006).

Our results suggest that, holding test scores constant in the two types of schools, Roma students in treated schools are graded lower due to the program than their counterparts in control schools (columns (1) and (2)), especially in schools in the first quartile of the percentage of Roma at application (columns (5) and (6)). In these schools, among students with the same standardized test score in Math, those in treated schools are marked 0.553 points less in the subject than those in control

[25] The positive and statistically significant coefficients of Table 4 for *female* suggests that we might observe different impacts depending on students' gender: girls perform better in any outcome of interest. We therefore look at the RTA's effects on boys and girls: Table A.1 in the Appendix report them separately, by row. The effects on marks, dropouts and test scores do not mask heterogeneity by gender: there are not statistically significant results. Nonetheless, girls are more likely to choose more demanding tracks. The overall result we observe in the main table is driven by them.

[26] In an alternative specification, we examined whether there are differential treatment effects depending on the number of years a child was treated during the lower four grades of primary school. Children in our sample have been treated from 1 to 4 years in lower grades. Overall, we find weak evidence of differential effects driven by the number of years treated, although the coefficients seem to suggest that only one year of treatment may not be enough to observe important changes. Results are available upon request.

[27] The results are confirmed if we test for multiple hypothesis.

[28] The results are confirmed if we test for multiple hypothesis.

[29] In Section A.4 of the Appendix we provide a graphical illustration of teachers' assessment and standardized test scores before and after program to further motivate our hypothesis.

schools.[30]

We speculate that the RTA program affected teachers in such a way that they over-grade fewer Roma students at the margin. Once the program is implemented, there is someone taking care of them and teachers may now raise their expectations and demand the same learning achievements as from Non-Roma pupils. The previous positive discrimination is now replaced by a more impartial behavior.[31] Students at the margin of passing a course are less likely to be over-graded due to the program, especially when the program is more intensively implemented.

5.2. Spillover effects: Impact of RTA on Non-Roma children

We then turn to Non-Roma children. The assistants do not work with them, and none of these students are directly treated by the RTA. Nonetheless, Roma and Non-Roma children are attending the same classes, interacting with each other in the school, and we can think of several channels for spillover effects in this setting. First, the program changes the quality of the peers of Non-Roma pupils and this could have direct effects. Then, the program may also free up teachers' time so they can work more with the low-performing Non-Roma students and it can affect teachers' behavior, e.g. evaluations of both Roma and Non-Roma pupils in their classes.

Table 7 reports the impacts for the entire sample of Non-Roma (Panel A), placebo effects for the eight graders not affected by the policy (Panel B), and the effects for schools with a percentage of Roma at application lower (higher) than the median (Panels C and D). Panel E reports the effects in the first quartile of Roma at application. The percentage of Roma at application could be as relevant for Non-Roma as for Roma children: the more Roma pupils at school, the weaker the impact of the program, the smaller the spillover effects for Non-Roma children and/or the time allocated to untreated low-performing students. We report the impact of the RTA on the marks in Language and Math (columns (1) and (2)), on the probability that students do not take the standardized exam (column (3)), on the standardized test scores in Language and Math (columns (4) and (5)), and on the probability of enrolling in secondary school and, conditional on continuing the studies, of choosing a longer track (columns (6) and (7)).

Overall, the results are not statistically significant: Non-Roma children neither perform better, nor are more likely to enroll in secondary education and choose a four-year track than their counterparts in control schools. The latter results are not particularly surprising since, on average, in control schools, almost 95% of Non-Roma continue to study in secondary education and 82% of them choose a four-year secondary school track.[32] The coefficients of the placebo regressions are not statistically significant, therefore validating the comparability of our control group. Largely, the program seems not to have had significant impacts on Non-Roma students.[33] Nonetheless, if we turn to the intensity of the program, we observe that being in a school with a low percentage of Roma students at application has a positive effect also for Non-Roma pupils, especially on the probability of choosing a more demanding track. As reported in Panel C of Table 7, students in treated schools are 4.8 percentage points more likely to choose a four-year track in schools with few Roma, reaching the 88% of students in treated schools. The program effects are even bigger in the first quartile of Roma at application: there, Non-Roma students perform 0.251 and 0.361 of a standard deviation more in the Language and Math test, and they are 7.5 percentage points more likely to choose longer tracks. The effects nullify in schools with a high percentage of Roma at application (Panel D).

As before, we test whether there is some evidence of over-grading induced by the program.[34] Table A.3 in the Appendix shows that teachers are more consistent and hence objective when evaluating their Non-Roma students. Coefficients for treatment in columns (1) and (2) are modest and much smaller than in the case of Roma and are not statistically significant. In the first quartile, we observe a similar behavior as with Roma. Nonetheless, it is worth noticing that the coefficient in column (6) is half the one recorded in Table 6, and that in these treated schools Non-Roma students are more likely to have higher - and not lower - marks (Panel E of Table 7).

Overall, we can conclude that in the medium-term there are significant effects of the program also on the not targeted group, especially in those schools where its intensity is higher, that is, where there are fewer Roma students per assistants. It can either be the result of positive peer learning spillover effects of the program on Non-Roma students or of more time now devoted by teachers to those among them who under perform.

5.3. Characteristics of the assistants

We are finally interested in understanding if the characteristics of the assistant can be important for our outcomes of interest. We would like to see which traits of the assistant correlate with better educational achievements of the pupils. We know their gender (49% are women), their age (36 years old, on average), whether they hold a university degree (23%), and their work experience in the school, in the year 2015 (4.3 years, on average).

Since overall we observe statistically significant impacts only on the probability of choosing a more demanding track, we concentrate the following analysis on secondary school choices.[35] In Table 8 we report the coefficients for enrollment (columns (1) to (8)), and for the probability of choosing a longer track (columns (9) to (16)).

Overall, teachers' characteristics seem to be weakly relevant, except for secondary school decisions. The probability of continuing to a four-year secondary school track is higher in schools with a female assistant compared to schools with a male assistant.[36] In both cases, students in treated schools are more likely to choose more difficult tracks, compared to students in control schools, but the effect is higher when the assistant is a woman. Also, having an assistant with a university degree increases

[30] Recall that students at the eight grade who receive an insufficient mark in a subject at the end of the academic year (May) are expected to take a make-up exam for that subject in June. The coefficient for treatment in column (6) suggests that on average students receiving the same test score, if in treated schools, were receiving 1.8 points (the mean in the control group is 2.313), that is, they were expected to take a make-up exam and passed it.

[31] The last two columns of the Table A.1 in the Appendix report the results of our test in specification (2) by gender. They show that, among boys with the same standardized test score in Language, those in treated schools are marked 0.358 points less in the subject than those in control schools: 2.25 instead than 2.6.

[32] The values unconditional of having sat the final examination suggest the same pattern. They are not reported, but they are available upon request. There are not statistically significant differences. In control schools, on average almost 94% of students go to secondary school and 80% of them choose a four-year track. These data are totally comparable with the whole population of students in Serbia in the period 2014 to 2018. Among pupils enrolled in the 8th grade of primary school, 98.47% sit the final exam. Similarly, among those enrolled in the 8th grade, 92.89% continue to secondary school and 75.91% choose four-year tracks (authors' calculation from MoESTD)).

