BUDAPEST WORKING PAPERS ON THE LABOUR MARKET

BWP – 2013/16

Are you on the right track?

The effect of educational tracks on student achievement in upper-secondary education in Hungary

ZOLTÁN HERMANN

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
HUNGARIAN ACADEMY OF SCIENCES
DEPARTMENT OF HUMAN RESOURCES, CORVINUS UNIVERSITY OF BUDAPEST
BUDAPEST, 2013

Budapest Working Papers on the Labour Market
BWP – 2013/16

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences
Department of Human Resources, Corvinus University of Budapest

Are you on the right track?
The effect of educational tracks on student achievement in upper-secondary education in Hungary

Author:

Zoltán Hermann
research fellow
Institute of Economics
Centre for Economic and Regional Studies
Hungarian Academy of Sciences
email: hermann.zoltan@krtk.mta.hu

November 2013

ISBN 978 615 5243 99 8
ISSN 1785 3788

Are you on the right track?

The effect of educational tracks on student achievement in upper-secondary education in Hungary

Zoltán Hermann

Abstract

The paper attempts to identify causal effects of being enrolled in different educational tracks on student achievement in upper-secondary education in Hungary. Rejected and admitted students are compared who applied to the same school and performed similarly in the previous grade. Average treatment effects on the treated are estimated with a matching method. Results indicate that higher track significantly raises student achievement. Beside the effect of tracks, the schools preferred by students within the tracks also provide better educational quality. Comparing the effects of tracks and differences within the tracks reveals that the advantage of the academic track does not differ from that of better schools in general. At the same time, the vocational track incurs substantial losses that are in part specific to that track.

Keywords: education, tracking, matching, equality of opportunity

JEL classification: I20, I21, I24

Acknowledgement:

I'm grateful for valuable comments for Gábor Kézdi and Júlia Varga.

Az iskolatípus hatása a tanulói teljesítményekre

Eredmények a magyarországi középfokú oktatásról

Hermann Zoltán

Összefoglaló

A tanulmány célja az iskolatípusok tanulói teljesítményekre gyakoroltoksági hatásának elemzése a magyar középfokú oktatásban. Az elemzés olyan elutasított és felvett diákok összehasonlítására épül, akik ugyanabba a középiskolába jelentkeztek és a korábbi tanulmányi eredményeik is hasonlóak. A kezelt csoportra gyakorolt átlagos hatásokat (average treatment effects on the treated) matching módszerrel becsüljük. Az eredmények azt mutatják, hogy a magasabb iskolatípusok javítják a diákok teljesítményét. Az iskolatípusokon belül a népszerűbb iskolák hatása szintén pozitív. A gimnázium hatása nem tér el az iskolatípuson belülijobb iskolák hatásától, míg a szakiskola negatív hatása erősebb ezeknél.

Tárgyszavak: oktatás, iskolatípus, matching, esélyegyenlőség

JEL kódok: I20, I21, I24

Köszönetnyilvánítás
Értékes megjegyzéseikért köszöntettel tartozom Kézdi Gábornak és Varga Júliának.

INTRODUCTION

Looking for institutional determinants of inequality of educational systems early tracking of students is often blamed. However, the impact of tracking is still far from unambiguous and causal evidence is scarce. This paper attempts to identify causal effects of being enrolled in different tracks[1] on student achievement in upper-secondary education in Hungary. The effect of preferred schools within tracks are also estimated and compared with track effects.

Tracking can reduce equality of opportunity if (1) poor students have a higher probability to enroll in a less prestigious track and (2) this track has a detrimental effect on them. The first condition is straightforward to confirm in most cases. In a cross-country comparison the highest segregation is found in countries with extensive tracking (Jenkins-Micklewright-Schnepf, 2008). At the same time, the second condition can not be judged easily due to overwhelming selection problems (Manning-Pischke, 2006). Since the selection into different tracks is most often merit based, academic tracks enroll students who will perform better in part because they are more able and motivated, and learn faster. The key question here is whether differences in student achievement across tracks later on are explained solely by this selection or the quality of education is different, as well. If the educational tracks do not exert an effect on achievement we can hardly expect that abandoning selection into different tracks would improve equality of opportunity. This way track effects on achievement provide indirect evidence on the overall impact of a tracking system, as a necessary condition for tracking to hinder equality of opportunity. The main research question of this paper is whether educational tracks do have a causal impact on student achievement.

However, other educational institutions may also produce similar outcomes than tracking, by sorting students with respect to family background and providing them education of different quality. Both theoretical models and empirical evidence suggest that free school choice may result in sorting of students across schools (Epple-Romano, 1998; MacLeod-Urquiola, 2012; Epple-Newlon-Romano, 2002). Ability sorting within schools (often referred as tracking in the US literature) may have similar effects. Again, as far as sorting is not independent of family background and more selective schools provide better educational quality, inequality of opportunity is endangered, as the opponents of school choice often argue (Ladd, 2002). At the other hand, a positive causal effect on achievement in the more popular schools seems indispensable for school choice to raise productivity. If schools can maintain high reputation and attract many students without true better quality,

[1] The term tracking is used differently in the US and European literature. In Europe tracking refers to streaming students into educational programs defined and regulated by the central government, usually with a more academic or vocational orientation, and sometimes providing different degrees. In the US it means various forms of ability sorting within schools. In this paper tracking refers to the former.

the incentives are not appropriate to improve productivity. Hence, exploring the effect of better schools provides evidence on the potential advantages of school choice for improving average quality and its side effects on equality of opportunity. The second research question of this paper addresses the impact of more popular and prestigious schools on student achievement.

Better school effects are also important for assessing track effects. Contrasting these reveals whether track effects are something special or similar to better school effects. If the latter is the case, ability sorting and stratification in a school choice regime may offset any improvement of equality of educational opportunity that can be expected from de-tracking.

The main contribution of the paper is to add causal evidence to the literature on tracking. Note that most of the analyses about tracking effects are flawed by serious methodological problems (see Betts, 2011 for a review). School-level cross-sectional analyses, asking questions similar to this paper, usually struggle with the endogeneity of tracking and selection issues. The results of US studies on ability sorting and tracking are mixed, both regarding the effect on tracking on equality of opportunity and on mean student achievement (Argy-Rees-Brewer, 1996; Betts-Shkolnik, 2000; Figlio-Page, 2002). Evidence from European countries is scarce and mixed. Dustmann-Puhani-Schönberg (2012) has analyzed the effect of tracks on long-term outcomes in Germany, identifying these from a school entry age rule, but found no significant track effects. Horn (2013) found a positive effect for a small, very selective elite tack in Hungary.

Cross-country analyses (Ammermueller, 2005; Brunello-Checchi, 2007; Schuetz-Ursprung-Woessmann, 2008) suggest that early tracking strengthens the impact of family background on student achievement, and hence it hinders equality of opportunity. Other studies, though using similar data, cast some doubt at this conclusion (Waldinger, 2006). However, cross-country comparison is difficult due to the small number of observations and many possible omitted country characteristics confounding the analysis.

More reliable causal evidence on tracking effects comes from analyses of educational reforms of tracking regimes. For Sweden and Finland respectively Meghir-Palme (2005) and Pekkarinen-Uusitalo-Kerr (2009) provide evidence on a positive effect of de-tracking on equality. However, it is not clear whether the conclusions from reforms several decades ago would still apply in the present context of education, and to what extent were the estimated impacts driven by other elements of the reforms. Guyon-Maurin-McNally (2010) reported that an expansion of enrollment in the academic track increased student achievement in Northern Ireland. At the same time, Malamud-Pop-Eleches (2010) analyzing long-term effect of a policy reform in Romania found that labor market outcomes are hardly affected by attending the general instead of the vocational track, thus observable differences in the outcomes between the graduates are probably mostly driven by selection.

Experimental evidence on tracking is scarce. Duflo-Dupas-Kremer (2011) has found a positive effect of tracking on low achievers in Kenya. However, in an experimental setting behavioral responses (most importantly school choices of students and teachers) are constrained, limiting the generalizability of these results. Altogether, the effect of tracking is still far from unambiguous and causal evidence is limited.

The other strand of related literature is that about the effect of more selective schools or in general better schools. Recently several studies employed innovative strategies to identify causal effects of better schools. Some use lottery-based access to these schools to eliminate selection problems (Cullen-Jacob-Levitt, 2006), but identification is most often built on the comparison of students just below and above the admission threshold in a merit-based selection regime. This strategy provides an estimate of the true causal impact of a certain school on student achievement. This approach has been mainly used in the recent literature to analyze the effect of elite schools. Overall, the findings of the better school literature are even more ambiguous than those of tracking. Cullen-Jacob-Levitt (2006), Abdulkadiroglu-Angrist-Pathak (2011) and Dobbie-Fryer (2011) for the US and Clark (2010) in the UK have found no elite school effect. At the same time, a positive effect of better schools was found by Jackson (2010) for Trinidad and Tobago, Pop-Eleches-Urquiola (2011) for Romania and Janvry-Dustan-Sadoulet (2012) for Mexico.

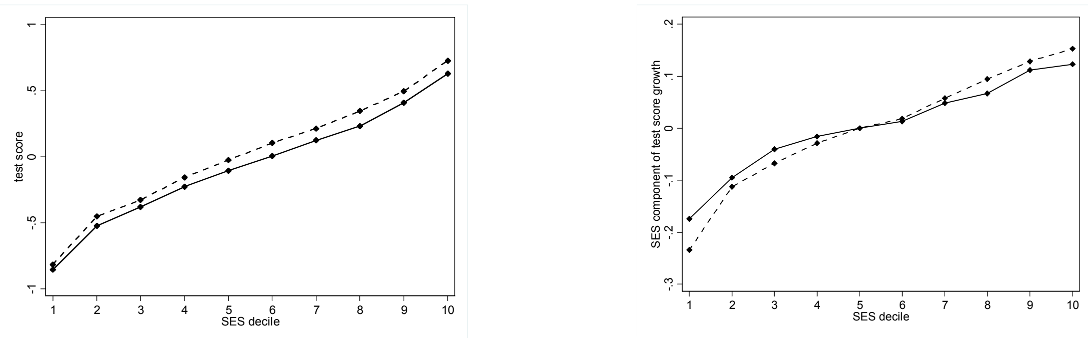
This paper attempts to provide causal evidence on the effect of educational tracks and better schools within the tracks. In order to avoid problems of selection on unobservable traits I use an empirical strategy similar to most papers in the better school literature. I compare rejected and admitted students who applied to the same school and are similar in terms of prior achievement. Identification mostly relies on students on the margin, assuming that these students, as they made similar application decisions, are also similar in unobserved characteristics like motivation, aspirations and self-confidence. However, as no rankings of students by the schools or admission test results are observed, the regression discontinuity method can not be applied. I estimate average treatment effects on the treated using a matching method. With matching based on prior student achievement essentially the same approach is employed, though using a different statistical method. Beside the effect of educational tracks I also estimate the impact of the preferred schools of the students within tracks.

Upper-secondary education in Hungary provides rich opportunities to analyze these questions. A rigid formal tracking regime with three tracks (academic, mixed, vocational) and widespread sorting both across tracks and schools within the tracks are present at the same time. School choice is limited only by capacity constraints, and the allocation of students into tracks and schools is mainly merit based, generating strong ability sorting at both levels. In order to compete for students and distinguish themselves from each other schools offer various educational programs within the three formal tracks.

In Hungary student achievement is exceptionally strongly related to family background, equality of educational opportunity is extremely weak in an international comparison (see the data in Schuetz-Ursprung-Woessmann, 2008). Figure 1 depicts the differences in student achievement over the deciles of students in terms of family socio-economic status. The test score gap between the poor and rich is substantial, exceeding one standard deviation of the test scores in grade 10 (Figure 1, left panel). These differences are already present before entering upper-secondary education, but poor students on average seem to suffer some further losses, as well, as suggested by the estimated family background effect on grade 10 test scores when grade 8 achievement is controlled for (Figure 1, right panel). At the same time, family background is strongly related to track placement in upper-secondary education, students with more disadvantaged family background concentrated in the lower tracks (Figure 2, left panel). Within track sorting is also apparent, as poor students on average attend classes with lower peer quality (Figure 2, right panel). However, sorting within tracks seems to be weaker than between track differences. Altogether this strong sorting implies that if educational tracks and school quality do matter for achievement, poor students can be expected to be disproportionately hurt.

Figure 1

Test scores and family background



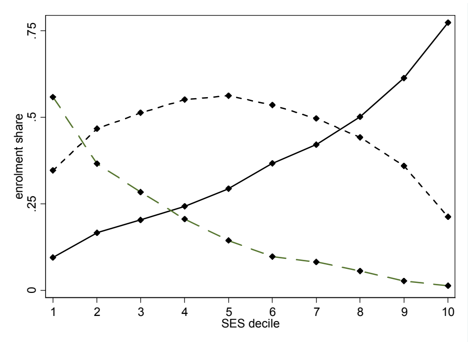
Test score in grade 10 SES component of test score growth from grade 8 to 10 [i]
——— : math - - - : reading
i: coefficients of SES decile dummies from regressions of test score in grade 10, controlling for prior test scores and gender

The analysis uses a large sample covering about one third of a student cohort in Hungary. Compared to the majority of the better school literature, this sample features huge variation in terms of both school quality and the ability level of marginal students, providing an opportunity for analyzing the heterogeneity in track and better school effects.

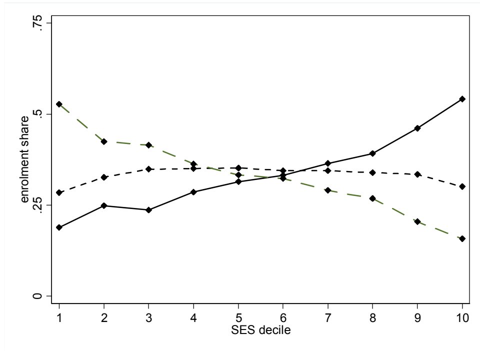
The results of the analysis reveal that higher tracks do indeed have a positive effect on student achievement in Hungary. Better schools within tracks also seem to matter. The test score gain associated with these is similar in magnitude to that of the academic vs. the mixed track (between 0.5 and 0.17 standard deviations). At the same time, attending the vocational instead of the mixed track conveys a 0.21-0.28 standard deviation loss in test scores. Most of this effect persists even if individual heterogeneity of rejected applications is controlled for.

Figure 2

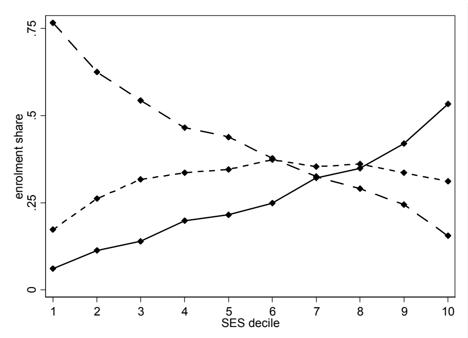
Sorting of students with different family background across and within tracks



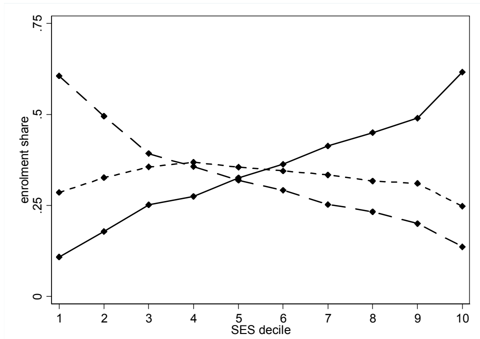
Enrollment share by track



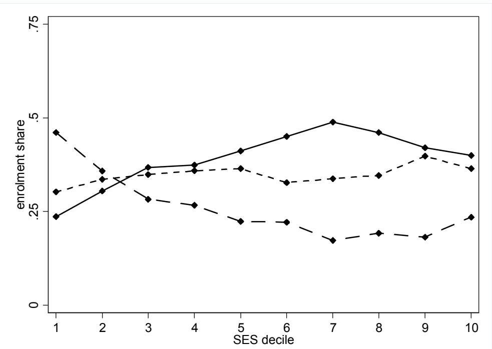
Enrollment share within track in the bottom, middle and top third of schools with respect to peer mean [i]



Enrollment share within the academic track in the bottom, middle and top third of schools with respect to peer mean [i]



Enrollment share within the mixed track in the bottom, middle and top third of schools with respect to peer mean [i]



Enrollment share within the vocational track in the bottom, middle and top third of schools with respect to peer mean [i]

——— : academic track / top third within track, - - - : mixed track / middle third within track,
– – : vocational track / bottom third within track

[i]: the average in the class of the average of math and reading test scores, measured before entering upper-secondary education

The large detrimental effect of the vocational track, together with a relative concentration of poor students in it, suggests that eliminating this track altogether could improve equality of opportunity to some extent. However, part of the track effect would probably be replaced by within track ability sorting. Estimated better school effects imply that the gain of de-tracking would be only about half of that derived solely from the vocational track effect.