[33] Table A.2 in the Appendix reports the effects for Non-Roma boys and girls separately, by row. The effects of RTA are not heterogeneous by gender, except for the probability of enrolling in secondary education. Boys are more likely to keep studying. The last two columns of the table report the results of our test in specification (2). All coefficients are small and not statistically significant.

[34] The distribution of marks and standardized test scores is provided in Section A.4 of the Appendix.

[35] Table A.4 in the Appendix reports the results for the other outcomes of interest.

[36] The p-value of the difference between the coefficients in columns (9) and (10) is 0.0000.

more the probability of choosing a longer track.[37] Finally, years spent in the same school and age are also important: younger and therefore, plausibly, less experienced assistants matter more for the choice of the more demanding track.[38] Moreover, if we concentrate our attention on the assistant's gender, we can further investigate its interaction with the student's gender. As reported in Table A.5 in the Appendix, girls' test scores, their enrollment and the probability of choosing a more demanding track increase when their assistant is a woman. Similarly, we find a positive impact on test scores for boys when the assistant is a man.[39] There is some evidence of the same-gender effect already observed in the literature.

6. Discussion of the mechanism of the program

Overall, we observe that the remedial program does not have transformative effects on Roma pupils schooling attainments in the medium-term, nor in terms of marks neither of standardized test scores, except for students in schools with a low percentage of Roma at application. Roma in treated schools are also no more likely to keep studying than their counterparts in control schools because of the program. Nonetheless, if they go to secondary school, they are more likely to choose a more demanding track.

This result is particularly important if we consider that choosing a four-year track allows to enter university, and that reaching this level of education is extremely rare for this minority. The remedial activities have not shown to change Roma students' average attainments, and the gap with Non-Roma pupils remains big. Nonetheless, we observe some unexpected behavior in grading and an improvement at the margin. Roma students get lower marks but higher test scores in Math in schools where the program is more intensively implemented. They are unlikely to be over-graded, but they learn more and perform better once externally tested. We speculate the RTA program affected teachers in such a way that they over-grade fewer Roma students at the margin of passing the course. With the program, there is someone taking care of them, and teachers may now raise their expectations and demand the same learning achievements as from Non-Roma pupils. We cannot exclude that lower marks result from higher stigmatization of Roma pupils in treated schools. Since the program targets a specific minority, one can be worried that it might stigmatize them with consequent negative effects. Yet, such explanations would not explain positive effects on standardized test score nor on the likelihood of choosing a more demanding track in secondary education.

Such more impartial behavior on the teacher side, if there, is however not enough to explain our overall results. Roma people usually attain very low education and their upper secondary school completion rates are much lower than Non-Roma. There is a clear low investment in education among them because of financial constraints, barriers of access to education, or low expectations for schooling to give them enough future opportunities. Having someone from their community who had to invest in education in the first place and, thanks to such an investment, has got a good full-time job in the formal sector may encourage them to follow a similar path. The teaching assistant is an important reference point for the community and can act as a role model. By sharing her successful experience with students, she can motivate them to believe that they can achieve analogous results. Graduating in a four-year secondary school track can be the first step, since it is a requirement for entering higher education. The evidence on the characteristics of the assistant again points at the importance of a role model: pupils, whose assistants are educated, young, women are more likely to choose a more demanding track. This is the case especially for girls, who may identify with their female assistants.

Non-Roma students also benefit from the program: in the medium-term there are significant effects on this not targeted group, especially in those schools where its intensity is higher, that is, where there are fewer Roma students per assistants. It can either be the result of a positive change in the quality of their peers that had positive learning spillovers on Non-Roma students or they can benefit from more time now devoted by teachers to those among them who underperform. We cannot disentangle which of the two possible channels is more important or if both channels play an equally important role. However, the finding that Non-Roma pupils benefit from the program is per se relevant.

7. Conclusions

The Roma Teaching Assistant program is one of the main program in Eastern Europe targeting the Roma population and aiming at increasing their educational attainment. We studied its impacts on children's schooling outcomes 5 to 8 years after its first implementation. Our study complements the results obtained in the short-term by Battaglia & Lebedinski, 2015 which have shown that the program improved school attendance and that younger children benefited more from it.

Roma ethnic minority has extremely low educational attainment and high poverty levels. A necessary condition to ensure a better life for the young generation of Roma is to provide them with an education that can help them rise the economic ladder. Remedial education programs as the RTA can help disadvantaged students escape the vicious cycle of poverty and attain better jobs than the generation of their parents. Since the assistant belongs to the same minority, she can act as an important reference for her community and as a role model for her students. Investing in education to achieve better future outcomes can appear now feasible and relevant. Our results suggest that, in the medium-term, more Roma students enroll in higher demanding tracks thanks to such program and that there can be some marginal improvements in learning, depending on the intensity of the program. Positive impacts are clearly observed when intensity is higher, both for the targeted group and, thanks to spillover effects, for their schoolmates. We also observe that pupils with young, educated, female assistants are more likely to choose more demanding tracks, especially if girls, reinforcing therefore the importance of the role model mechanism.

The results of the medium-term evaluation are aligned with the previous findings from the study on short-term effects (Battaglia & Lebedinski, 2015). It is confirmed that the number of Roma pupils that an assistant works with is essential for the success of the program. The program works only when the number of pupils per assistant is not higher than roughly 40. We thus conclude that the overall effects of RTA are quite modest and limited to schools with a small percentage of Roma, and the program is not achieving its aim of raising educational attainments of all Roma pupils. Our results show that this program is insufficient to bridge the gap between Roma and Non-Roma pupils and that additional measures are required to achieve the goal of integration for this disadvantaged minority.

Credit Author Statement

Both authors equally contributed to the manuscript.

Acknowledgements

We are grateful to Marisa Hidalgo-Hidalgo, Marko Vladisavljević and Sunčica Vujić for helpful suggestions. We benefited from valuable comments of seminar participants at 2019 SAEe Meeting, at the Universitat de les Illes Balears, and at the Graduate Institute Geneva. Financial support from Ministerio de Economía y Competitividad (ECO2017-83069-P) and from the Serbian Ministry of Education, Science and Technological Development is gratefully acknowledged. All opinions expressed are of the authors, all errors are our own.

[37] The p-value of the difference between the coefficients in columns (11) and (12) is 0.0000.

[38] The p-value of the difference between the coefficients in columns (13) and (14) is 0.0000.

[39] Girls' test scores increase by 0.390 of a standard deviation in Language up to 0.547 in Maths when their assistant is a woman, while boys' test scores increases in Math by 0.418 of a standard deviation when their assistant is a man.

Appendix A

A1. Additional tables

Table A.1
Roma children by gender.

	Language	Math	Dropout	Stand. Test Scores Language	Stand. Test Scores Math	Secondary School Enrollment	Secondary School Four-year track	Conditional on St. Test Score Language	Conditional on St. Test Score Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
treatedpost (male)	-0113	-0.063	-0.025	0.297	0.127	0.015	0.142	-0.358**	-0.183
	(0.148)	(0.096)	(0.078)	(0.194)	(0.144)	(0.031)	(0.087)	(0.177)	(0.144)
treatedpost (female)	-0.190	0.076	0.021	0.399	0.142	0.034	0.230**	-0.235	0.138
	(0.320)	(0.284)	(0.093)	(0.318)	(0.205)	(0.045)	(0.097)	(0.219)	(0.258)
Observations	4135	4135	4135	3443	3443	3443	2995	3443	3443
Mean in control	2.551	2.297	0.162	-1.189	-1.092	0.866	0.545	2.689	2.407

Notes: The table presents the treatment effects on education outcomes for Roma children. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2
Non-Roma children by gender.

	Language	Math	Dropout	Stand. Test Scores Language	Stand. Test Scores Math	Secondary School Enrollment	Secondary School Four-year track	Conditional on St. Test Score Language	Conditional on St. Test Score Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
treatedpost (male)	-0.046	-0.008	0.047	0.079	0.090	0.022*	0.003	-0.042	0.007
	(0.084)	(0.070)	(0.036)	(0.090)	(0.102)	(0.012)	(0.031)	(0.093)	(0.124)
treatedpost (female)	-0.006	0.078	0.009	0.115	0.105	-0.006	0.028	-0.054	0.030
	(0.094)	(0.070)	(0.020)	(0.086)	(0.124)	(0.012)	(0.018)	(0.107)	(0.120
Observations	40,824	40,824	40,825	39,453	39,453	39,453	37,557	39,453	39,453
Mean in control	3.712	3.368	0.042	-0.169	-0.178	0.951	0.826	3.768	3.421

Notes: The table presents the treatment effects on education outcomes for Non-Roma children. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3
Non-Roma children: Marks and standardized test scores.

	Language	Math	Lower than median Language	Lower than median Math	First quartile Language	First quartile Math
	(1)	(2)	(3)	(4)	(5)	(6)
Std. testscore Serbian	0.738***		0.731***		0.728***	
	(0.012)		(0.015)		(0.008)	
treatedpost	-0.049	0.020	0.012	-0.153	0.030	-0.279***
	(0.096)	(0.116)	(0.120)	(0.151)	(0.051)	(0.055)
post	-0.089	-0.130	-0.171**	-0.020	-0.265***	0.027
	(0.091)	(0.089)	(0.078)	(0.107)	(0.045)	(0.051)
Std. testscore Math		0.882***		0.886***		0.886***
		(0.010)		(0.013)		(0.008)
Observations	39,453	39,453	21,149	21,149	11,580	11,580
Mean in control	3.768	3.421	3.760	3.384	3.760	3.285

Notes: The table presents the treatment effects on education outcomes for Non-Roma children. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4
Characteristics of the assistant.

	Language								Math							
	Assistant		University		Tenure		Age		Assistant		University		Tenure		Age	
	Male	Female	Without	With	Longer	Shorter	Higher	Lower	Male	Female	Without	With	Longer	Shorter	Higher	Lower
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
treatedpost	-0.148	-0.143	-0.131	-0.115	-0.121	-0.164	-0.152	-0.139	0.040	-0.020	0.034	-0.041	0.067	-0.041	-0.028	0.034
	(0.180)	(0.175)	(0.171)	(0.197)	(0.174)	(0.179)	(0.197)	(0.169)	(0.153)	(0.165)	(0.153)	(0.181)	(0.148)	(0.163)	(0.160)	(0.157)
post	-0.279	-0.278	-0.287	-0.282	-0.271	-0.288	-0.316*	-0.253	-0.311*	-0.290*	-0.294*	-0.324*	-0.303*	-0.300*	-0.295*	-0.305*
	(0.174)	(0.182)	(0.174)	(0.185)	(0.174)	(0.181)	(0.179)	(0.176)	(0.162)	(0.170)	(0.162)	(0.172)	(0.164)	(0.167)	(0.169)	(0.163)
Observations	2742	2396	3506	1583	2664	2474	2160	2978	2742	2396	3506	1583	2664	2474	2160	2978
Mean in control	2.551	2.551	2.551	2.551	2.551	2.551	2.551	2.551	2.297	2.297	2.297	2.297	2.297	2.297	2.297	2.297
Placebo 2009 to 2012 versus 2008																
treatedpost	-0.144	-0.271	-0.173	-0.285	-0.197	-0.205	-0.201	-0.203	-0.007	-0.080	-0.022	-0.100	0.007	-0.083	-0.072	-0.018
	(0.197)	(0.195)	(0.191)	(0.204)	(0.199)	(0.194)	(0.219)	(0.185)	(0.175)	(0.174)	(0.168)	(0.186)	(0.175)	(0.172)	(0.194)	(0.164)
Observations	2081	1826	2705	1165	2129	1778	1599	2308	2081	1826	2705	1165	2129	1778	1599	2308
Mean in control	2.564	2.564	2.564	2.564	2.564	2.564	2.564	2.564	2.299	2.299	2.299	2.299	2.299	2.299	2.299	2.299

	Dropout							
	Assistant		University		Tenure		Age	
	Male	Female	Without	With	Longer	Shorter	Higher	Lower
	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
treatedpost	-0.020	0.002	-0.035	0.056	-0.028	0.001	-0.019	-0.009
	(0.067)	(0.058)	(0.059)	(0.062)	(0.067)	(0.058)	(0.063)	(0.061)
post	-0.284***	-0.266***	-0.279***	-0.266***	-0.281***	-0.268***	-0.274***	-0.273***
	(0.047)	(0.046)	(0.046)	(0.047)	(0.047)	(0.046)	(0.046)	(0.046)
Observations	2742	2396	3506	1583	2664	2474	2160	2978
Mean in control	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.162
Placebo 2009 to 2012 versus 2008								
treatedpost	-0.033	0.024	-0.032	0.090	-0.017	0.002	-0.035	0.012
	(0.056)	(0.046)	(0.048)	(0.057)	(0.056)	(0.048)	(0.058)	(0.048)
Observations	2083	1828	2707	1167	2131	1780	1601	2310
Mean in control	0.357	0.357	0.357	0.357	0.357	0.357	0.357	0.357

Stand. Test Scores

	Language								Math							
	Assistant		University		Tenure		Age		Assistant		University		Tenure		Age	
	Male	Female	Without	With	Longer	Shorter	Higher	Lower	Male	Female	Without	With	Longer	Shorter	Higher	Lower
	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)
treatedpost	-0.005	0.230	0.097	0.234	0.134	0.092	0.093	0.125	0.337	0.286	0.303	0.363	0.370	0.253	0.290	0.318
	(0.180)	(0.164)	(0.160)	(0.192)	(0.171)	(0.167)	(0.158)	(0.171)	(0.258)	(0.242)	(0.244)	(0.262)	(0.248)	(0.245)	(0.245)	(0.248)
post	-0.467***	-0.347**	-0.432***	-0.375**	-0.419***	-0.403***	-0.404***	-0.415***	-0.545**	-0.402	-0.504**	-0.434	-0.495*	-0.461*	-0.483*	-0.477*
	(0.139)	(0.151)	(0.140)	(0.156)	(0.142)	(0.148)	(0.148)	(0.142)	(0.249)	(0.261)	(0.246)	(0.273)	(0.251)	(0.258)	(0.259)	(0.251)
Observations	2248	2039	2892	1350	2192	2095	1770	2517	2248	2039	2892	1350	2192	2095	1770	2517
Mean in control	-1.092	-1.092	-1.092	-1.092	-1.092	-1.092	-1.092	-1.092	-1.189	-1.189	-1.189	-1.189	-1.189	-1.189	-1.189	-1.189
Placebo 2009 to 2012 versus 2008																
treatedpost	-0.005	0.174	0.085	0.087	0.124	0.032	0.049	0.103	0.231	0.013	0.147	0.049	0.229	0.022	0.130	0.132
	(0.221)	(0.187)	(0.195)	(0.219)	(0.217)	(0.193)	(0.196)	(0.206)	(0.314)	(0.293)	(0.299)	(0.312)	(0.310)	(0.295)	(0.315)	(0.300)
Observations	1347	1202	1732	784	1346	1203	1017	1532	1347	1202	1732	784	1346	1203	1017	1532
Mean in contro	l -1.086	-1.086	-1.086	-1.086	-1.086	-1.086	-1.086	-1.086	-1.166	-1.166	-1.166	-1.166	-1.166	-1.166	-1.166	-1.166