Finally, the results reveal substantial heterogeneity in better school and track effects. The less close substitute is the actual school of the student for the preferred one, the larger is the magnitude of the effect. Moreover, the effect diminidhies when the student is farther below the peer mean of the preferred class in terms of prior achievement. Weak students may not benefit from highr standards when these are too demanding. Note that this may explain why no consistent positive effect for elite schools is found in the literature: if the admission cut-off is set far below the mean of admitted students, no significant effects around the cut-off can be expected.

Note that by matching rejected applicants to similar enrolled students the analysis focuses on students on the margin between school tracks. Though these margins prove to be broad, covering a wide range of the ability distribution, marginal students still represent a special group, thus the results can not be assumed to hold for the whole student population. However, the results for the marginal students can be used to assess the effect of small changes in the share of the tracks. There is much less empirical evidence on this than on the existence of tracking (see Brunello-Checchi, 2007; Guyon-Maurin-McNally, 2010), though this seems to be a more accessible policy option than replacing tracking with a comprehensive school system as a whole.

The paper is structured as follows. In the second section the relevant features of secondary education in Hungary are reviewed. The next section introduces the data and presents descriptive analysis. The fourth section covers the empirical strategy. Then the results are summarized and discussed, with a special focus on equality of opportunity. This is followed by robustness and sensitivity analysis and concluding remarks.

UPPER-SECONDARY EDUCATION IN HUNGARY

Primary and secondary education in Hungary is traditionally organized in a two-tier system. Primary and lower-secondary education is provided in general schools, covering 1st-8th grade. At the end of grade 8 students apply for an upper-secondary school. They can choose from three tracks, the academic track, a mixed academic-vocational track and a vocational track.

Enrollment shares in the tracks are similar in magnitude, with the academic and mixed tracks above one third and the vocational track somewhat below that.

The academic and mixed tracks are rather similar. In both cases at the end of grade 12 students take the final maturity exam, which qualifies for higher education. The two tracks follow similar curricula, with only a minor share related to vocational training in the mixed track. After the final exam in grade 12 students can opt for higher education, enter the labor market with their general upper-secondary degree or continue in the mixed track for one or two additional years to receive a vocational qualification.

At the other hand, general curricula in the vocational track are less demanding that in the other two tracks. Vocational training starts after grade 10, either in school workshops or at firms. Students can not take a maturity exam, thus the vocational degree they get does not qualify for higher education. In this respect the vocational track is a dead-end.

There are apparent differences in the prestige of the three tracks. Students almost exclusively rank the academic track before the mixed track and the mixed before the vocational track. Opposite ranking occurs only as an exception in student applications, e.g. when placing a class in the mixed track specialized for arts before one in the academic track. Preferring a class in the vocational track to the mixed track is even more exceptional.

Within the tracks there is significant sorting across schools. Also, schools stream students into different programs, usually organized as separate classes. E.g. within the academic track one class may have a specialization for extra foreign language education, while the other is dedicated to science at a more advanced level. Classes with no specialization are also offered. In the mixed or the vocational track usually the type of occupations defines the program, e.g. technical occupations in agriculture, health care services etc. Though an educational program in a school may include more than one class, for the sake of simplicity I use the term *class* for a program within a school henceforward.

The application and admission to upper-secondary education and the allocation of students is a centralized process. First, students apply for as many school as they wish, ranking their applications from the most to the least preferred one. An application consists of a given class in a given school in a given track. Thus it is also possible to apply for several classes of the same school, or to apply for classes in different schools but in the same track. Second, schools rank the applicants for each different class from the most to the least preferred, but potentially accepted one or strictly reject some of the applicants (the latter implies that the school is not willing to accept the student even if the class will end up with only few students). Schools are relatively free to establish their admittance criteria and to rank the applicants. They can rely on the results from a central entrance exam and the marks received in grade 8 in the general school. Some very popular and selective schools also arrange an oral entrance exam, while others do not require even taking the central entrance exam test, but rank the applications solely based on their marks. Schools may also consider other student characteristics, e.g. religious affiliation when the school is owned by a church or a sibling already attending the school. Finally, students and classes of schools are matched using a centralized allocation algorithm. This is a Gale-Shapely type of algorithm (Kóczy, 2010), providing no incentives for rational students to deviate from their true preferences in the ranking of classes.

Two other features of the institutional setup deserve to be mentioned. First, compulsory education ends at age 18. Thus the vast majority of the cohort analyzed here is in school, selection problems related to dropping out are not considerable. Second, there are few very selective classes within the academic track that start in grade 5 or 7. About 6-8% of students are enrolled in this type of classes and these are not included in the analysis here.

DATA AND DESCRIPTIVE ANALYSIS

The analysis covers a single cohort of students, who finished the general school and started upper-secondary education in 2006. Data comes from the National Assessment of Basic Competences (NABC) and the Upper-Secondary Education Application and Admission (USEAA) data files. NABC files include math and reading literacy scores from standardized tests, average marks in general school and several variables on students' family background. The USEAA file reports the class, school and track for each application, students' preference ranking and the application where the student has been finally admitted as the result of the admission algorithm. However, ranking of applicant students by the schools is not available in the 2006 USEAA file.

Both the NABC and USEAA files are administrative data, covering the full student population with only a few exceptions (e.g. students absent on the day of the test or students with no application). However, the final sample consists only of the set of students whose data can be linked from both of the NABC 2006, the NABC 2008 and the USEAA 2006 files. The final sample covers about one third of the student cohort. The number of students in the original data files exceeds 100000, while the final sample includes 34084 students. The major source of missing data is a random fail to link data for students. The NABC 2006 and NABC 2008 files were linked by general schools on a voluntary basis. Though only about half of the schools were willing to do that, this non-response can be regarded as more or less random. At the same time, even for these schools some of the 8th grader students in 2006 do not appear in the 10th grade NABC 2008 file. Students who drop out or repeat a grade at grade 9 are not observed either. This can be a problem at the lower tail of the ability distribution. Moreover, in grade 8 some students with special education needs were also tested, while in grade 10 they were excluded.

Comparing the sample and the population test score distributions suggests that sample represents well both the 8th and 10th grade student populations. Figure A1 of the Appendix compares the distribution of test scores in the population and the sample. Regarding grade 10 the sample reproduces the population distribution fairly closely. In case of grade 8 the lower tail of the population distribution is somewhat thicker, mainly due to the fact that students with special education needs were included in the grade 8 population, but excluded from grade 10 testing and hence not present in the sample. Altogether the sample represents well the student population with no special education needs.

The information on students' track and school comes from the USEAA file, thus it is measured at the entrance to upper-secondary education. Student movements between tracks and schools in the following years are not considered.

In the estimation of track and school effects student achievement is measured by math and reading literacy test scores, standardized to have a mean of 0 and standard deviation of 1 for the total student population. Table A1 of the Appendix summarizes student achievement in grade 8 and 10 by educational track. Descriptive statistics reveal substantial differences between the tracks. The average student of the academic track has an advantage compared to the mixed track of about one half of a standard deviation, while the average student in the vocational track lags behind that in the mixed track by close to a full standard deviation. These differences are similar for grade 8 and 10.

As the admission to upper-secondary classes is essentially merit-based a strong sorting of students across and within tracks emerges with respect to prior achievement. The variation between the three tracks in grade 9 exceeds the between school variation in grade 8 (Appendix Table A2).

Table 1 presents OLS estimates for track and school quality effects. In these simple specifications test scores in grade 10 are regressed on track dummies, gender and the two prior test scores. The results suggest substantial tracking effects. The expected gain in the academic compared to the vocational track amounts to about one third and one half of a standard deviation in math and reading respectively. The major part of this gap is between the vocational and the mixed track.

Note that these naïve estimates do not account for self-selection of students related to unobserved characteristics. Specifications 2 and 3, including dummies for students rejected from a higher track provide some indirect evidence indicating that selection on unobservables is indeed present. Rejected students tend to outperform both their track- and class-mates with similar prior achievement, who have not applied for a higher track, suggesting that applicants and non-applicants are indeed different in terms of some unobserved characteristics.

Table 1

OLS estimates of test scores in grade 10

	math			reading		
	(1)	(2)	(3)	(1)	(2)	(3)
track (ref.: mixed)						
academic	0.103***	0.106***		0.147***	0.156***	
	(0.010)	(0.010)		(0.009)	(0.010)	
vocational	-0.242***	-0.253***		-0.366***	-0.390***	
	(0.012)	(0.014)		(0.013)	(0.014)	
rejected students (track of preferred – actual class)						
mixed – vocational		0.049***	0.046**		0.115***	0.067***
		(0.018)	(0.021)		(0.021)	(0.022)
academic – mixed		0.017	0.012		0.060***	0.042***
		(0.013)	(0.015)		(0.014)	(0.016)
controls: prior math and reading score, gender	yes	yes	yes	yes	yes	yes
class fixed effects	no	no	yes	no	no	yes
N	33,459	33,459	33,459	33,462	33,462	33,462
R^2	0.687	0.687	0.761	0.670	0.670	0.742

Standard errors clustered for classes are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

EMPIRICAL STRATEGY

A. IDENTIFICATION

The identification of track and better school effects here starts from the fact that if there is some kind of merit-based selection of students into schools, students just below and above the admission threshold are similar to each other. The comparison of these two groups provides an estimate of the true causal impact of a certain school on student achievement. This approach has been used in the recent literature to analyze the effect of elite schools (Abdulkadiroglu-Angrist-Pathak, 2011; Dobbie-Fryer, 2011; Clark, 2010; Jackson, 2010; Pop-Eleches-Urquiola, 2011; de Janvry-Dustan-Sadoulet, 2012). The most straightforward and often applied estimation strategy in this setting is regression discontinuity.

Identification in this paper also relies on the similarity of students close to the admittance cut-off, but the estimation is built on a matching method. Successful applicants are matched with unsuccessful ones with similar prior achievement. Note that here unsuccessful application does not imply that the school definitely rejected the student; instead it simply means that the student were not allocated to her preferred school as a result of the admission algorithm. In fact the schools might have accepted these students, as well, but happened to be able to enroll enough students that they preferred even more. However, for the sake of simplicity I will use the term ‘rejected’ for these students throughout the paper. The matching method is a substitute for regression discontinuity estimation here, since the rankings of students by schools and admission cut-offs are not observed. However, prior student achievement provides a proxy for the unobserved rankings. Therefore the approach underlying regression discontinuity can be implemented by matching. The fact that schools apply various criteria to rank the applicants and the ranking may not reflect exclusively prior achievement, makes using the matching method even more warranted.

The outcome variable is student achievement in grade 10, measured by standardized test results. Rejected students are considered the treatment group, while similar, but successful applicants form the control group. Thus the effect of lower tracks and dispreferred schools within tracks are estimated. The effects are calculated for the rejected students, i.e. average treatment effects for the treated (ATT) are reported.

Formally the effect of being rejected as opposed to being admitted to a given type of school is estimated as the difference of the outcome between the treated student and the mean of its matched control pairs:

(1) $$(A^{10} \mid S = 0) - (A^{10} \mid S = 1) = A_T^{10} - \frac{1}{k} \sum_j A_{Cj}^{10}$$

where *A* denotes test score, *S* the type of school. The upper index is grade, lower indices T and C are the treated or control status of the student and *k* is the number of control students. The average treatment effect for the treated is calculated as the mean of this counterfactual difference for the group of rejected students.

Identification by the matching method is possible if two conditions are met. First, sufficient number of control students should be found who are fairly similar to the treated students in terms of the major determinants of later student achievement. Second, there should be no such differences in unobserved characteristics between the treated and the control group that might affect achievement later on.

Possibly confounding unobserved student characteristics evidently emerge as a major problem when school effects are to be estimated. Even if prior achievement is similar, students in different tracks can be also different regarding their aspirations, motivation and self-confidence. At he same time, these characteristics can evidently affect later achievement (see Christofides et al., 2012; Hastings-Neilson-Zimmerman, 2012; Brunello – Schlotter, 2011). E.g. more motivated students can be expected to take more ambitious choices and perform better in part due to their motivation, which confound the estimation of the true school or track effect. For this reason controlling for prior achievement solely can not eliminate selection bias (Manning-Pischke, 2006).

By matching rejected and admitted students these confounding factors can be eliminated. It is assumed, that both students applying to the same school are similarly motivated, self-confident etc. However, this assumption need not hold for a heterogeneous group of schools. Students applying for different schools within the academic track can well differ: those choosing the elite schools are usually more motivated, confident etc. than those opting for a less prestigious school within the same track. For this reason, treated and control students are matched exactly with respect to school and class. I.e. for each treated case (a rejected application) control students are chosen from those enrolled to the very same class where the treated student applied and was not admitted to.

Do students vary with respect to further unobservable cognitive or non-cognitive traits that do not affect their application decisions but affect achievement later on? I assume that these factors are random comparing the treated and control groups. Since the schools rank the applicants mainly based on entrance exam scores and marks in the general school, further unobservable cognitive or non-cognitive traits are unobservable not only for the researcher but for the schools, as well, thus selection can hardly be based on these. As an exemption, this assumption may not hold for the few elite schools arranging oral entrance exams.

Regarding the overlap condition, identification relies on three features of the application and admission process. First, for each secondary school, some of the rejected students do not meet, but are close to the admission threshold, while few enrolled students are just above this. Prior achievement of students on the margin can be expected to be similar. Second, admission tests and schools measure student achievement and aptitude with error, thus it is possible, that a smart student performs poorly on the day and ends up being rejected, while an even weaker student is admitted. Third, schools may apply somewhat different criteria when ranking applicants, e.g. evaluating the same grade points from the 8th grade differently or giving different weight to non-academic factors (sibling in the school, religious affiliation). The importance of the last two features can be clearly seen from school rankings in other years; it is quite common that schools rank the same applicants differently[2].

When interpreting the results of the matching estimates here, three caveats have to be kept in mind. First, estimates are mainly based on students on the margin. Even though there is no well-defined cut-off between the tracks and thus the margin is a wide band, the students

[2]No school ranking data are available for the cohort analyzed here, but in the next year, 2007 hundreds of thousands of pairs of students can be identified, who applied to the same two schools and both schools ranked each student (i.e. neither school definitely rejected either student). In 29% of these pairs the two schools ranked the two students differently relative to each other. This kind of inconsistency can be observed for each track roughly to the same extent, i.e. it is not a particular feature of a special group of schools.

considered here do not represent the entire population. Second, applicants maximizing their utility can be assumed to pick close substitutes for their preferred schools as second, third etc. options. Thus the estimated effects are not the difference between the *average* classes of two tracks. Third, the estimated effect of educational tracks can be interpreted as a weighted average of individual class effects, weighted by the number of unsuccessful applicants.