Notes: The table presents the treatment effects on education outcomes for Roma children depending on the characteristics of the assistant. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5
Assistant's and Student's Gender.

	Language		Math		Dropout		Stand. Test Scores				Secondary School			
							Language		Math		Enrollment		Four-year Track	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
treatedpost (male)	-0.060	-0.185	-0.002	-0.138	-0.046	-0.011	0.099	0.085	0.418**	0.046	0.030	0.004	0.126	0.166*
	(0.167)	(0.158)	(0.103)	(0.123)	(0.092)	(0.079)	(0.185)	(0.181)	(0.209)	(0.221)	(0.043)	(0.030)	(0.101)	(0.090)
treatedpost (female)	-0.267	-0.098	0.062	0.106	0.017	0.021	-0.120	0.390*	0.245	0.547*	0.007	0.072*	0.181*	0.272**
	(0.326)	(0.329)	(0.287)	(0.295)	(0.097)	(0.103)	(0.225)	(0.204)	(0.349)	(0.309)	(0.059)	(0.042)	(0.105)	(0.106)
Observations	2742	2396	2742	2396	2742	2396	2248	2039	2248	2039	2248	2039	1925	1807
Mean in control	2.551	2.551	2.297	2.297	0.162	0.162	-1.092	-1.092	-1.189	-1.189	0.866	0.866	0.545	0.545

Notes: The table presents the treatment effects on education outcomes for Roma children by gender depending on the gender of the assistant. All regressions control for an indicator variable for the year of the survey and school fixed effects. They also include a dummy for gender, a dummy for being a pupil born in a different district from the one where the school is located, age at test, class size and class size squared. Robust standard errors are clustered at the school level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A2. Additional figures

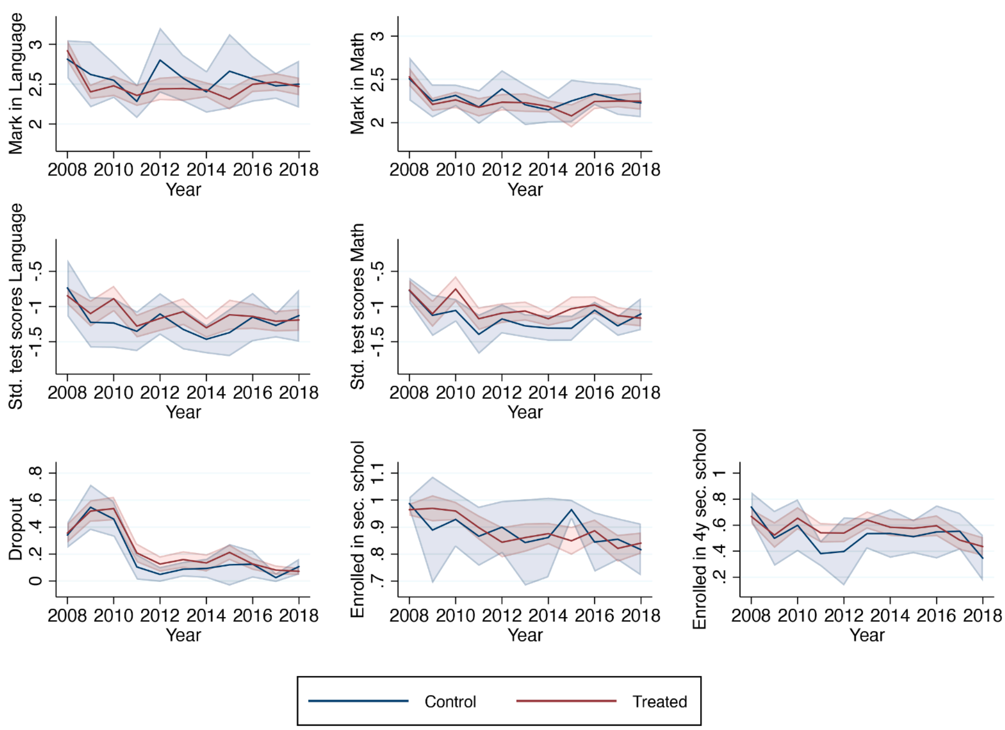

Fig. A.1. Roma: Trends of outcome variables.

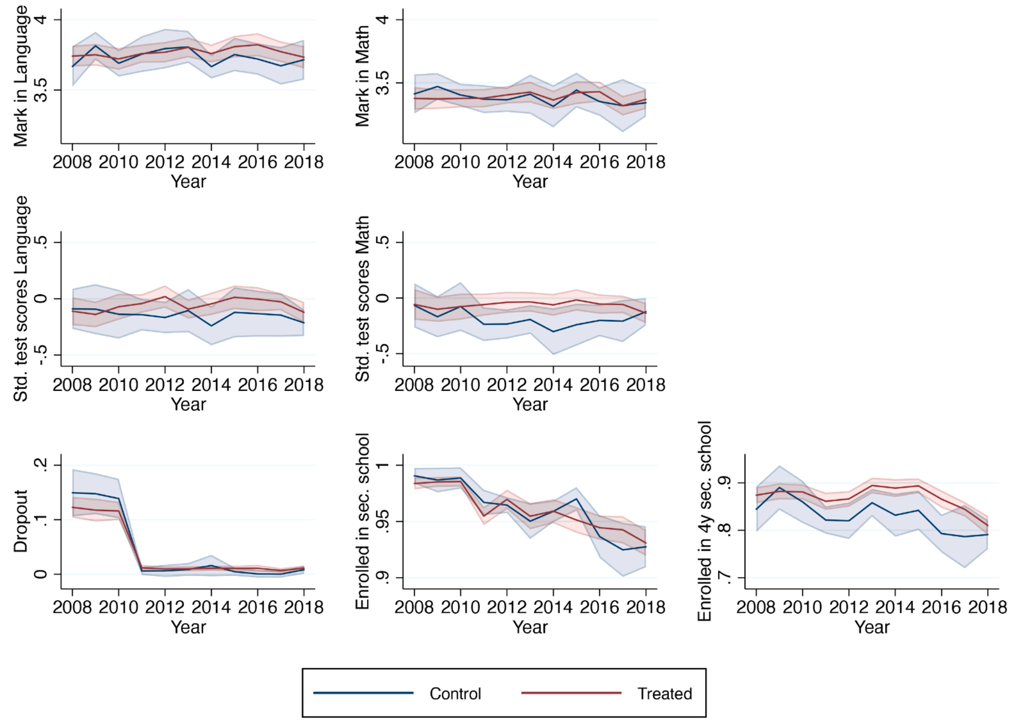

Fig. A.2. Non-Roma: Trends of outcome variables.

A3. Overall impact of RTA

Table A.6 reports the RTA impacts for the overall sample and allows to observe differential effects between Roma and Non-Roma pupils. Columns (1) to (3) report the program impacts on the marks in Language, columns (4) to (6) on the marks in Math, columns (7) to (9) on the probability that students do not take the standardized exam and therefore do not finish primary school, columns (10) to (12) on the standardized test scores in Language, columns (13) to (15) on the standardized test in Math, columns (16) to (18) on enrollment in secondary school, and columns (19) to (21) on the probability of choosing a four-year track. More precisely, the first column of each outcome reports the raw impact of the RTA, the second adds the control variables (gender, whether the pupil was born in a different district from the one where the school is located, age at test, class size and class size squared), and the third the interaction with the ethnicity of the child.

The RTA does not have statistically significant effects on any of the outcomes of interest for the all sample (Panel A). When we look at heterogeneous impacts by ethnicity (third column of each outcome), we again observe not statistically significant effects, although the direction of the coefficients is the one expected and suggests higher positive effects in treated schools for Roma children. Panel B reports the coefficients of the placebo regressions, confirming again the comparability of our control group. All these results together seem to suggest modest average effects of the program and hardly indicate important reductions in the achievement gap between Roma and Non-Roma.

Table A.6
All sample.

							Dropout			Standardized Test Scores						Secondary School					
	Language			Math						Language			Math			Enrollment			Four-year Track		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Panel A: Main results																					
treatedpost	-0.023	-0.045	-0.022	0.031	0.024	0.037	0.028	0.025	0.031	0.123	0.107	0.108	0.116	0.099	0.105	0.008	0.008	0.009	0.021	0.024	0.019
	(0.081)	(0.082)	(0.082)	(0.062)	(0.061)	(0.062)	(0.027)	(0.026)	(0.026)	(0.081)	(0.081)	(0.076)	(0.109)	(0.110)	(0.112)	(0.009)	(0.011)	(0.011)	(0.019)	(0.019)	(0.019)
post	0.002	-0.043	-0.013	-0.069	-0.105**	-0.083	-0.153***	-0.148***	-0.138***	-0.077	-0.131*	-0.103*	-0.151*	-0.168*	-0.150*	-0.051***	-0.073***	-0.068***	-0.042**	-0.094***	-0.078***
	(0.073)	(0.086)	(0.087)	(0.051)	(0.052)	(0.051)	(0.025)	(0.023)	(0.023)	(0.065)	(0.068)	(0.062)	(0.090)	(0.088)	(0.089)	(0.007)	(0.009)	(0.009)	(0.018)	(0.017)	(0.018)
roma	-1.187***	-1.129***	-0.834***	-1.104***	-1.054***	-0.855***	0.120***	0.107***	0.183***	-1.000***	-0.959***	-0.567***	-0.904***	-0.872***	-0.606***	-0.076***	-0.062***	0.004	-0.285***	-0.269***	-0.078
	(0.043)	(0.042)	(0.093)	(0.037)	(0.037)	(0.071)	(0.011)	(0.010)	(0.037)	(0.051)	(0.048)	(0.146)	(0.036)	(0.034)	(0.087)	(0.010)	(0.009)	(0.009)	(0.023)	(0.022)	(0.065)
treatedpost × roma			-0.176			-0.085			-0.028			0.054			-0.034			0.008			0.095
			(0.140)			(0.122)			(0.046)			(0.183)			(0.107)			(0.027)			(0.060)
postroma			-0.305**			-0.230**			-0.109***			-0.377**			-0.228***			-0.066***			-0.211***
			(0.129)			(0.111)			(0.036)			(0.172)			(0.087)			(0.022)			(0.053)
treatedroma			0.053			0.035			0.034			-0.161			-0.084			-0.023			-0.106
			(0.112)			(0.090)			(0.048)			(0.155)			(0.102)			(0.014)			(0.070)
Observations	44,959	44,959	44,959	44,959	44,959	44,959	44,960	44,960	44,960	42,896	42,896	42,896	42,896	42,896	42,896	42,896	42,896	42,896	40,552	40,552	40,552
Mean in control	3.624	3.624	3.624	3.287	3.287	3.287	0.051	0.051	0.051	-0.237	-0.237	-0.237	-0.239	-0.239	-0.239	0.946	0.946	0.946	0.808	0.808	0.808
Panel B: Placebo 2009 to 2012 versus 2008																					
treatedpost	-0.100	-0.085	-0.064	-0.009	-0.001	0.012	0.017	0.024	0.024	0.075	0.073	0.071	0.082	0.090	0.094	0.005	0.005	0.004	-0.005	-0.001	-0.009
	(0.094)	(0.091)	(0.092)	(0.066)	(0.065)	(0.065)	(0.018)	(0.019)	(0.018)	(0.071)	(0.077)	(0.070)	(0.107)	(0.112)	(0.112)	(0.005)	(0.005)	(0.005)	(0.013)	(0.014)	(0.014)
treatedpost × roma			-0.200			-0.112			0.015			0.027			-0.042			0.027			0.137
			(0.153)			(0.140)			(0.045)			(0.229)			(0.124)			(0.031)			(0.083)
Observations	42,765	42,765	42,765	42,765	42,765	42,765	42,770	42,770	42,770	38,413	38,413	38,413	38,413	38,413	38,413	38,413	38,413	38,413	29,677	29,677	29,677
Mean in control	3.651	3.651	3.651	3.317	3.317	3.317	0.118	0.118	0.118	-0.193	-0.193	-0.193	-0.199	-0.199	-0.199	0.976	0.976	0.976	0.858	0.858	0.858

Notes: The table presents the treatment effects on education outcomes for the entire sample. All regressions control for an indicator variable for the year of the survey and school fixed effects. Columns (1), (4), (7), (10), (13), (16) and (19) report the estimates without controls. Standard errors are clustered at the school level (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

A4. Additional results subjective versus objective grading

Roma In the main text, we show an empirical investigation of marks versus standardized test scores. Here, we provide additional graphical evidence and discussion. Fig. A.3 reports the school mark distribution of Roma pupils in both Language and Math, before and after the introduction of the program, in treated (black) and control (gray) schools. These schools have a percentage of Roma at application lower than the median.[40] In both subjects, after the implementation of the program, we observe a polarization of marks around 2, the sufficient mark to pass the course. It comes at the cost of less medium marks (3 and 4), especially in treated schools, where however the number of children who get an insufficient mark also increases. Interestingly, if we look at standardized test scores by mark received in the subject (Fig. A.4), we observe that Roma in treated schools perform better than their counterparts in control schools, and better than before the program. Such a shift to the right