B. CORRECTING FOR UNBALANCED COVARIATES AND ESTIMATING AVERAGE TREATMENT EFFECTS

Though treated and control students are similar with respect to prior achievement, they are not identical, especially due to students below and above the margin. Rejected students in general tend to somewhat lag behind the admitted ones. This kind of unbalance may lead to biased estimates (Abadie-Imbens, 2002). In order to control for this bias, I use two alternative methods.

On the one hand, I use the bias-correction proposed by Abadie and Imbens (2002). When the treatment effect is estimated for the treated, first the effect of individual characteristics on the outcome should be estimated parametrically for the control group:

$$(2) \qquad A_{Cj}^{10} = \beta X_{Cj} + \mu_j$$

where X denotes the vector of individual characteristics, β is the estimated vector of parameters and μ is the error term. In the present case I use prior student achievement; math and reading test scores and the average marks score for the correction (see the calculation of the marks score below). Equation (2) is estimated separately by gender.

Then the outcome for the control students is corrected as if their observable characteristics were identical to those of the treated students. The corrected version of eq. 1 is:

$$(3) \qquad (A^{10} \mid S=0) - (A^{10} \mid S=1) = A_T^{10} - \frac{1}{k}\sum_j (A_{Cj}^{10} - \beta X_{Cj} + \beta X_T)$$

The value of this correction depends on whether the specification of the parametric achievement model is correct. Unfortunately, regarding student achievement this proves to be a rather strong assumption when prior achievement is included on the left hand side (Todd-Wolpin, 2003). If the lagged outcome is correlated with the error term, β will be biased, introducing a bias into the treatment effect, as well. The direction of this potential bias is not straightforward in the current case.

On the other hand, a difference-in-difference approach can be employed as an alternative method. Here I measure the effect of being rejected by the difference in the growth of test scores from grade 8 to 10, i.e. whether the achievement growth of the treated student lags behind that of the control students:

$$(4) \qquad [(A^{10} \mid S=0) - A^{8}] - [(A^{10} \mid S=1) - A^{8}] = [A_{T}^{10} - A_{T}^{8}] - \frac{1}{k}\sum_{j}[A_{Cj}^{10} - A_{Cj}^{8}]$$

This second approach assumes that small differences in prior achievement do not affect achievement growth. However, it may be slightly biased by regression to the mean, related to measurement error in the prior test score. Regression to the mean implies that students with lower prior test scores tend to produce larger growth. Since test scores for the treated are typically lower than that of the controls, the difference measure can be expected to slightly underestimate the magnitude of the true effect. If it is biased, the bias is towards zero.

Average treatment effects for the treated are estimated as the mean of differences between treated and control students, calculated for the treated students. For the achievement level and growth models respectively:

$$(5) \qquad ATT^{level} = \frac{1}{n}\sum_{i}\left(A_{Ti}^{10} - \frac{1}{k_i}\sum_{j} A_{Cij}^{10}\right)$$

and

$$(6) \qquad ATT^{diff} = \frac{1}{n}\sum_{i}\left([A_{Ti}^{10} - A_{Ti}^{8}] - \frac{1}{k_i}\sum_{j}[A_{Cij}^{10} - A_{Cij}^{8}]\right)$$

where index i and j denotes treated and control students respectively, n is the number of treated students and k_i is the number of control students belonging to treated student i.

C. MATCHING METHOD

The matching method applied here is radius matching with respect to prior achievement within class and gender. For each rejected applicant control students are matched, who (1) are enrolled in exactly the preferred class, where the treated student was rejected from, (2) the difference between the two students in grade 8 achievement is below a given threshold, and (3) belong to the same gender.

Matching is exact with respect to gender and the class and school the students applied to. In other words, unsuccessful applicants are matched with students enrolled in the very same class in the same school. Exact matching by gender is necessary, since boys and girls may differ not only in achievement level, but also in achievement growth (see e.g. Leahey - Guo, 2001, for the development of math skills). Moreover, since the gender of the teacher may have a different effect on boys' and girls' achievement (Dee, 2005), matching students by gender can reduce measurement error in estimated treatment effects.

When estimating the track and school effects below, student achievement is measured by math and reading literacy test scores. However, when matching treated and control students I use a different achievement measure, the average of marks in grade 8. The reason for this is the presumably substantial measurement error in test scores. Even if the measurement error is distributed randomly across students, it may bias matching estimates. As students scatter around the admission cut-off, measurement error may create false matches, where prior achievement is overestimated for the rejected student and/or underestimated for the control student. Since there is a prior achievement gap within these matched pairs, disguised by the measurement error, their performance later will diverge more than that of true matched students. The problem is that this positive error is not offset by measurement error of the opposite sign, as when the difference in measured prior scores is larger than its true value, no match occurs. To see this assume that the true score of student *A* is relatively far below the admission cut-off in her preferred school, thus if there is no measurement error, student *A* will be not matched with any control students. If measurement error is negative there is still no match, but a positive measurement error may result in finding a seemingly similar control student, *B*. However, the difference between student *A* and *B* in grade 10 can be expected to be quite large, since *A* is in fact a much less able student than *B*, resulting in an upward bias in the effect of the school. The problem is that positive error is not offset by negative measurement error, since in the latter case the student does not appear in the matched sample. Note that, as expected, matching by test scores do indeed tend to provide somewhat larger treatment effects than the preferred matching method, especially for test score growth (see the section on sensitivity analysis below).

In order to avoid this bias, matching is based on average marks in grade 8, which measure student performance on a longer time period. However, average marks may also contain some measurement error, due to different grading standards in different schools or other, unobservable factors. To eliminate the potential bias from this I calculate an average marks score that is adjusted for grading standard differences across schools (see Appendix section A.2 for details).

Following Betts-Grogger (2003) and Figlio-Lucas (2004) grading standards of schools can be empirically estimated by a model including school fixed effects:

$$(7) \qquad \overline{A}_{ics}^{8} = \beta_1 M_{ics} + \beta_2 M_{ics}^{2} + \beta_3 M_{ics}^{3} + \beta_4 X_{ics} + \beta_5 Z_{cs} + \sum \gamma_s S_s + \varepsilon_{ics}$$

where $\overline{A}^{8}$ denotes the mean of the math and reading test scores in grade 8, M is average marks in grade 8, S stands for the school dummies, and X and Z are vectors of student and class level control variables for student i in class c in school s. The estimated fixed effects represent the grading standard of schools. Control variables are gender, special education needs status, classes specialized for advanced education of foreign languages or other subjects and share of special education needs students in the class[3]. The predicted values from this model provide a prior achievement score measured by average marks and corrected for differences in grading standards:

$$(8) \qquad \hat{A}_{ics}^{8} = \beta_1 M_{ics} + \beta_2 M_{ics}^{2} + \beta_3 M_{ics}^{3} + \beta_4 X_{ics} + \beta_5 Z_{cs} + \sum \gamma_s S_s$$

$\hat{A}^{8}$ is the expected value of the mean test score given the actual average marks of the student and the grading standard in her school (and the value of the control variables). Note that the grading standard is estimated from all students' test scores in a school, and is not affected by student level measurement error within the school.

Treated and control students are matched with respect to this adjusted average marks measure. The marks score is measured on a scale similar to that of the test scores. The value of the radius is set to 0.25 standard deviation of test scores[4]. Two students can be matched only if the difference between their corrected average marks scores are less than or equal to 0.25.

Additionally, in order to mitigate the problem of potential measurement error in average marks within schools, control students with an 8th grade test score that is statistically significantly different (at the 5% level) from that of the treated student are dropped. That is, even if two students are similar with respect to average marks, if either their math or reading test scores in grade 8 is almost certainly different, they are not matched. The significance of the difference in the test scores is evaluated using the analytical standard errors provided for each score in the NABC file.

---

[3] The control variables may modify the grading standard within the school or regarding the individual student. See e.g. Bonesronning (2008) on gender related differences in grading practices. In fact, including or excluding these variables makes no real difference neither in the predicted achievment nor in the estimated fixed effects. The correlations of these from the two specifications are 0.99.

[4] The results are quite insensitive to the radius value (see the section on sensitivity analysis below.

Note that matching would be more reliable if test scores and average marks were taken into account simultaneously. However, increasing the number of covariates decreases the number of matched observations substantially, though it produces similar results (see the section on sensitivity analysis below).

When using the radius matching the number of controls belonging to one treated student varies. At the same time, matching with replacement implies that one control student may be matched to several treated students. Standard errors of the ATTs are calculated taking into account these features, following the method in Abadie et al. (2004).

D. RESULTS OF MATCHING AND ASSESSING THE IDENTIFICATION ASSUMPTIONS

In order to produce reasonable estimates of treatment effects a matching method requires a non-negligible overlap of treatment and control cases. In this case about half of the treated observations; 8718 rejected applications of 5971 students were successfully matched to control students, given the preferred parameters of matching. Students in the treatment group cover 17.5% of the total sample. The matched applications were rejected from 1941 classes of 691 schools, representing 40% of the total classes in the sample.

The shares of successful matches are similar across and within tracks (Appendix Table A3). Five groups of rejected applications are considered: students rejected from the mixed track and enrolled in a vocational class, students in the mixed track who aspired to the academic track, and students who preferred another class within each of the three tracks[5].

Table 2 provides descriptive data on the matched sample. As expected, rejected applicants on average performed somewhat better than their actual class-mates but lagged behind those in the preferred class. Similarly, students who ended up in a lower track than they preferred perform on average between the averages of the two tracks, but closer to the track they finally enrolled in. Comparing the peer means of the preferred and actual classes reveals a significant jump in peer quality. This suggests that there is large enough variation within the matched pairs of applicants to identify the effect of school quality. This is important, as in a stratified school system with strong sorting across schools students can be expected to relatively easily find close substitutes to their preferred choices. This is a threat for the identification strategy applied here, since if rejected students were enrolled in schools that offer only slightly inferior educational quality, no effect on later achievement could be expected to be detected. However, the diversity of supply at local school markets is always

[5] Note that a minor share of students in the first group in fact applied to an academic track class, but for sake of simplicity this group is considered to have chosen and been rejected from the mixed track. An even smaller number of students made a reverse ranking of tracks. These exceptional cases are treated as if the two applications were within the same track.

limited and students probably include safe options on their preference list, as well. Moreover, the peer mean difference between preferred and actual classes is on average smaller than the peer mean difference between the tracks. This suggests that students rejected from a higher track do indeed end up in above average classes of the lower track.

Table 2

Prior test scores of matched rejected applicants and the peer mean in the preferred and actual classes and tracks

track of preferred - actual class	rejected applicant test score	peer mean in preferred class	peer mean in actual class	difference between classes	peer mean in preferred track	peer mean in actual track	difference between tracks
math							
mixed – vocational	-0.56	-0.19	-0.80	0.61	-0.03	-0.86	0.83
academic – mixed	0.20	0.55	0.06	0.48	0.53	-0.03	0.56
academic – academic	0.46	0.82	0.39	0.43	0.53		-
mixed - mixed	-0.15	0.14	-0.19	0.34	-0.03		-
vocational – vocational	-0.95	-0.76	-0.97	0.21	-0.86		-
reading							
mixed – vocational	-0.47	-0.14	-0.77	0.63	-0.02	-0.87	0.85
academic – mixed	0.35	0.65	0.11	0.53	0.63	-0.02	0.64
academic – academic	0.64	0.84	0.52	0.32	0.63		-
mixed - mixed	-0.11	0.13	-0.19	0.31	-0.02		-
vocational – vocational	-0.92	-0.73	-0.97	0.24	-0.87		-

Figure A2 of the Appendix represents rejected students prior achievement in more detail. It depicts the distribution of test scores for the entire student population and rejected applicants by educational track for grade 8. In terms of prior achievement applicants rejected within each track closely represent the entire student population of the track, on average these students hardly differ from the representative student of the track. At the same time, the distribution of applicants who preferred a higher track is somewhat skewed as these students performed better than the typical student of their actual track, however, test scores in this group of students are still very diverse, representing each part of the distribution of the track except the lower tail. This confirms the observation that there is no sharp cut-off between the tracks. First, the cut-off in the least popular school in the higher track may be different at each local school market. Second, some students do not apply to the least popular school in the higher track, but choose instead a school in the lower track as their second or third option, thus facing a different cut-off.

Identification from rejected applicants assumes that these students are close to the margin. If they were scattered evenly over the prior achievement distribution of their preferred class, that measured test scores were only very weakly related to the unobservable selection criteria of schools. Comparing their prior test scores to the peer mean and distribution of the preferred class confirm that rejected applicants are typically not far from the margin. Prior test scores are on average 0.2-0.35 standard deviation below the peer mean of the preferred class (Table 2), and half of them would be landed in the bottom third of the prior achievement distribution in that class if admitted (Appendix Table A4).

The crucial assumption of the matching approach is that the treated and control groups do not differ in unobserved characteristics that may have an impact on the outcome analyzed. Naturally, this assumption can not be tested directly. Comparing means of prior achievement and family background variables for the treated and control groups shows no major differences between the two groups, which is reassuring regarding unobserved characteristics (Appendix Table A5). However, since rejected applicants tend to be below the margin while control students are above that, prior scores of the latter group are slightly higher and family background is somewhat more favorable. There are minor differences in the average number of books and years of parental education between the two groups. This can also reflect being at the other side of the margin, since prior achievement and family background are strongly correlated. Altogether, the matched sample is to some extent unbalanced in terms of student characteristics, which makes a correction for this necessary.

RESULTS

A. MAIN RESULTS

The estimated average treatment effects of being enrolled in a less preferred class as opposed to a more preferred one are summarized in Table 3. The first two columns of the table presents the estimated effect on the test score level in grade 10, the next two is the same estimator with parametric bias correction (see in the section on empirical strategy). Here the bias correction is based on prior achievement solely (see sensitivity analysis for other specifications). Column 5 and 6 presents results on the test score difference, i.e. achievement growth from grade 8 to grade 10.

In each panel the first two rows indicate the effect of educational tracks. The results suggest that being rejected from the mixed track and enrolled in a vocational class conveys a huge disadvantage. Disregarding the level estimates without bias correction, on average students suffer a 0.21-0.26 and 0.27-0.29 standard deviation loss in 10th grade test scores in math and reading respectively. Another way to asses the size of the effect is to see how the position of the typical student in the test score distribution could have changed. The average rejected applicant is located at the 25th percentile of the math distribution in grade 10 and at the 23rd of reading. If she had been enrolled in a mixed track class, she could have reached the 32th-34th percentile for both subjects.

The estimated negative impact of the mixed track relative to the academic track is smaller, but still statistically significant. Here the loss is about one tenth – one fifth of a standard deviation.

Note that the negative treatment effects for being rejected from a higher track are surprisingly similar to the track effects estimated by OLS; the size of the treatment effect is almost identical to thatt for math and only somewhat smaller for reading (Table 1).

Table 3

Estimated effect of being rejected from the preferred class on student achievement

	level, no bias correction		level, with bias correction		difference		N	
track of preferred - actual class	ATT	se	ATT	se	ATT	se	rejected appl.	control
math								
mixed – vocational	-0.288	0.020	-0.262	0.017	-0.210	0.022	1203	1623
academic – mixed	-0.215	0.019	-0.158	0.016	-0.112	0.021	1234	1534
within any track	-0.178	0.009	-0.148	0.008	-0.097	0.011	6050	7047
academic – academic	-0.218	0.015	-0.166	0.013	-0.113	0.017	2161	2447
mixed - mixed	-0.172	0.014	-0.149	0.012	-0.105	0.016	2568	3275
vocational – vocational	-0.096	0.024	-0.115	0.022	-0.051*	0.028	911	1207
reading								
mixed – vocational	-0.337	0.022	-0.285	0.019	-0.266	0.025	1255	1706
academic – mixed	-0.227	0.020	-0.171	0.017	-0.139	0.023	1245	1557
within any track	-0.160	0.010	-0.123	0.009	-0.101	0.011	6216	7270
academic – academic	-0.157	0.016	-0.111	0.014	-0.078	0.018	2182	2465
mixed - mixed	-0.169	0.015	-0.138	0.013	-0.120	0.017	2627	3344
vocational – vocational	-0.176	0.027	-0.139	0.024	-0.126	0.029	987	1344

*: significant at 10%, all other treatment effects significant at the 1% level

The next four rows display the effect of a less preferred class compared to a more preferred one within the same educational track. Overall, the impact is negative, implying that students indeed choose better schools and classes first. The effect is similar within the three tracks, except for math in within the vocational track where the preferred class makes smaller difference, and for math score growth it is statistically significant only at the 10% level.