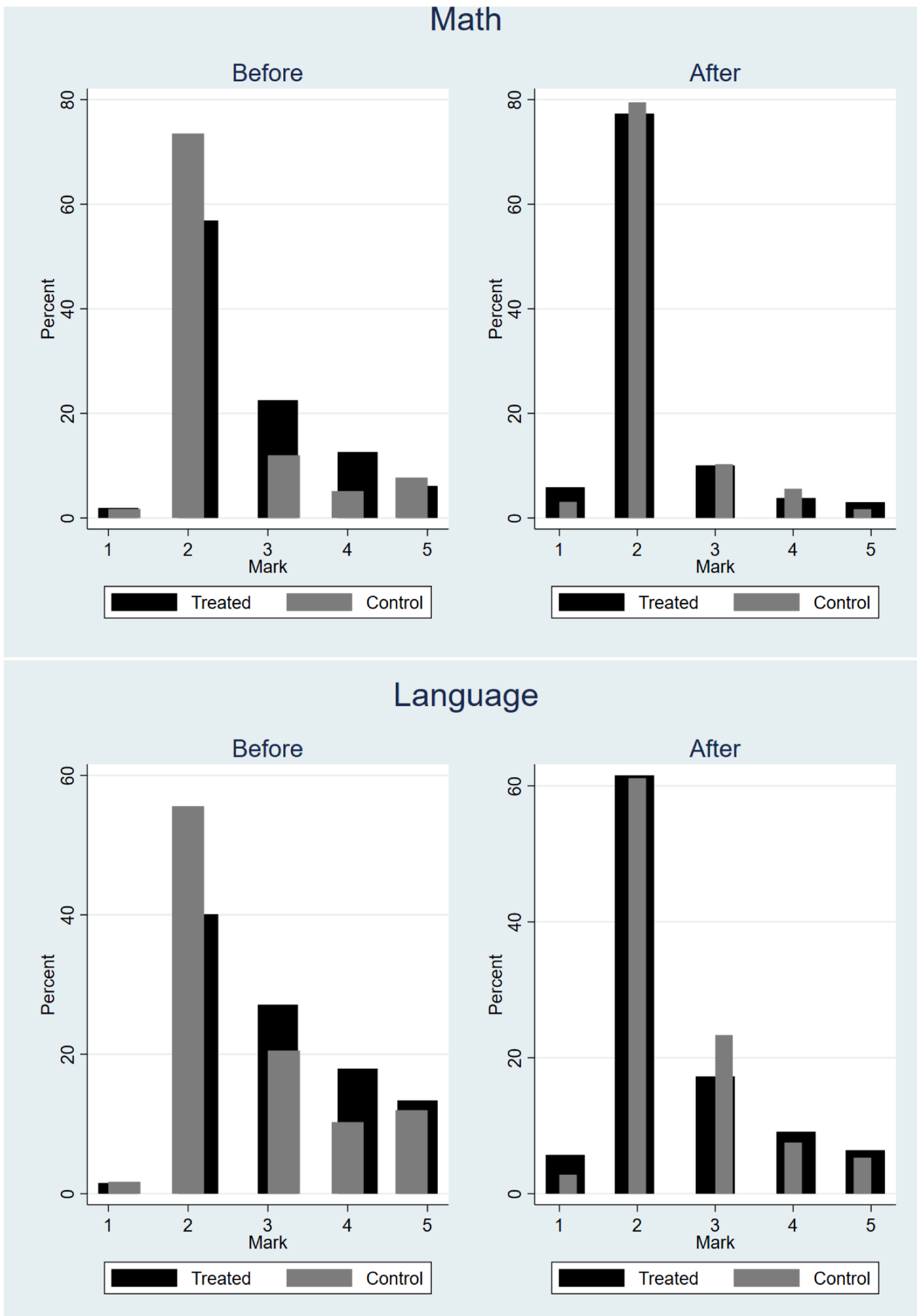

Fig. A.3. Marks of Roma people (Lower than the median).

[40] For the entire sample of schools, the graphs are entirely comparable.

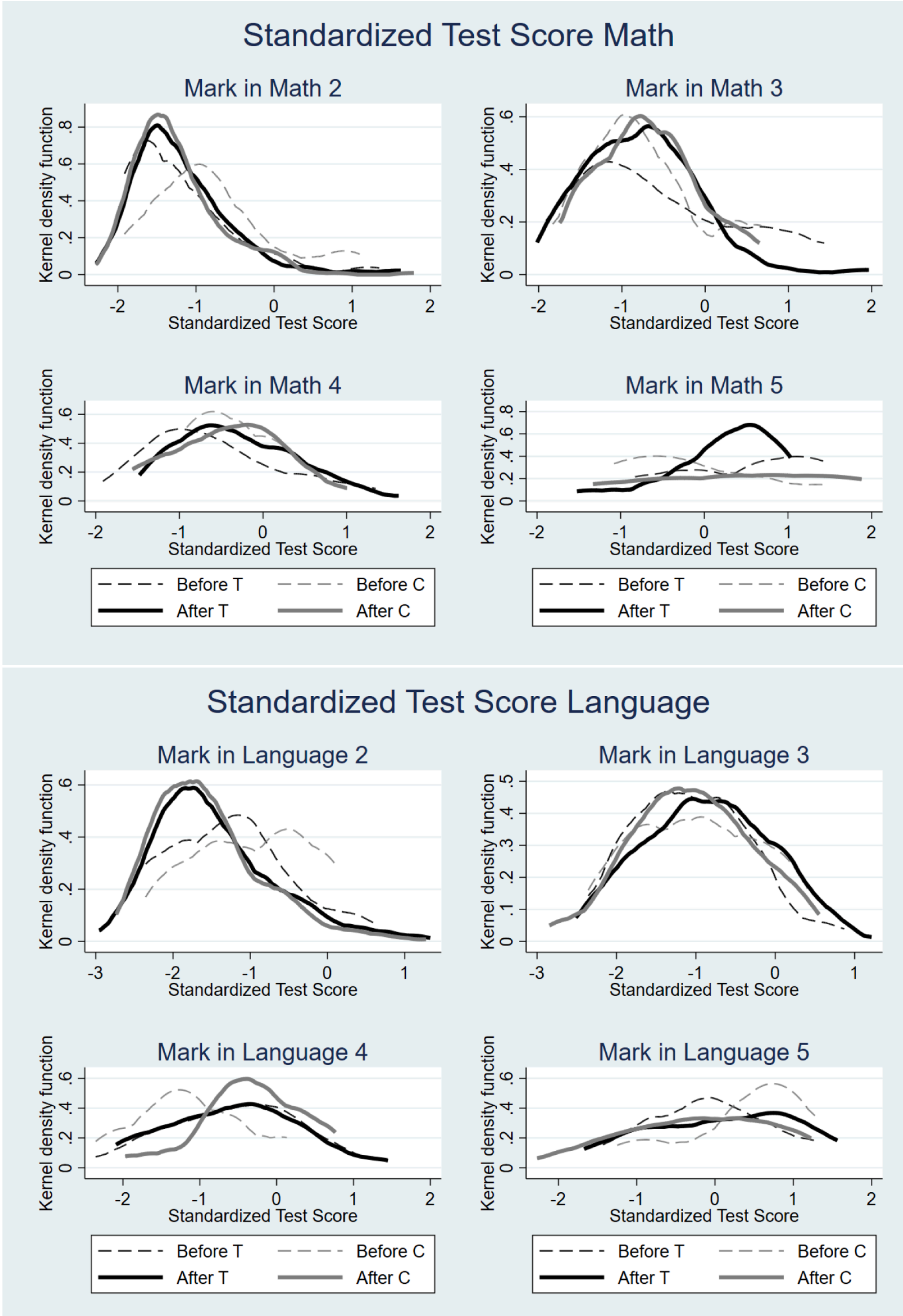

Fig. A.4. Standardized test score of Roma people (Lower than the median).

in the score distribution is bigger for those children graded the lowest passing marks, 2 and 3. The impacts are stronger for Math than Language, but overall they seem to suggest that lower marks in treated schools than in control are not followed by worse performances in the final exam.