Regarding the three estimation methods the results are rather similar. Parametric bias correction of test score level estimates moderates somewhat the average treatment effects, but for most cases the change is insignificant. As expected, estimates for test score growth tend to produce smaller treatment effects, in accordance with the assumed bias toward zero due to regression to the mean. The growth estimates seem to provide conservative results.

How do the effects of educational tacks and preferred classes within tracks compare? Within track effects are smaller in magnitude than the effect of the vocational track, but similar to that of the mixed relative to the academic track. This suggests that the academic track conveys no more advantage than a better school in general, while the losses associated with the vocational classes are in part specific.

However, comparing the average treatment effects is not the best way to assess the difference between track and better school effects. One reason for this is that individual heterogeneity may make the picture more complex. Being rejected can affect individual students heterogeneously. First, more or less able students may profit more or less from a better class. Note that not only the absolute, but also the relative level of ability may matter. If a student's ability is far below that typical in the class, the standards and expectations of teachers may prove to be too difficult for her to meet. Comparing the location of rejected applicants in the prior achievement distribution of the preferred class suggests that there are indeed some differences across the preferred-actual track types (see Table 2 and Appendix Table A4). Within the vocational track rejected students appear to be less far from the typical student in the preferred class than in the other cases.

Second, and more importantly, the difference in the quality of the preferred and the actual classes may also affect the loss incurred by being rejected. If a student is enrolled in a class nearly as good as her preferred one her achievement may not drop too much. At the other hand, if she faces a much worse class than her first choice, she can be expected to suffer severe losses. It seems plausible, that it is easier to find close substitutes within tracks. Thus, when a student gets into a lower track this might imply a larger difference between the quality of the preferred and the actual class than enrolling in a less preferred class within the preferred track. If this is the case, comparing the size of the estimated treatment effects directly may tell only half of the story. The vocational track versus the mixed track *overall* has a stronger negative effect than being rejected from a better school within any track or the mixed track compared to academic one. But is this merely due to the fact that vocational track classes are usually not as close substitutes for the preferred choice as the less preferred options within the same track? If yes, the effect of higher tracks and preferred schools is essentially similar. Or is there any track specific element in the large negative effect of the vocational track?

In order to answer these questions one should compare across and within track effects with a similar degree of substitutability of the actual for the preferred class. This requires controlling for the distance between the preferred and the actual class in terms of expected educational quality. Unfortunately, the expected quality of the classes can not be observed. However, differences in peer means (i.e. the student composition with respect to prior achievement) provide a proxy for quality differences. The two are closely related for several reasons. First, the presence of more able class-mates may generate positive peer-effects (see e.g. Hoxby, 2000; Hanushek et al, 2001; Lavy et al, 2009). Second, if there is positive matching between students and teachers, better students on average also implies better teachers (see e.g. Lankford et al, 2002; Clotfelter et al 2005, 2006). Third, classes with more able students are generally more popular, that is why they can pick the best applicants, thus, these classes are deemed by the public to provide above average quality. Comparing the peer means confirms that the distance is larger when the two classes are not of the same track, especially for students rejected from the mixed and enrolled in the vocational track (Table 2).

In order to take these factors into account I estimate regression models for the treatment effects at the individual level. The dependent variable is the difference in the outcome between the treated and the control students calculated in terms of test score growth from grade 8 to 10, as defined in eq. 4[6]. The observations are the rejected applications. On the left-hand side dummy variables stand for the combinations of the track of the preferred and actual classes. Observations with extreme values of the individual treatment effect; below the 1st and above the 99th percentiles are excluded[7]. Standard errors are clustered for the preferred class.

The model is first estimated with no controls, representing the average treatment effects in this restricted sample (specification 1). Then three controls are included in order to take individual heterogeneity into account: gender; the difference between the peer mean of the preferred and actual class; and prior student achievement relative to the preferred class (specification 2). The latter is measured as the distance of the student's prior test score from the peer mean of the preferred class. As the effect of this relative ability measure turned out to be slightly non-linear, a squared term is also added. Note that as opposed to this relative measure the test score level turned out to be insignificant in preliminary estimates and hence not included. Finally, micro-region fixed-effects are also added to control for differences in the local supply of and demand for schools, especially the diversity of the available options on the local school market (specification 3). The difference of peer means and distance from the preferred class is measured according to the dependent variable, i.e. in math for the math equation, and reading for reading.

Note that since the equations are estimated with no constant term, the coefficients of the track combinations directly represent the estimated average treatment effects when the value of the control variables is set to zero. The control variables are centered to have zero mean for the within track rejected applications. This way comparing the second and third specifications to the first one directly presents how track effects do change when the control variables shifts to values that are typical within track.

Results are summarized in Table 4. Including the micro-region fixed effects does not change the results, thus the second and third specifications are not discussed separately.

[6] Estimates for the bias-corrected level effects are not shown here, but provide similar results.
[7] Estimates for the full sample or that restricted to the 5th - 95th percentiles produce qualitatively similar results.

The results confirm that the difference between the preferred and actual classes does indeed matter. As expected, if the student manages to find a close substitute for her preferred choice, being rejected makes less harm. A higher difference, i.e. being enrolled in a class farther below the preferred one in terms of peer mean is associated with a larger negative effect. For a rejected student a 0.33 s.d. weaker second choice (the average case within tracks) compared to a close substitute in terms of peer mean increases the negative impact of being rejected by about 0.04 s.d., regarding both math and reading achievement.

The distance of the student from the preferred class seems to matter, as well. Being rejected has stronger negative effect when the student is not far below the peer mean of the preferred class. At the same time, a larger distance mitigates the negative effect. In other words, if the student had lagged far behind most of her classmates in the preferred class, she would have benefitted much less or nothing from being admitted. Note that the effect is slightly non-linear for math, resulting in some flattening when the distance from the preferred class is quite large. This implies that beyond some point being far below the level of the preferred class or even farther makes less difference.

Table 4

Regression estimates for the effect of being rejected from the preferred class

	math			reading		
	(1)	(2)	(3)	(1)	(2)	(3)
track of preferred - actual class						
mixed / vocational	-20.133***	-18.130***	-18.414***	-24.142***	-22.157***	-21.960***
	(1.998)	(2.025)	(2.044)	(2.330)	(2.432)	(2.399)
academic / mixed	-10.511***	-9.359***	-9.833***	-15.258***	-13.774***	-13.614***
	(1.828)	(1.814)	(1.821)	(2.063)	(2.096)	(2.115)
academic / academic	-10.307***	-9.777***	-9.284***	-7.389***	-6.895***	-7.122***
	(1.492)	(1.468)	(1.507)	(1.556)	(1.557)	(1.556)
mixed / mixed	-11.463***	-11.751***	-11.929***	-12.069***	-12.370***	-12.467***
	(1.408)	(1.433)	(1.386)	(1.589)	(1.601)	(1.618)
vocational / vocational	-5.005**	-5.705**	-5.663**	-13.033***	-13.545***	-12.913***
	(2.494)	(2.531)	(2.554)	(2.393)	(2.388)	(2.482)
difference in peer mean		-0.136***	-0.135***		-0.129***	-0.132***
		(0.020)	(0.020)		(0.023)	(0.023)
distance from peer mean in preferred class		0.193***	0.195***		0.192***	0.194***
		(0.017)	(0.017)		(0.015)	(0.015)
distance from preferred class squared		-0.00032**	-0.00031**		-0.00021	-0.00023
		(0.00016)	(0.00016)		(0.00016)	(0.00016)
gender: female		-3.826**	-3.596**		-0.477	0.524
		(1.763)	(1.744)		(1.873)	(1.889)
micro-region FE	no	no	yes	no	no	yes
N (rejected applications)	8,319	8,319	8,319	8,542	8,542	8,542
R2	0.038	0.063	0.064	0.044	0.068	0.069

Standard errors clustered for preferred classes are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

These results suggest that too ambitious choices, even if realized, do not necessarily pay off. This is probably related to standards and the level of required effort set by teachers with respect to the capabilities of the typical students in the class (see Moen-Tjelta, 2010 for evidence on setting grading standards this way in Norway). A student who fits well into the class in terms of her skills, abilities and prior knowledge may benefit from higher standards. At the same time, if these are far too demanding for the student, little benefits can be expected. This is also consistent with grading standards heterogeneously affecting the performance of students, as shown by Figlio-Lucas (2004). This raises concern, that higher tracks and better schools may not benefit everyone and could have even negative side effects. For elite schools in Mexico City de Janvry, Dustan and Sadoulet (2012) found that admission increased student achievement, but also increased the probability of dropping out of school. Note that in Hungary this sort of side effect is not plausible, since it is relatively easy for students to move into a lower track or a less demanding school within a track. Hence dropping out from education altogether typically occurs at the low end, in the vocational track, from schools with very poor reputation.

Note that the effects of the distance from the preferred class and of the preferred-actual class difference holds for each preferred-actual track combinations, when estimated separately for the five types (Appendix Table A6).

Gender appears to be related to the size of the treatment effect for math. Girls on average loose more with being rejected in terms of later math achievement, but for reading there are no significant gender differences.

Most interesting is the pattern of differences between the track and better school effects. Altogether the regression results seem to confirm the picture suggested by the average treatment effects in Table 3. When effects are made comparable, being rejected from the mixed track still has much larger negative impact than being rejected within track or from the academic track. The higher track effects shrink only mildly as the controls are set to the average values within track; by about 0.02 standard deviations for being rejected from the mixed track, and even less for the academic track.

In order to assess the differences in the estimated treatment effects of being rejected from a higher track or from a preferred class within tracks, pair wise comparisons of the coefficients of the track combination variables are tested (Appendix Table A7). The effect of being rejected from the mixed track is always statistically significant from any of the within track effects or the average of those or from being rejected from the academic track. At the same time, the negative impact of a less preferred class within the same track is not consistently statistically different for the three tracks. Being rejected from the academic track also implies similar disadvantage as a less preferred class within the same track (except compared to the within academic track effect for reading).

The fact, that when the difference in peer mean and the relative student ability is controlled for, the difference between the vocational track effect and the other effects decreases somewhat confirms that the large negative average treatment effect estimated for the vocational track is indeed in part due to vocational classes being less close substitutes for the preferred choices than alternatives within the same track. However, the small change in the vocational track effect implies that this explains only a minor part of the story. Being rejected from the mixed and enrolled in the vocational track incurs a marked *track specific* negative effect.

Altogether, the results above indicate a significant negative effect of being rejected in general, and an additional negative vocational track effect. For education policy the next question is what makes better schools better? The patterns of the effect of the difference in peer means and the distance of the student from the preferred class are tempting to invoke a pure peer group effect explanation (on peer group effects see e.g. Hoxby, 2000; Hanushek et al, 2001; Lavy et al, 2009). Differences in teacher quality together with positive student-teacher matching are another natural candidate (see e.g. Lankford et al, 2002; Clotfelter et al 2005, 2006 and Varga, 2009 for Hungary). Unfortunately, the available data do not allow for distinguishing the mechanisms behind the effect of better schools and tracks.

A larger gap in peer mean producing a larger negative treatment effect is in accordance with peer composition affecting the performance of students. Also, the finding that the distance of the student from the peer mean does matter is in line with heterogeneous peer effects (Lavy et al, 2009). However, the estimated effects in the regressions above can not be regarded as direct evidence for peer effects, but reflects at best upper bounds for these. First, teacher quality is omitted from the estimates, though it can be correlated with peer quality as far as students and teachers are non-randomly matched. Positive student-teacher matching can be expected when teacher wages are not compensating for student composition and is often observed in the US, see e.g. Lankford et al, 2002; Clotfelter et al 2005, 2006. Varga (2009) provides similar evidence for Hungary, as well. If better students indeed attract better teachers, the peer mean may in part represent teacher quality. Second, it is possible that not the better peer group has a positive effect, but schools providing better education for any reason attract more students and can select the best ones. Again, the incidence of teacher quality is a natural candidate for explaining school quality. In this case the endogenously formed peer mean is simply an indicator for school quality, but not the cause of it. Overall not too much can be said about the mechanisms driving the track and better school effects. The presence of peer effects can not be ruled out, but the available data do not allow for identifying these.

However, peer effects and teacher-student matching can be assumed to be equally at work regarding educational tracks and better schools within tracks. Hence, the outstanding vocational track effect should be triggered by other factors, specific for this track. Differences in curricula and the lower standards required to successfully finish the school are natural candidates.

B. THE IMPACT ON EQUALITY OF OPPORTUNITY

As argued above, track and better school effects are important for educational policy since both have an impact on equality of opportunity. However, the estimated track and better school effects are not enough to assess this impact, as equality of opportunity also depends on who does gain or is hurt by these effects. Tracking is often argued to hinder equality of opportunity, i.e. strengthening the effect of family background on student achievement. As shown above (Figure 2) poor students are sorted into the lower tracks disproportionately and these tracks do have a negative impact on achievement. But how large is the impact on equality of opportunity?

The estimated average treatment effects makes it possible to assess the impact of a marginal change in the shares of the tracks or better schools within tracks, under the assumption that only marginal students are affected and indirect effects on the other students are negligible. In order to explore the implications on equality of opportunity, I simulated the distribution of the gains of a marginal increase in enrollment in higher tracks and better schools. Five cases are analyzed separately: allowing more students into the higher tracks at the expense of the vocational track, expanding the academic track with decreasing enrollment in the mixed track, and increasing the supply of better schools within each of the three tracks. In each case 2 percent of the total student population is assigned to the treatment, chosen random among the rejected applicants.

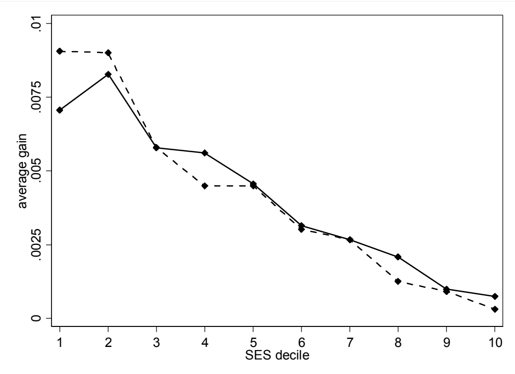
For drawing the 2 percent subsets two scenarios are considered. First, the affected students are chosen random from the rejected applicants. In a second scenario students are chosen with no respect to their actual applications, but with some restrictions with respect to prior achievement. Students assigned to a higher track are chosen from the top third of the prior score distribution of their actual track, in accordance with meritocratic selection among tracks. Students assigned to a better school within a track are chosen from the middle and bottom third of the prior achievement distribution of the track, assuming that there is less room for increasing the supply of better schools in the top third, as those students already attend the best schools of the track.

Counterfactual improvement in test scores is predicted using the estimated average treatment effects. Students assigned to a higher track or better school are assumed to experience a gain in grade 10 test scores equal to the average treatment effects on the treated from the difference model (Table 3). Finally, the average test score gain is calculated for the deciles of students with respect to family socio-economic status.

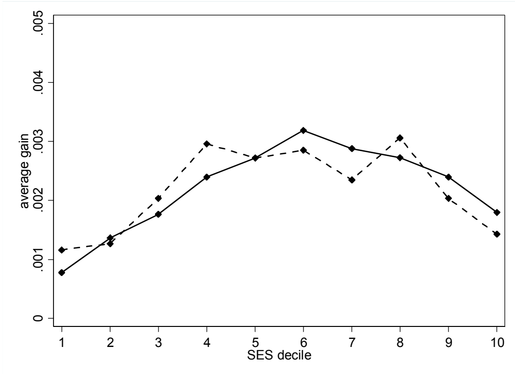
Figure 3 depicts the results of these simple back-of-the-envelope calculations about the effect of admitting additional marginal students into their preferred classes on inequality of opportunity. The results reveal that decreasing the share of the vocational track has the strongest positive impact on equality of opportunity. Students from poor families would benefit most, while the rich remains unaffected. The magnitude of the gain from a marginal change is of course modest in absolute terms, when compared to the overall level of inequality (see Figure 1). Nevertheless, it is much larger than the impact of other changes in enrollment shares.

Figure 3a

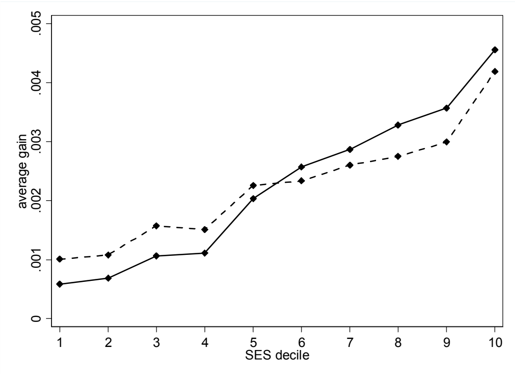
The estimated average gain in 10th grade test scores from a 2% expansion of higher tracks and better schools within tracks for deciles of family background, math



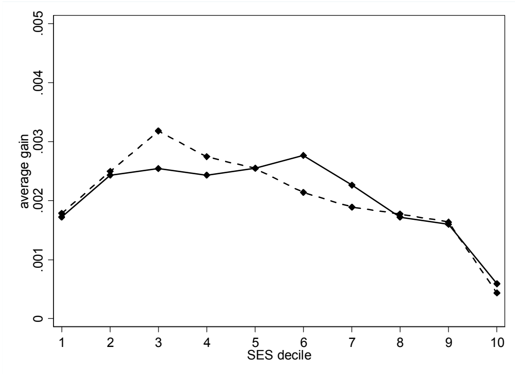
more mixed track, less vocational



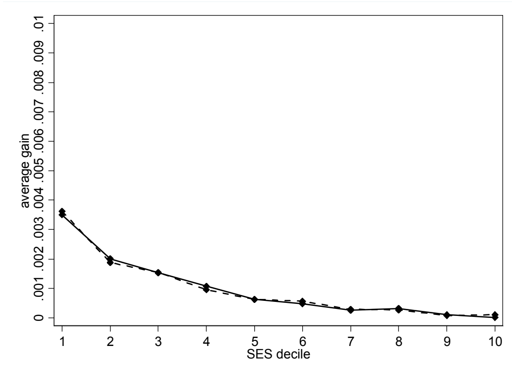
more academic track, less mixed



more better schools in the academic track



more better schools in the mixed track



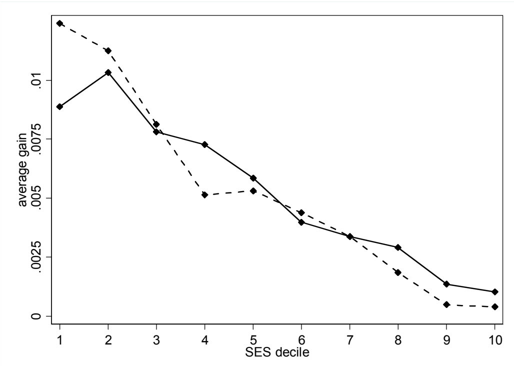
more better schools in the vocational track

——— : the extra 2% of admitted students chosen random from rejected applicants

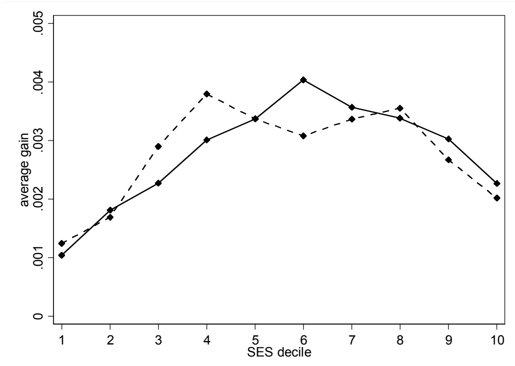
– – – : the extra 2% chosen random from the top third (higher track) or the bottom and middle third (within tracks)

Figure 3b

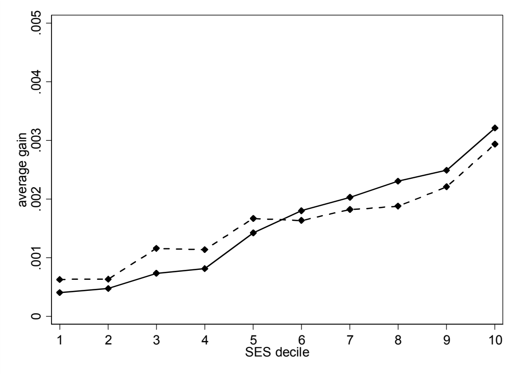
The estimated average gain in 10th grade test scores from a 2% expansion of higher tracks and better schools within tracks for deciles of family background, reading



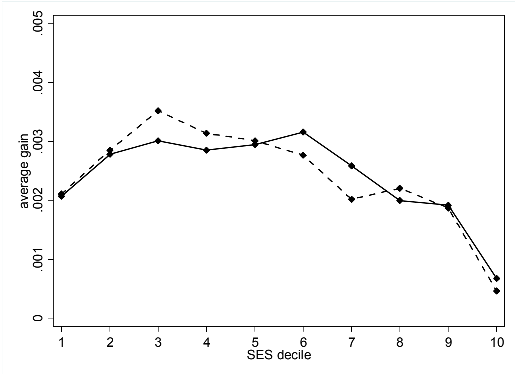
more mixed track, less vocational



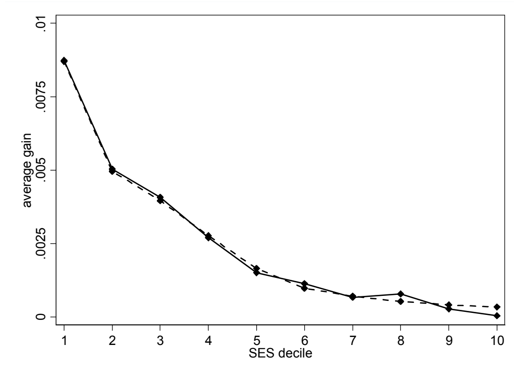
more academic track, less mixed



more better schools in the academic track



more better schools in the mixed track



more better schools in the vocational track

——— : the extra 2% of admitted students chosen random from rejected applicants

– – – : the extra 2% chosen random from the top third (higher track) or the bottom and middle third (within tracks)

Admitting more students to the academic track has a more ambiguous and weak impact. Students with a middle class background gain most, and the poor benefit the least. The impact of expanding better schools within tracks is mixed. Increasing enrollment in better schools in the academic track impedes, while in the vocational track it improves equality of opportunity. In the mixed track there is no impact, since the gain is almost evenly distributed with respect to social status, except the poorest and richest who are less affected. Altogether, if enrollment in better schools were expanded within all the three tracks at the same, equality of opportunity would remain unaffected.

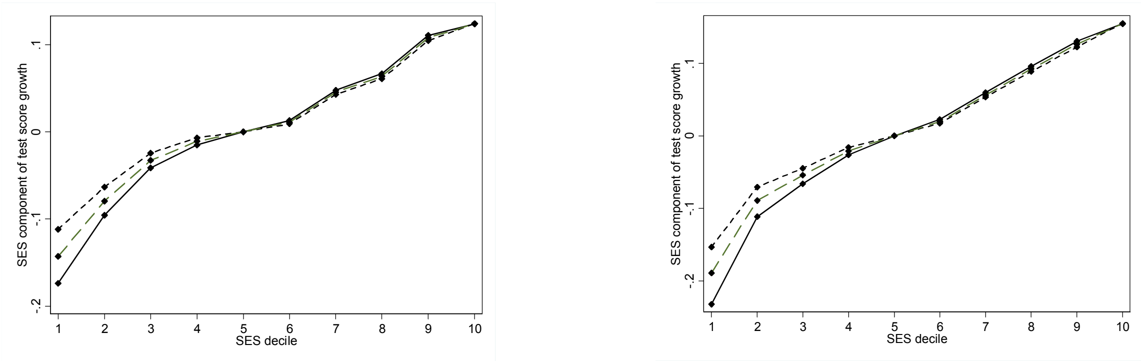
Finally, note that the two scenarios of selecting the students for the expanded enrollment provide almost identical results.

Overall, the simulated impact of marginal changes suggests, that tracking, and more specifically tracking at the low end is more closely related to inequality of opportunity than ability sorting within tracks. The reason for this is twofold. First, the estimated vocational track effect is larger than better school or academic track effects, and second, family background is much more different between the vocational and the mixed track than in the other cases. However, the improvement in equality of opportunity from a marginal change in enrollment shares is still small. This is not only due to the small change in the share of tracks, but also because within the vocational track students who applied to the mixed track (or those located in the top third of the prior achievement distribution in the second scenario) tend to have a more favorable family background than those who settle for vocational education when making their application decisions.

Beside the effect of marginal changes in enrollment the more general question is 'what would happen with no tracking at all?'. The large detrimental effect of the vocational track, together with a relative concentration of poor students in it, suggests that eliminating this track altogether could significantly improve equality of opportunity. Figure 4 compares the actual family background related component in test score growth between grade 8 and 10 to that simulated for no vocational track, assuming again that only vocational track students would be affected. If the gains in test scores are calculated solely from the estimated vocational track effect, improvement of equality of opportunity is non-negligible indeed (dashed line). However, the track effect would probably be replaced by within track ability sorting. When this is taken into account and the estimated gains are attenuated by within track better school effects (long dashed line), the improvement from de-tracking would be only about half of that derived from the vocational track effect solely.

Figure 4

The estimated SES component of test score growth from grade 8 to 10 with the vocational track eliminated



math reading

Coefficients of SES decile dummies from regressions of test score in grade 10, controlling for prior test scores and gender.

——— : actual

- - - : simulated; from vocational track effect solely

– – – : simulated; from vocational track effect, within mixed track effect subtracted

ROBUSTNESS AND SENSITIVITY ANALYSIS

A. ROBUSTNESS TO ALTERNATIVE EXPLANATIONS

A possible argument against the causal interpretation of the track and preferred class effects reported above could refer to the special effect of being rejected on student motivation. Being rejected can cause frustration and hence diminish student effort. This explanation is difficult to test directly. However, one would expect frustration to grow with the number of rejections: being admitted to the class ranked as a 4th or 6th choice is probably more frustrating than getting a place in the 2nd one. This hypothesis can be tested by including dummy variables in the regression models for the treatment effect representing the preference rank given by the student to her actual class. If a frustration effect were at work, enrollment in higher ranked classes should aggravate the negative effect of being rejected, i.e. the coefficients of higher ranked classes should be negative and increasing in size. Results in Table A8 of the Appendix do not appear to support the frustration hypothesis, as ending up in a lower ranked class does not consistently increase the magnitude of the negative treatment effect.

When comparing track and better school effects, some omitted school characteristics can bias the results. It is possible, that student preferences depend on other factors than academic quality, as well, and these preferences may play a larger role in within track choices than track placement. Two candidates are the geographical distance from the school and the owner of the school. Some students may rank higher a weaker school that is closer than a somewhat better one located farther away. Also, church or foundational schools may be preferred for non-academic reasons. However, it is less plausible that these factors would similarly affect track choice. In order to mitigate the bias due to non-academic elements in preferences, I repeat the regression estimates for the sample excluding applications when the preferred and actual class is located in distinct towns or the type of the owner (government, church, non-church private) is different. Then the estimation sample is further restricted to government-run schools, preferred and actual classes located in the same town. Results basically remain unchanged, suggesting that not these preferences make the within track effects smaller than the vocational track effect (Appendix Table A9).

B. SENSITIVITY TO MATCHING METHOD AND PARAMETERS

The sensitivity of the results is examined with respect to the matching method, the parameters of matching and the covariates used in the matching and for the bias correction.

Table A10 of the Appendix summarizes the estimation results with alternative matching methods for achievement level with no parametric bias correction and achievement growth. Three alternative methods are used. First, nearest neighbor matching with replacement is used instead of radius matching. Second the set of rejected applications is restricted to one application per student (that with the highest rank), as opposed to the baseline model, where all rejected applications of each student are included. Third, the set of control students is restricted to those with a more similar application profile than in the baseline estimation. E.g. a treated student who applied for a class in a mixed school but enrolled in a vocational class is matched to control students who managed to enroll in the mixed track but also applied to a vocational class. The first and second alternatives produce virtually identical results to the baseline method. In the third case the treatment effects for students enrolled in the vocational track are somewhat smaller compared to the baseline, but the results are qualitatively similar.

Table A11 exhibits the results for matching by other covariates; first by prior test score of the subject (math or reading), then by average marks and the two prior test scores together. Matching by the prior score provides similar treatment effects, which tend to exceed the baseline estimates, as expected, due the bias related to measurement error in test scores. In the second case control students are selected to have similar marks, prior math test scores and prior reading test scores at the same time to the rejected applicant. A matched pair is constructed if the distance between students is below the radius for each of the three covariates. Since this approach implies more stringent criteria for matching, it provides a lower number of observations with the same radius value. Compared to the baseline estimates the number of successful matches drops by about a factor of five, however, the estimated treatment effects remain similar in qualitative terms. The only exception is the effect of better schools within the vocational track for math, which falls to zero and insignificant.

Furthermore, the results appear to be not sensitive to the values of the matching parameters (the radius between 0.1 and 0.4) and the set of covariates used for the parametric bias correction. Figure A3-A7 of the Appendix depicts the estimated treatment effects, standard errors and the number of observations for different values of the matching parameters. The radius is allowed to take on values between 0.10 and 0.40 standard deviations, while the difference of test scores is judged both at the 5 and 10 percent significance levels. Overall, the results seem to be robust with respect to these parameters. Changing the radius causes a difference in the estimated treatment effect of at most 0.02 points compared to the baseline. The number of treated and control observations increases in the radius, while the standard error decreases. The significance level makes not much difference either.

Table A12 of the Appendix shows the effect of an alternative specification for the bias correction. In the baseline estimates only prior achievement is used in the bias correction regression. In the alternative specification family background variables; i.e. dummies for mother's and father's educational attainment and the number of books at home are also included. This modification does not have any considerable effect on the estimated treatment effects.

Altogether the results do not appear to be sensitive to the matching method, the set of covariates and the parameter values.

CONCLUSIONS

In this paper I estimated the causal impact of educational tracks and preferred schools of students within tracks on student achievement in upper-secondary education in Hungary. Identification relies on comparing similar students who applied to a given school but happened to be finally rejected with otherwise similar students who were enrolled in the same school. Average treatment effects on the treated were estimated using a matching method.

The results reveal that the higher tracks have a positive impact on student achievement. The negative impact of the vocational track on basic skills is considerable. It amounts to 0.21-0.28 standard deviation of test scores. The benefit provided by the academic over the mixed track is smaller, about half of the vocational track effect, but still substantial. Better schools within the tracks, i.e. those preferred by the applicants also seem to improve student achievement. The magnitude is similar to the academic track effect.

Raw comparison of the estimated effects can not reveal whether track effects are particular to a tracking regime or the tracks simply works as labels for schools of especially good or poor quality. In order to distinguish pure track effects I compared the estimated higher track and better school effects while controlling for the difference between the preferred and actual school of the rejected applicant in terms of peer quality and the distance of the rejected applicant from the peers of the preferred school. Although the track effects slightly diminish when these factors are taken into account, altogether the regression results are in line with the conclusions drawn from the average treatment effects. The outstanding negative effect of the vocational track persists.