Non-Roma We show the graphical distribution of marks and standardized test scores for Non-Roma children. In Fig. A.5, there are no relevant differences in neither subject between Non-Roma students in treated and control schools, before and after the introduction of the program. We observe some improvements in treated schools in Language, since we find a small shift from mark 2 to mark 3, but not much differences in standardized tests (Fig. A.6).

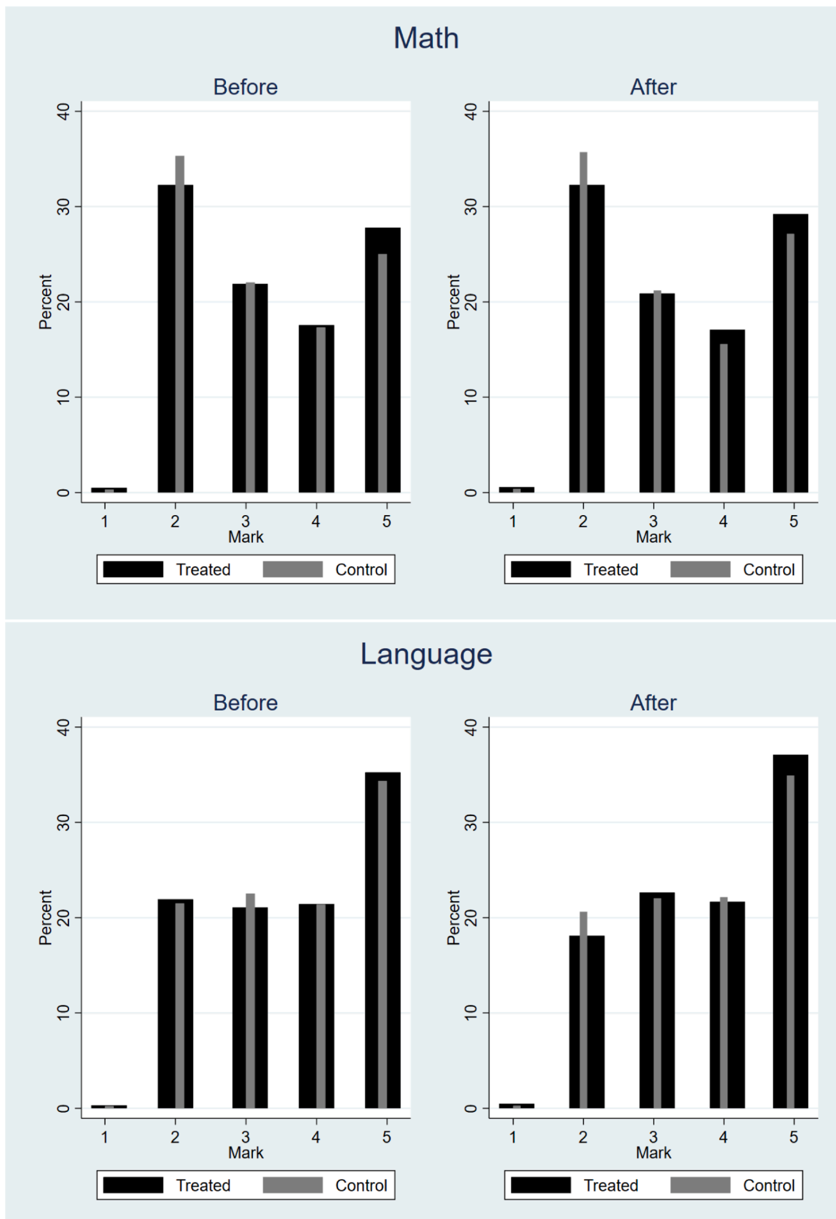

Fig. A.5. Marks of Non-Roma people (Lower than the median).