Comparing track and better school effects provides a mixed picture. On the one hand, the positive impact of the academic track does not differ consistently from better school effects within tracks. This suggest that if the academic and mixed tracks were replaced by a single general track, the average achievement level and equality of opportunity would probably do not change, as school choice and the selection of students would reproduce the present stratified school system with similar outcomes. At the same time, the vocational track does have a specific negative effect, additional to that of less preferred schools in general. This implies that in some cases tracking may have more severe impact than school choice, especially if one of the tracks provides much more inferior opportunities and prospects for its students than the others. However, tracking does not necessarily matter when school choice is present, as results for the academic track suggests.

The available data do not allow for distinguishing among the mechanisms behind the track and better schools effects. Peer effects and better teachers, due to positive student-teacher matching are likely explanations for the advantage of better schools in general. These may be equally at work regarding educational tracks and better schools. Hence, the outstanding vocational track effect should be triggered by other factors, specific for this track. Differences in curricula and the lower standards required to successfully finish the school are natural candidates. These factors are not only track-specific, but differ from peer effects and teacher quality in another way, as well. If the share of good and bad teachers is given and peer effects are not strongly asymmetric, then admitting more students into the preferred tracks or schools may result in some redistribution, but will not raise average achievement. At the same time, eliminating or mitigating disadvantages of the vocational track by changing the curricula and the standards set for students or reducing its share in upper-secondary education may improve equality of opportunity and average achievement at the same time.

To assess implications for equality of opportunity I calculated the distribution of the gains of a *marginal* increase in enrollment in higher tracks and better schools within tracks over groups of students with different family background. The simulated distribution of gains shows that equality of opportunity can be improved mostly by shrinking the vocational track. Considering marginal changes in enrollment, the academic track and ability sorting within tracks harm equality much less, if at all.

The large negative effect of the vocational track, together with a relative concentration of poor students in it, suggests that eliminating this track altogether could indeed improve equality of opportunity to some extent. However, about half of the track effect eradicated in this case can be expected to be replaced by within track ability sorting. This would attenuate the possible gains of de-tracking.

An important question for policy is whether the better school or track effects are specific for a particular group of students, being on the margin or can apply for others, as well. In other words, can the benefits of higher tracks or better schools be easily extended to other students? First note here that rejected students are not restricted to be a special group in terms of prior achievement located close to a sharp cut-off, but represent the greater part of the achievement distribution. At the same time, applicants far below the peer mean of a class appear to benefit little when admitted to a better school, suggesting that the gains of expanding higher tracks might be decreasing.

However, regarding their unobserved characteristics like motivation and self-confidence, the group of rejected applicants can be assumed to be quite different from those students who settled with less ambitious options. Hence the treatment effects for the treated group can not be simply assumed to hold for everyone. On the other hand, non-cognitive skills are more prone to be improved in schools compared to cognitive skills, as evidence on remediation programs targeting disadvantaged adolescents suggests (see Brunello – Schlotter, 2011 and Cunha et al., 2006 for a review). This implies that there is some room for education policy to make even less motivated, ambitious or confident students benefit from higher tracks and better schools.

It is important to note though, that the estimated effects are probably can not be replicated indefinitely by extending enrollment in higher tracks or the schools preferred by most students. Admitting more and more weak students would affect directly or indirectly the school quality faced by every student, via peer group effects and the access to the limited set of good teachers, and would trigger complex behavioral responses.

Finally note that the results above do not necessarily holds for every tracking and school choice regime, of course. A crucial feature of upper-secondary education in Hungary is the strong merit-based selection of students into schools. This is not an indispensible component of school choice, and even tracking regimes may differ in this respect (see e.g. Checchi-Flabbi, 2007 on differences in selection into tracks between Italy and Germany).

REFERENCES

Abadie, A. - D. Drukker - H. Herr - G. Imbens (2004): Implementing Matching Estimators for Average Treatment Effects in STATA, The Stata Journal (2004) 4, Number 3, pp. 290–311

Abadie, A. - G. Imbens (2002): Simple and Bias-Corrected Matching Estimators for Average Treatment Effects, NBER technical working paper no. 283

Abdulkadiroglu, A. – J. D. Angrist – P. A. Pathak (2011): The elite illusion Achievement effects at Boston and New York exam schools, NBER wp17264

Ammermueller, A. (2005): Educational opportunities and the role of institutions, ZEW DP No. 05–44.

Argys, L. M. - Rees, D. I. - Brewer, D. J. (1996): Detracking America's schools: equity at zero cost? Journal of Policy Analysis and Management, 15(4), 623–645.

Betts, J. R. (2011): The Economics of Tracking in Education, in: E. A. Hanushek – S. Machin – L. Woessmann (eds.): Handbook of the Economics of Education, Vol. 3, Elsevier, Pages 341–381.

Betts, J. R. – J. Grogger, 2003): The impact of grading standards on student achievement, educational attainment, and entry-level earnings, Economics of Education Review 22 (2003) 343–352.

Betts, J. R. – J. L. Shkolnik (2000): The effects of ability-grouping on student math achievement and resource allocation in secondary schools, Economics of Education Review, vol. 19(1), pp. 1–15.

Bonesronning, H. (2008): The Effect of Grading Practices on Gender Differences in Academic Performance, Bulletin of Economic Research, Vol. 60, Issue 3, pp. 245-264

Brunello, G. – D. Checchi (2007): Does school tracking affect equality of opportunity? New international evidence, Economic Policy October 2007 pp. 781–861

Brunello, G. – M. Schlotter (2011): Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems, IZA DP 5743

Checchi, D. – L. Flabbi (2007): Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks, IZA Dp. 2876

Clark, D. (2010): Selective Schools and Academic Achievement, The B.E. Journal of Economic Analysis & Policy, Vol. 10/1.

Clotfelter, C.T. – H.F. Ladd – J.L. Vigdor (2005): Who Teaches Whom? Race and the Distribution of Novice Teachers, Economics of Education Review 24 (2005) 377–392

Clotfelter, C. T. – Ladd, H. F. – Vigdor, J. L. (2006): Teacher-student matching and the assessment of teacher effectiveness. Journal of Human Resources, XLI(4), 778–820

Christofides, L. N. – M. Hoy – J. Milla – T. Stengos (2012): Grades, Aspirations and Post-Secondary Education Outcomes, IZA Dp 6867.

Cunha, F. - J. J. Heckman - L. J. Lochner - D. Masterov (2006): Interpreting the Evidence on Life Cycle Skill Formation, in E. Hanushek and F. Welch, eds., The Handbook of Economics of Education, 697–812, North Holland, Amsterdam

Cullen, J. B., B. A. Jacob, and S. Levitt (2006): The effect of school choice on participants: Evidence from randomized lotteries, Econometrica74 (5): 1191– 1230.

Dee, T. S. (2005): A Teacher like Me: Does Race, Ethnicity, or Gender Matter?, The American Economic Review, Vol. 95, No. 2, pp. 158-165.

Dobbie, W. – R. G. Fryer (2011): Exam High Schools and Academic Achievement: Evidence from New York City

Duflo, E. – P. Dupas – M. Kremer (2011): Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, American Economic Review 101: 1739–1774

Dustmann, C. – P. A. Puhani – U. Schönberg (2012): The Long-term E ects of School Quality on Labor Market Outcomes and Educational Attainment, CReAM Discussion Paper No 08/12

Epple, D. – R. E. Romano (1998): Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects, Ameracan Economic Review, 62 (1), 33–62.

Epple, D. – E. Newlon – R. Romano (2002): Ability tracking, school competition, and the distribution of educational benefits, Journal of Public Economics, 83 1-48.

Figlio, D. N. – M. E. Lucas (2004): Do high grading standards affect student performance? Journal of Public Economics 88 (2004) 1815– 1834

Figlio, D.N. - Page, M.E. (2002): School choice and the distributional effects of achievement tracking Does separation increase inequality?, Journal of Urban Economics, vol. 51(3), pp. 497–514.

Imbens, G. W. (2004): Nonparametric estimation of average treatment effects under exogeneity A review, The Review of Economics and Statistics, February 2004, 86(1): 4–29

Hanushek, E. A. – J. F. Kain – J. M. Markman – S. G. Rivkin (2001): Does peer ability affect student achievement? NBER wp 8502.

Hastings, J. S. – C. A. Neilson – S. D. Zimmerman (2012): The effect of school choice on intrinsic motivation and academic outcomes, NBER wp18324

Horn, D. (2013): Diverging Performances: the detrimental effects of early selection on equality of opportunity in Hungary, Research in Social Stratification and Mobility, vol 32. pp. 25-43.

Hoxby, C. (2000): Peer effects in the classroom: Learning from gender and race variation, NBER wp 7867.

Guyon, N. – E. Maurin – S. McNally (2010): The effect of tracking students by ability into different schools A natural lexperiment, Fondazione Eni Enrico Mattei, Nota di Lavoro 152.2010

Jackson, C. K. (2010): Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago, The Economic Journal, 120, 1399–1429.

de Janvry, A. de – A. Dustan – E. Sadoulet (2012): The benefits and hazards of elite high school admission Academic opportunity and dropout risk in Mexico City

Jenkins, Stephen P. - John Micklewright - Sylke V. Schnepf (2008): Social Segregation in Secondary Schools: How Does England Compare with Other Countries? Oxford Review of Education, Vol. 34/1, 21-37.

Kóczy Á. László (2010): A magyarországi felvételi rendszerek sajátosságai, Közgazdasági Szemle, LVII.(2) 142—164.

Ladd, H. F. (2002): Market-Based Reforms in Urban Education, Economic Policy Institute, Washington

Lankford, H. - Loeb, S. - Wyckoff, J. (2002): Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis, Educational Evaluation and Policy Analysis, v24 n1 p37-62

Lavy, V. – O. Silva – F. Weinhardt (2009): The Good, The Bad and The Average: Evidence on the Scale and Nature of Ability Peer Effects in School, NBER WP15600

Leahey E. - Guo G. (2001): Gender differences in mathematical trajectories, Social Forces 80:713–732.

MacLeod, W. B. – M. Urquiola (2012): Anti-lemons: School reputation, relative diversity, and educational quality, IZA Dp 6805

Malamud, O. – C. Pop-Eleches (2010): General Education versus Vocational Training: Evidence from an Economy in Transition, Review of Economics and Statistics 92(1): 43–60.

Manning, Alan, and Jorn-Steffen Pischke (2006): Comprehensive versus Selective Schooling in England and Wales: What Do We Know? IZA Discussion Paper No. 2072.

Meghir, C. – M. Palme (2005): Educational Reform, Ability, and Family Background, American Economic Review 95(1): 414–424.

Moen, J. – M. Tjelta (2010): Grading Standards, Student Ability and Errors in College Admission, Scandinavian Journal of Educational Research, 54/3, 221-237.

Pekkarinen, T. – R. Uusitalo – S. Kerr (2009): School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform, Journal of Public Economics 93(7–8): 965–973.

Pop-Eleches, C. – M. Urquiola (2011): Going to a Better School: Effects and Behavioral Responses, NBER wp16886

Schuetz, G. – H. W. Ursprung – L. Woessmann (2008): Education Policy and Equality of Opportunity, Kyklos 61(2): 279-308.

Todd, P. E. – K. I. Wolpin (2003): On the specification and estimation of the production function for cognitive achievement, The Economic Journal, 113(February), F3–F33.

Varga J. (2009): A tanárok elosztása a különböző szociokulturális hátterű tanulókat tanító iskolák között, Fazekas K. (szerk.): Oktatás és foglalkoztatás, MTA Közgazdaságtudományi Intézet, Budapest, 65-82.

Waldinger, F. (2006): Does tracking affect the importance of family background on students'test scores?, mimeo

APPENDIX

A1. DESCRIPTIVE STATISTICS FOR THE SAMPLE

Table A1

Summary statistics of student achievement by educational track

		Grade 8			Grade 10	
		math	reading	marks score	math	reading
Academic track	mean	0.530	0.626	0.510	0.407	0.576
	s.d.	0.91	0.83	0.55	0.85	0.74
	N	12365	12430	12431	12424	12427
Mixed track	mean	-0.030	-0.018	-0.034	-0.098	-0.029
	s.d.	0.81	0.79	0.53	0.78	0.76
	N	15149	15320	15321	15317	15316
Vocational track	mean	-0.858	-0.870	-0.748	-0.966	-1.031
	s.d.	0.67	0.71	0.42	0.74	0.75
	N	5924	6276	6278	6273	6274
Total	mean	0.031	0.060	0.033	-0.074	0.007
	s.d.	0.96	0.95	0.68	0.93	0.94
	N	33438	34026	34030	34014	34017

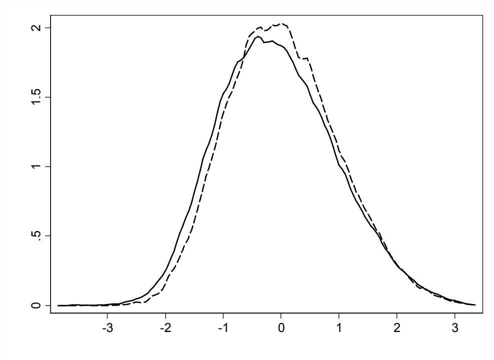
Table A2

Within and between track, school and class variation of grade 8 student achievement

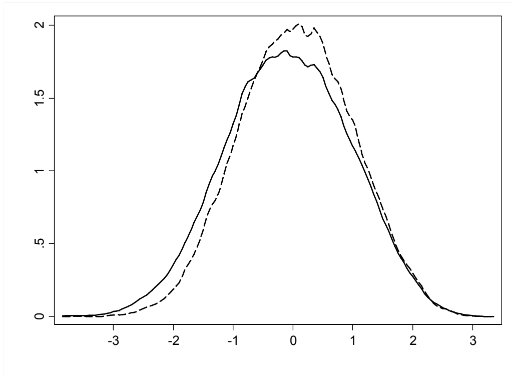
	grade 8		grade 9		
	school	class	track	School x track	class
math					
share of between variation	0.204	0.325	0.257	0.464	0.544
share of within variation	0.795	0.675	0.743	0.536	0.456
N of groups	1460	2910	3	1478	4166
N of students	33478	33478	33478	33478	33478
reading					
share of between variation	0.181	0.286	0.311	0.472	0.536
share of within variation	0.819	0.714	0.689	0.528	0.464
N of groups	1465	2942	3	1486	4194
N of students	34080	34080	34080	34080	34080

Figure A1

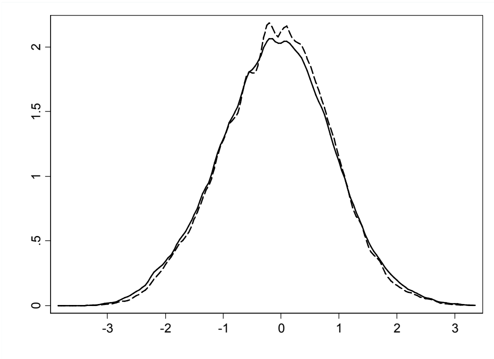
The distribution of test scores in the student population and the sample



math, grade 8

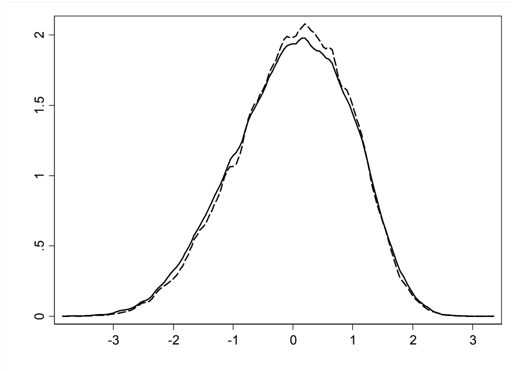


reading, grade 8



math, grade 10

- - : sample, ——— : population



reading, grade 10

A2. DESCRIPTIVE STATISTICS: MATCHED REJECTED APPLICANTS

Table A3

The number and share of matched observations in the treatment group

	rejected applications			rejected students		
track of preferred - actual class	N	N matched	r matched	N	N matched	r matched
mixed – vocational	2475	1255	0,51	1482	890	0,60
academic – mixed	2970	1245	0,42	1693	887	0,52
academic – academic	5408	2602	0,48	2895	1782	0,62
mixed - mixed	4855	2628	0,54	3019	1930	0,64
vocational – vocational	1556	988	0,63	1103	779	0,71
total	17264	8718	0,50	9453	5971	0,63

Table A4

Rejected applicants in the prior test score distribution of the preferred class

	math			reading		
track of preferred - actual class	low	medium	top	low	medium	top
mixed – vocational	0.538	0.305	0.158	0.447	0.348	0.204
academic – mixed	0.528	0.296	0.176	0.521	0.296	0.182
academic – academic	0.494	0.325	0.181	0.477	0.327	0.196
mixed - mixed	0.552	0.308	0.140	0.531	0.321	0.148
vocational – vocational	0.426	0.359	0.215	0.421	0.360	0.219

Table A5

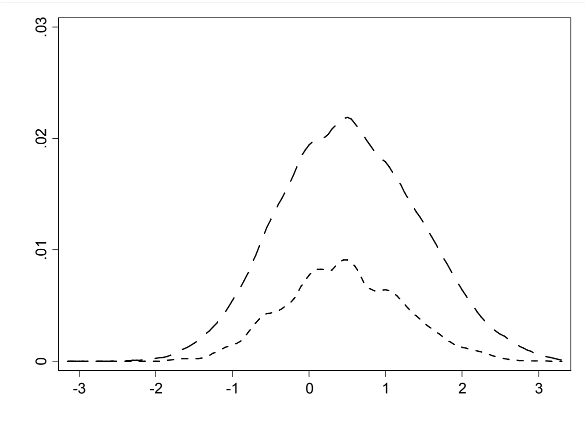
Average prior achievement and family background characteristics of rejected applicants and control students

	Prior student achievement					
	math		reading		marks	
	treated	control	treated	control	treated	control
track of preferred - actual class						
mixed – vocational	-0.56	-0.48	-0.47	-0.40	-0.47	-0.42
academic – mixed	0.20	0.30	0.35	0.44	0.31	0.35
academic – academic	0.46	0.57	0.64	0.71	0.55	0.58
mixed - mixed	-0.15	-0.08	-0.11	-0.06	-0.13	-0.09
vocational – vocational	-0.95	-0.91	-0.92	-0.87	-0.79	-0.77
	Family background					
	books at home		mother's years of educ		father's years of educ	
	treated	control	treated	control	treated	control
track of preferred - actual class						
mixed – vocational	3.19	3.28	10.44	10.79	10.32	10.64
academic – mixed	4.11	4.32	12.12	12.57	11.62	12.07
academic – academic	4.53	4.64	12.91	13.10	12.47	12.65
mixed - mixed	3.61	3.68	11.25	11.47	10.93	11.15
vocational – vocational	2.61	2.74	9.70	9.91	9.79	9.92

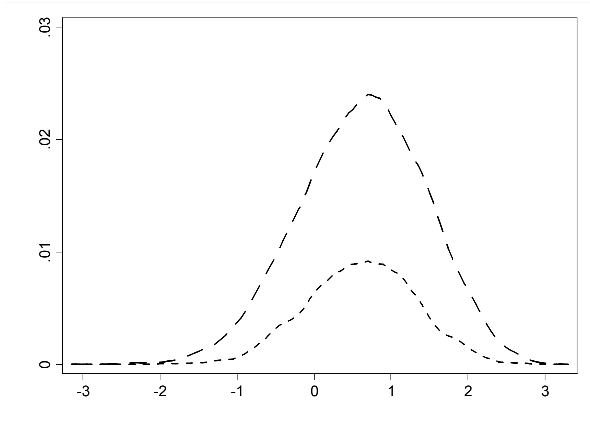
Average values for rejected applications. Books at home: number of books at home, measured in 7 categories (from 1 to 7).

Figure A2

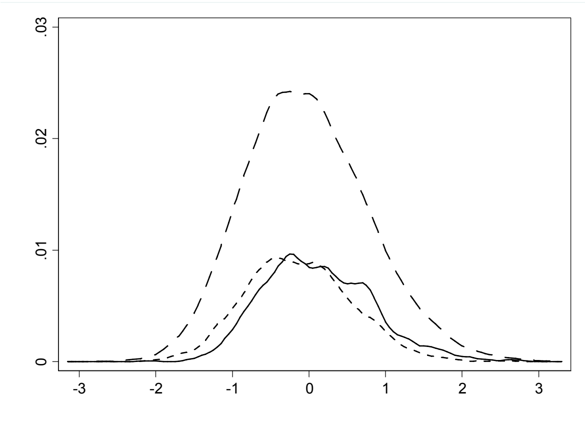
The distribution of test scores in grade 8 for the sample population of students and for rejected applicants by educational track



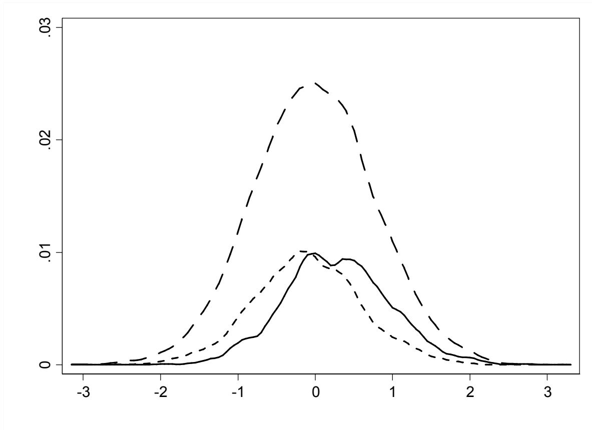
math, academic track



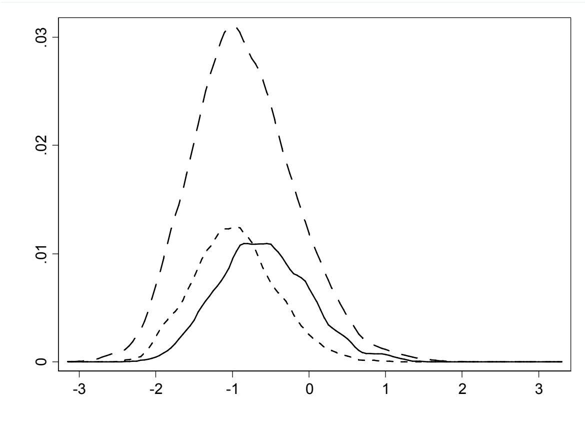
reading, academic track



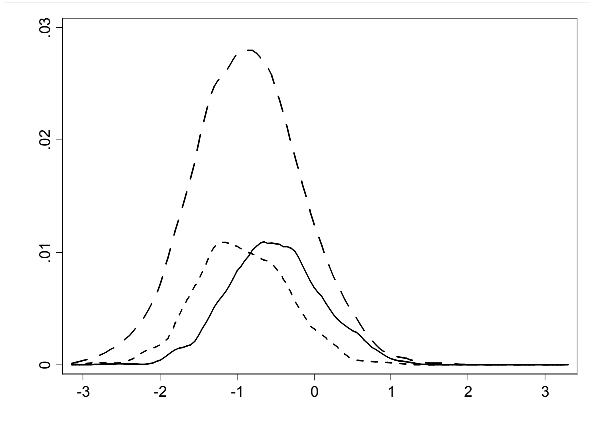
math, mixed track



reading, mixed track



math, vocational track



reading, vocational track

– – – : total student population of the track
- - - - : rejected applicants, preferred class in the same track
——— : rejected applicants, preferred class in a higher track
The distribution of rejected applicants is shrunk by a factor of three.

A3. REGRESSION ESTIMATES OF THE EFFECT OF BEING REJECTED

Table A7

Testing the differences in the effect of being rejected

	math			reading		
	(1)	(2)	(3)	(1)	(2)	(3)
track effects vs within track effects						
mix/voc - within any	-11.208**	-9.053***	-9.455***	-13.312**	-11.220**	-11.125**
aca/mix - within any	-1.586	-0.281	-0.875	-4.428*	-2.837	-2.780
pairwise comparison of coefficients						
mix/voc - aca/mix	-9.622***	-8.772***	-8.581***	-8.884***	-8.383***	-8.346***
mix/voc - aca/aca	-9.826***	-8.354***	-9.130***	-16.753**	-15.261**	-14.837**
mix/voc - mix/mix	-8.670***	-6.379***	-6.485***	-12.073**	-9.787***	-9.493***
mix/voc - voc/voc	-15.129**	-12.425**	-12.751**	-11.109**	-8.611**	-9.046***
aca/mix - aca/aca	-0.204	0.418	-0.549	-7.869***	-6.878***	-6.492***
aca/mix - mix/mix	0.952	2.393	2.096	-3.189	-1.404	-1.147
aca/mix - voc/voc	-5.507*	-3.653	-4.170	-2.225	-0.228	-0.701
aca/aca - mix/mix	1.156	1.975	2.645	4.680**	5.474**	5.344**
aca/aca - voc/voc	-5.302*	-4.071	-3.621	5.644**	6.650**	5.791*
mix/mix - voc/voc	-6.458**	-6.046**	-6.266**	0.964	1.176	0.447

Differences in the combination of estimated coefficients of Table 4.
F-tests: *** p<0.01, ** p<0.05, * p<0.1

Table A8

Regression estimates of the effect of the rank of the actual class in rejected student's preference list on the effect of being rejected from the preferred class

	math			reading		
	(1)	(2)	(3)	(1)	(2)	(3)
the rank of the actual class in students' preference list (ref.: 2nd)						
3rd	0.000	0.004	0.002	0.009	0.013	0.013
	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
4th	-0.017	-0.010	-0.016	0.001	-0.001	-0.001
	(0.019)	(0.019)	(0.019)	(0.021)	(0.021)	(0.021)
5th	0.019	0.025	0.020	0.069***	0.074***	0.074***
	(0.026)	(0.025)	(0.026)	(0.027)	(0.027)	(0.027)
6th	0.022	0.027	0.020	-0.008	0.018	0.019
	(0.028)	(0.027)	(0.028)	(0.032)	(0.032)	(0.031)
7th or higher	0.075**	0.078***	0.071**	0.014	0.013	0.013
	(0.030)	(0.030)	(0.030)	(0.032)	(0.032)	(0.032)

Standard errors clustered for preferred classes of rejected applicants are in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Control variables as in Table 4.

Table A6

Regression estimates for the effect of being rejected from the preferred class, by track of preferred and actual class

track of preferred - actual class:	mixed	vocational	academic	mixed	academic	academic	mixed	mixed	vocational	vocational
	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)
math										
difference in peer mean	-0.1347**	-0.1595**	-0.1477***	-0.1342***	-0.1464***	-0.1382***	-0.1523***	-0.1569***	0.0354	-0.0073
	(0.0682)	(0.0674)	(0.0423)	(0.0427)	(0.0309)	(0.0308)	(0.0385)	(0.0372)	(0.0749)	(0.0756)
distance from peer mean in preferred class	0.1670***	0.1628***	0.2539***	0.2537***	0.2140***	0.2125***	0.1623***	0.1716***	0.1924***	0.1847***
	(0.0421)	(0.0428)	(0.0367)	(0.0359)	(0.0326)	(0.0328)	(0.0271)	(0.0268)	(0.0595)	(0.0582)
distance from preferred class squared	0.0003	0.0003	-0.0006	-0.0006*	-0.0006**	-0.0006**	-0.0004	-0.0004	0.0010	0.0009
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0007)	(0.0006)
gender: female	-4.0407	-4.0026	-7.0475	-6.3577	-5.0253	-5.1130	-2.4936	-1.8843	-4.0838	-3.6892
	(4.0758)	(4.0385)	(4.2883)	(4.2573)	(3.1849)	(3.1948)	(2.9360)	(2.9131)	(5.0707)	(5.1309)
micro-region fixed-effects	no	yes	no	yes	no	yes	no	yes	no	yes
constant	-18.1266***	-17.3952***	-9.1580***	-9.8896***	-9.5215***	-9.2105***	-11.7153***	-11.7631***	-1.5985	-2.6267
	(2.5098)	(2.4919)	(1.8995)	(1.8113)	(1.4970)	(1.4824)	(1.4559)	(1.3909)	(2.5492)	(2.5535)
N	1,180	1,180	1,217	1,217	2,532	2,532	2,504	2,504	886	886
R2	0.0256	0.0252	0.0424	0.0416	0.0265	0.0259	0.0212	0.0234	0.0356	0.0318
reading										
difference in peer mean	-0.2857***	-0.3344***	-0.0970*	-0.0637	-0.0901**	-0.1022***	-0.1231***	-0.1182***	-0.2292***	-0.2434***
	(0.0795)	(0.0738)	(0.0575)	(0.0583)	(0.0370)	(0.0365)	(0.0439)	(0.0445)	(0.0760)	(0.0729)
distance from peer mean in preferred class	0.1893***	0.1957***	0.2041***	0.1943***	0.2355***	0.2385***	0.1793***	0.1816***	0.1186***	0.1310***
	(0.0445)	(0.0436)	(0.0368)	(0.0367)	(0.0258)	(0.0260)	(0.0268)	(0.0271)	(0.0395)	(0.0393)
distance from preferred class squared	-0.0002	-0.0004	-0.0005	-0.0005	-0.0002	-0.0002	-0.0002	-0.0001	0.0003	0.0002
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
gender: female	-2.8177	-1.8402	-3.9472	-1.4673	-0.2823	0.5020	1.5201	3.2241	2.9067	2.2090
	(4.7267)	(4.6545)	(4.0291)	(4.0109)	(3.3219)	(3.3226)	(3.0988)	(3.1159)	(5.3277)	(5.5603)
micro-region fixed-effects	no	yes	no	yes	no	yes	no	yes	no	yes
constant	-17.1855***	-14.9851***	-14.1737***	-14.9449***	-6.9227***	-7.6286***	-12.1693***	-12.2135***	-13.9510***	-13.0151***
	(3.4142)	(3.1206)	(2.3359)	(2.2525)	(1.5810)	(1.5329)	(1.5765)	(1.5729)	(2.3126)	(2.3407)
N	1,218	1,218	1,218	1,218	2,567	2,567	2,584	2,584	955	955
R2	0.0256	0.0295	0.0271	0.0237	0.0376	0.0379	0.0230	0.0241	0.0172	0.0192

Standard errors clustered for preferred classes are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A9

Regression estimates for the effect of being rejected from the preferred class, sample restricted to preferred and actual classes in the same town and with the same type of owner

	math public and private schools			math public schools only			reading public and private schools			reading public schools only		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
track of preferred - actual class												
mixed – vocational	-0.199***	-0.184***	-0.186***	-0.199***	-0.184***	-0.190***	-0.235***	-0.213***	-0.211***	-0.217***	-0.194***	-0.191***
	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)	(0.025)	(0.026)	(0.027)	(0.026)	(0.027)	(0.027)
academic – mixed	-0.124***	-0.114***	-0.117***	-0.127***	-0.117***	-0.117***	-0.151***	-0.135***	-0.133***	-0.143***	-0.127***	-0.123***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.023)	(0.024)	(0.025)	(0.023)	(0.024)	(0.025)
academic – academic	-0.096***	-0.091***	-0.083***	-0.091***	-0.087***	-0.084***	-0.063***	-0.059***	-0.058***	-0.062***	-0.058***	-0.059***
	(0.018)	(0.018)	(0.019)	(0.019)	(0.018)	(0.020)	(0.018)	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)
mixed / mixed	-0.102***	-0.105***	-0.108***	-0.102***	-0.104***	-0.105***	-0.114***	-0.115***	-0.117***	-0.112***	-0.113***	-0.114***
	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)	(0.016)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
vocational – vocational	-0.017	-0.026	-0.033	-0.023	-0.030	-0.032	-0.111***	-0.120***	-0.114***	-0.106***	-0.116***	-0.110***
	(0.032)	(0.032)	(0.033)	(0.031)	(0.032)	(0.033)	(0.034)	(0.034)	(0.035)	(0.034)	(0.035)	(0.036)
N (rejected applications)	5,908	5,908	5,908	5,724	5,724	5,724	6,080	6,080	6,080	5,898	5,898	5,898
R2	0.039	0.064	0.065	0.039	0.066	0.067	0.042	0.070	0.068	0.038	0.065	0.063

Standard errors clustered for preferred classes are in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Control variables as in Table 4.