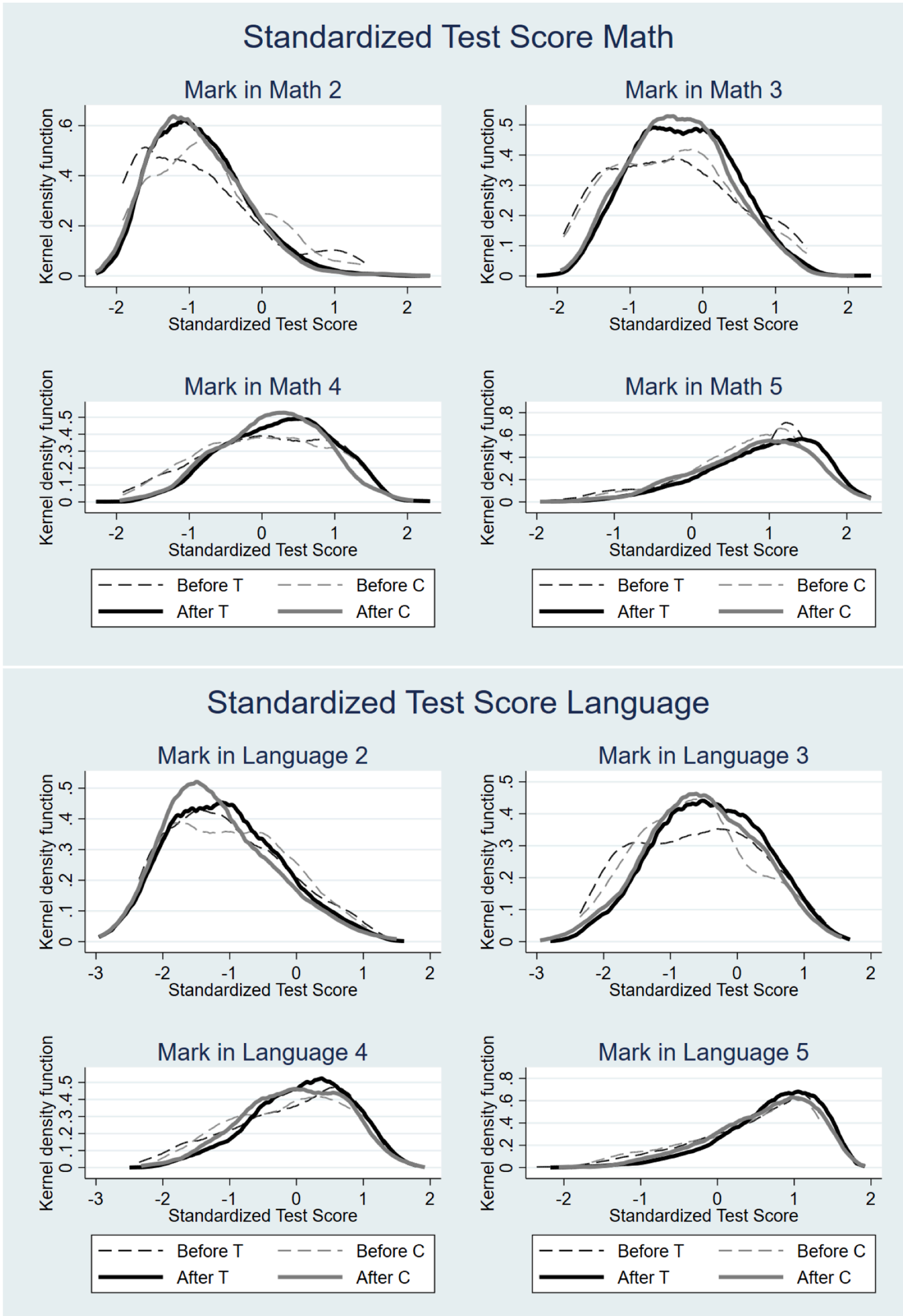

Fig. A.6. Standardized test score of Non-Roma people (Lower than the median).

References

Abdullah, A., Doucouliagos, H., & Manning, E. (2015). Does education reduce income inequality? ameta-regression analysis. *Journal of economic surveys, 29*(2), 301–316.

Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? evidence from boston's metco program. *The American economic review, 94(5)*, 1613–1634.

Ashenfelter, O., & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *The American economic review, 84*(5), 1157–1173.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The quarterly journal of economics, 122* (3), 1235–1264.

Battaglia, M., & Lebedinski, L. (2015). Equal access to education: An evaluation of the Roma Teaching Assistant program in serbia. *World development, 76*, 62–81.

Battaglia, M., & Lebedinski, L. (2017). The curse of low expectations: Remedial education and perceived returns to education of Roma people. *Economics of Transition, 25*(4), 681–721.

Baucal, A. (2006). Development of mathematical and language literacy among roma students. *Psihologija, 39*(2), 207–227.

Bhabha, J., Matache, M., Chernoff, M., Fuller, A., Lloyd McGarry, S., Simić, N., … Mihajlović, M. (2018). One in one hundred: Drivers of success and resilience among college-Educated romani adolescents in serbia. *FXB Center for Health and Human Rights, Harvard University*.https://cdn2.sph.harvard.edu/wp-content/uploads/sites/114/2018/12/OneinOneHundred.pdf

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Education Evaluation and Policy Analysis, 31(4)*, 416–440.

Burgess, S., & Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of labor economics, 31*(3), 535–576.

Calsamiglia, C., & Loviglio, A. (2019). Grading on a curve: When having good peers is not good. *Economics of education review, 73*, 101916. https://doi.org/10.1016/j.econedurev.2019.101916

Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human resources, 48*(1), 236–264.

Currie, J., & Thomas, D. (1995). Does head start make a difference? *American Economic Review, 85*(3), 341–364.

Diamond, R., & Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. *NBER Working Papers, 22207*.

Domina, T. (2005). Leveling the home advantage: Assessing the effectiveness of parental involvement in elementary school. *Sociology of education, 78(3)*, 233–249.

Fryer, R. G. (2010). Racial inequality in the 21st century: The declining significance of discrimination. *Handbook of Labor Economics, 4*, 855–971.

García-Pérez, J. I., & Hidalgo-Hidalgo, M. (2017). No student left behind? evidence from the programme for school guidance in spain. *Economics of education review, 60*, 97–111.

Gershenson, S., Hart, C. M. D., Hyman, J., Lindsay, C., & Papageorge, N. W. (2018). The Long-Run Impacts of Same-Race Teachers. *Working Paper*. National Bureau of Economic Research.

Gutiérrez, E., & Rodrigo, R. (2014). Closing the achievement gap in mathematics: Evidence from a remedial program in mexico city. *Latin American economic review, 23* (1), 14.

Hainmueller, J. (2012). Entropy balancing for causal effects: A Multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis, 20* (01), 25–46.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of education review, 30*(3), 466–479.

Harmon, C., Oosterbeek, H., & Walker, I. (2003). The returns to education: Microeconomics. *Journal of economic surveys, 17*(2), 115–156.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *American Economic Review, 94 (1)*, 233–258.

Keeley, B. (2015). *Income inequality*. OECD. https://doi.org/10.1787/9789264246010-en

Kenrick, D., & Taylor, G. (1998). Historical dictionary of the gypsies (romanies). In *European Historical Dictionaries*. Scarecrow Press. https://books.google.rs/books?id=BlDYAAAAMAAJ

Kertesi, G., & Kézdi, G. (2011). The Roma/non-Roma test score gap in hungary. *American Economic Review: Papers & Proceedings, 101*(3), 519–525.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? evidence from a natural experiment. *Journal of public economics, 92*(10–11), 2083–2105.

Lavy, V., Kott, A., & Rachkovski, G. (2018). Does Remedial Education at Late Childhood Pay Off After All? Long-Run Consequences for University Schooling, Labor Market Outcomes and Inter-Generational Mobility. *Technical Report*. National Bureau of Economic Research.

Lavy, V., & Schlosser, A. (2005). Targeted remedial education for underperforming teenagers: Costs and benefits. *Journal of labor economics, 23*(4), 839–874.

Lebedinski, L. (2019). The effect of residential segregation on formal and informal employment of roma in serbia. *Eastern European economics, 0*(0), 1–29.

Lindahl, E. (2007). Comparing Teachers' Assessments and National Test Results: Evidence from Sweden. *Technical Report*. Working Paper.

Milivojević, Z. (2015). Analiza Pravnog Okvira i Aktuelnog Statusa i Prakse Pedagoških Asistenata. *Technical Report*. Social Inclusion and Poverty Reduction Unit.

O'Higgins, N. (2012). Roma and Non-Roma in the Labour Market in Central and South Eastern Europe. *Roma Inclusion Working Papers*. Bratislava: United Nations Development Programme.

Perić, T. (2012). The Housing Situation of Roma Communities: Regional Roma Survey 2011. *Roma Inclusion Working Papers*. Bratislava: United Nations Development Programme.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management, 26(4)*, 909–949.

Rodríguez-Pose, A., & Tselios, V. (2009). Education and income inequality in the regions of the european union. *Journal of regional science, 49*(3), 411–437.

Rus, C. (2006). The Situation of Roma School Mediators and Assistants in Europe. COE.

Ryan, A., Kontopantelis, E., Linden, A., & Burgess, J. (2018). Now trending: Coping with non-Parallel trends in difference-in-Differences analysis. *Statistical methods in medical research.* https://doi.org/10.1177/0962280218814570962280218814570

Sanbonmatsu, L., Kling, J. R., Duncan, G. J., & Brooks-Gunn, J. (2006). Neighborhoods and academic achievement: Results from the moving to opportunity experiment. *The Journal of human resources, 41(4)*, 649–691.