A4. SENSITIVITY ANALYSIS

Table A10

Estimated effect of the preferred track and the preferred school on student achievement: sensitivity for the matching method

	math, level (no bias corr.)		math, diff				reading, level (no bias corr.)		reading, diff			
track of preferred - actual class	ATT	se	ATT	se	N treated	N control	ATT	se	ATT	se	N treated	N control
baseline												
mixed – vocational	-0.288	0.020	-0.210	0.022	1203	1623	-0.337	0.022	-0.266	0.025	1255	1706
academic – mixed	-0.215	0.019	-0.112	0.021	1234	1534	-0.227	0.020	-0.139	0.023	1245	1557
within track	-0.178	0.009	-0.097	0.011	6050	7047	-0.160	0.010	-0.101	0.011	6216	7270
academic – academic	-0.218	0.015	-0.113	0.017	2161	2447	-0.157	0.016	-0.078	0.018	2182	2465
mixed - mixed	-0.172	0.014	-0.105	0.016	2568	3275	-0.169	0.015	-0.120	0.017	2627	3344
vocational – vocational	-0.096	0.024	-0.051*	0.028	911	1207	-0.176	0.027	-0.126	0.029	987	1344
nearest neighbour												
mixed – vocational	-0.286	0.022	-0.206	0.024	1198	935	-0.341	0.024	-0.271	0.027	1249	972
academic – mixed	-0.211	0.021	-0.113	0.023	1225	954	-0.227	0.021	-0.131	0.025	1236	964
within track	-0.169	0.010	-0.090	0.011	5974	4171	-0.156	0.011	-0.103	0.012	6140	4275
academic – academic	-0.209	0.016	-0.104	0.018	2125	1525	-0.159	0.017	-0.083	0.019	2146	1538
mixed - mixed	-0.160	0.015	-0.100	0.017	2537	1836	-0.166	0.016	-0.122	0.018	2596	1872
vocational – vocational	-0.107	0.027	-0.049+	0.031	907	669	-0.171	0.029	-0.133	0.032	983	720
treated: one appl. per student only												
mixed – vocational	-0.296	0.027	-0.216	0.030	643	1008	-0.319	0.029	-0.259	0.033	665	1064
academic – mixed	-0.204	0.026	-0.118	0.027	610	902	-0.242	0.027	-0.168	0.031	615	916
within track	-0.166	0.011	-0.092	0.013	3513	5126	-0.154	0.012	-0.098	0.013	3620	5304
academic – academic	-0.210	0.019	-0.111	0.020	1226	1729	-0.143	0.020	-0.062	0.022	1240	1741
mixed - mixed	-0.165	0.016	-0.106	0.019	1507	2373	-0.165	0.018	-0.129	0.020	1543	2437
vocational – vocational	-0.088	0.028	-0.038+	0.033	583	919	-0.184	0.031	-0.122	0.034	634	1023
control: lower track applicants only												
mixed – vocational	-0.225	0.029	-0.171	0.035	619	586	-0.250	0.032	-0.212	0.038	652	619
academic – mixed	-0.184	0.024	-0.097	0.027	871	866	-0.218	0.024	-0.135	0.027	881	882

*: significant at 10%, +: not statistically significant, all other treatment effects significant at the 1% level

Table A11

Estimated effect of the preferred track and the preferred school on student achievement: sensitivity for the matching covariates

	math, level (no bias corr.)		math, diff				reading, level (no bias corr.)		reading, diff			
track of preferred - actual class	ATT	se	ATT	se	N treated	N control	ATT	se	ATT	se	N treated	N control
baseline												
mixed – vocational	-0.288	0.020	-0.210	0.022	1203	1623	-0.337	0.022	-0.266	0.025	1255	1706
academic – mixed	-0.215	0.019	-0.112	0.021	1234	1534	-0.227	0.020	-0.139	0.023	1245	1557
within track	-0.178	0.009	-0.097	0.011	6050	7047	-0.160	0.010	-0.101	0.011	6216	7270
academic – academic	-0.218	0.015	-0.113	0.017	2161	2447	-0.157	0.016	-0.078	0.018	2182	2465
mixed - mixed	-0.172	0.014	-0.105	0.016	2568	3275	-0.169	0.015	-0.120	0.017	2627	3344
vocational – vocational	-0.096	0.024	-0.051*	0.028	911	1207	-0.176	0.027	-0.126	0.029	987	1344
test score												
mixed – vocational	-0.276	0.014	-0.257	0.018	1688	2744	-0.356	.016	-0.340	0.020	1738	2749
academic – mixed	-0.196	0.014	-0.178	0.017	1693	2183	-0.227	.015	-0.209	0.018	1786	2299
within track	-0.154	0.007	-0.140	0.009	7552	9233	-0.192	.008	-0.178	0.009	7872	9458
academic – academic	-0.172	0.012	-0.160	0.015	2618	3029	-0.191	.013	-0.177	0.015	2804	3161
mixed - mixed	-0.152	0.010	-0.136	0.013	3343	4643	-0.200	.011	-0.186	0.014	3399	4649
vocational – vocational	-0.134	0.020	-0.124	0.025	1062	1489	-0.210	.022	-0.195	0.027	1100	1492
marks score, math score, reading score												
mixed – vocational	-0.253	0.034	-0.243	0.045	267	290	-0.258	0.039	-0.248	0.050	300	327
academic – mixed	-0.107	0.035	-0.096	0.049	209	214	-0.163	0.038	-0.165	0.046	215	223
within track	-0.110	0.015	-0.107	0.020	1336	1378	-0.130	0.016	-0.127	0.020	1441	1499
academic – academic	-0.155	0.027	-0.155	0.034	404	411	-0.092	0.027	-0.094	0.034	412	421
mixed - mixed	-0.123	0.022	-0.117	0.029	616	657	-0.155	0.024	-0.156	0.029	652	700
vocational – vocational	-0.009+	0.039	-0.003+	0.053	219	239	-0.168	0.042	-0.145	0.047	276	305

*: significant at 10%, +: not statistically significant, all other treatment effects significant at the 1% level

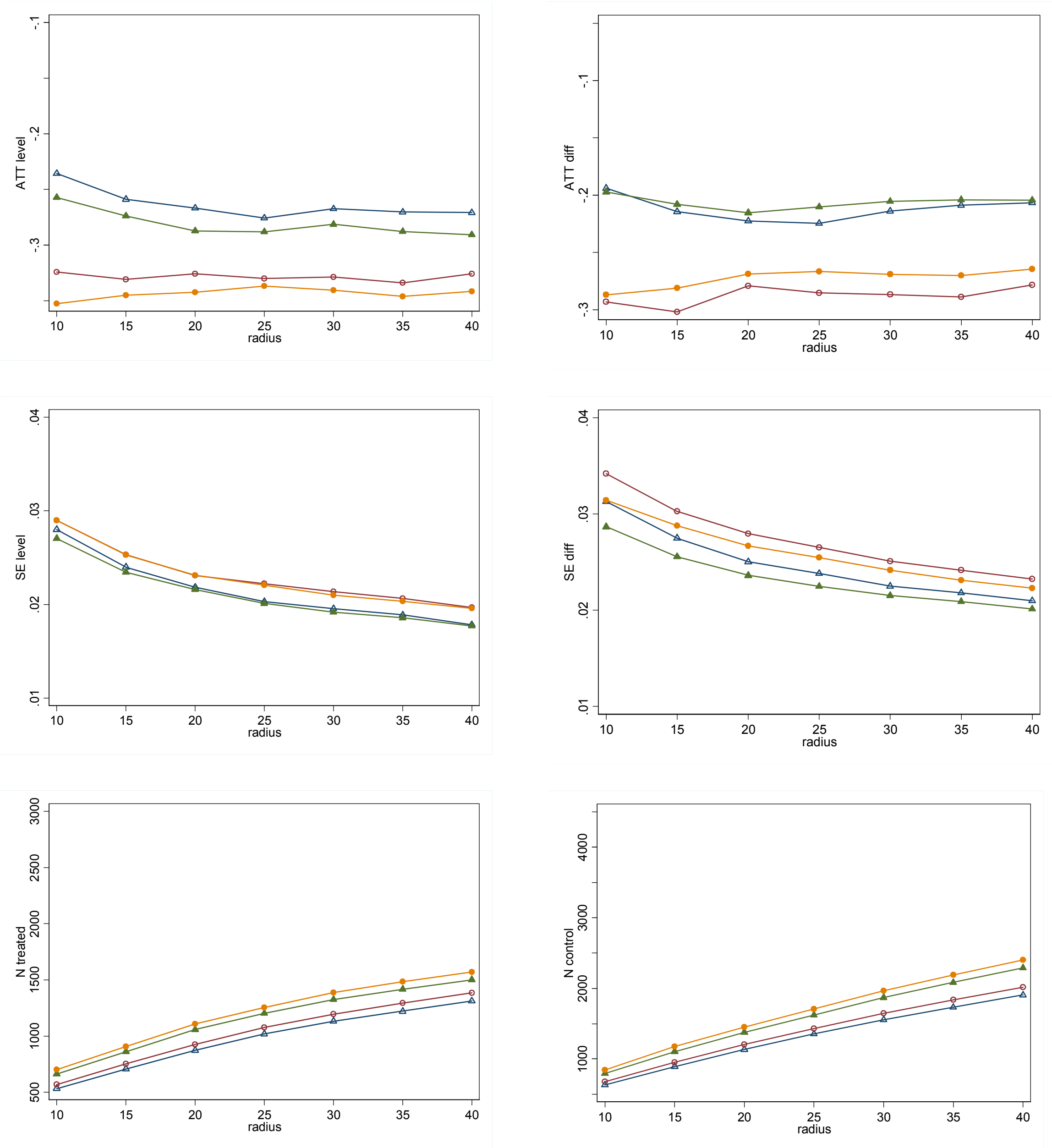
Table A12

Sensitivity for the specification of bias correction

track of preferred - actual class	math ATT	math se	math N treated	math N control	reading ATT	reading se	reading N treated	reading N control
baseline: bias correction based on prior achievement only								
mixed – vocational	-0.262	0.017	1203	1623	-0.285	0.019	1255	1706
academic – mixed	-0.158	0.016	1234	1534	-0.171	0.017	1245	1557
within track	-0.148	0.008	6050	7047	-0.123	0.009	6216	7270
academic – academic	-0.166	0.013	2161	2447	-0.111	0.014	2182	2465
mixed - mixed	-0.149	0.012	2568	3275	-0.138	0.013	2627	3344
vocational – vocational	-0.115	0.022	911	1207	-0.139	0.024	987	1344
bias correction based on prior achievement + family background								
mixed – vocational	-0.264	0.018	1203	1623	-0.263	0.019	1255	1706
academic – mixed	-0.156	0.017	1234	1534	-0.169	0.017	1245	1557
within track	-0.151	0.008	6050	7047	-0.119	0.009	6216	7270
academic – academic	-0.173	0.014	2161	2447	-0.115	0.014	2182	2465
mixed - mixed	-0.146	0.012	2568	3275	-0.134	0.013	2627	3344
vocational – vocational	-0.113	0.023	911	1207	-0.115	0.024	987	1344

Figure A3

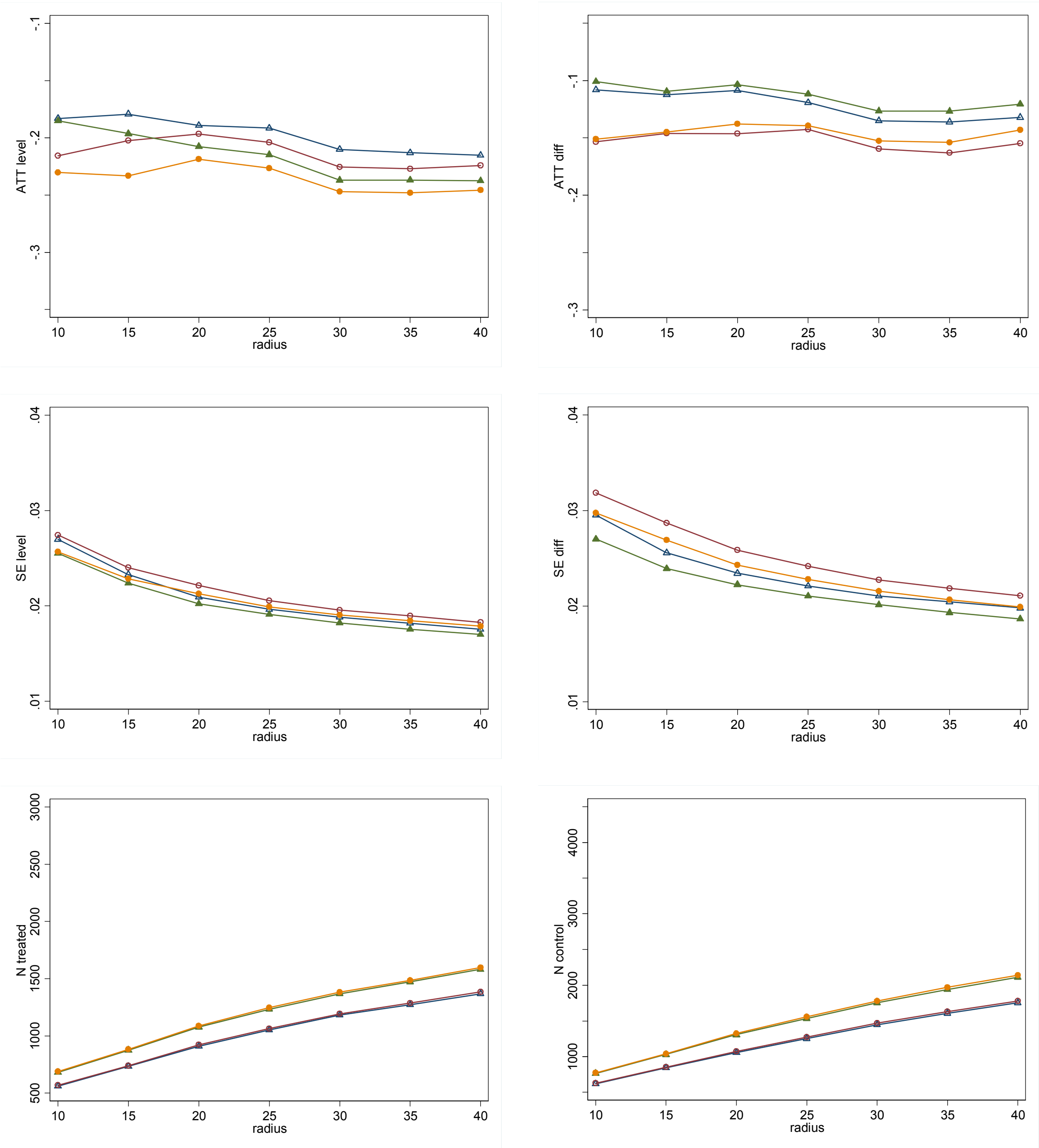
Estimated treatment effects, standard errors and number of observations with different matching parameters, preferred class: mixed, actual class: vocational



▲ math, p=0.05, Δ math, p=0.1, ● reading, p=0.05, o reading, p=0.1

Figure A4

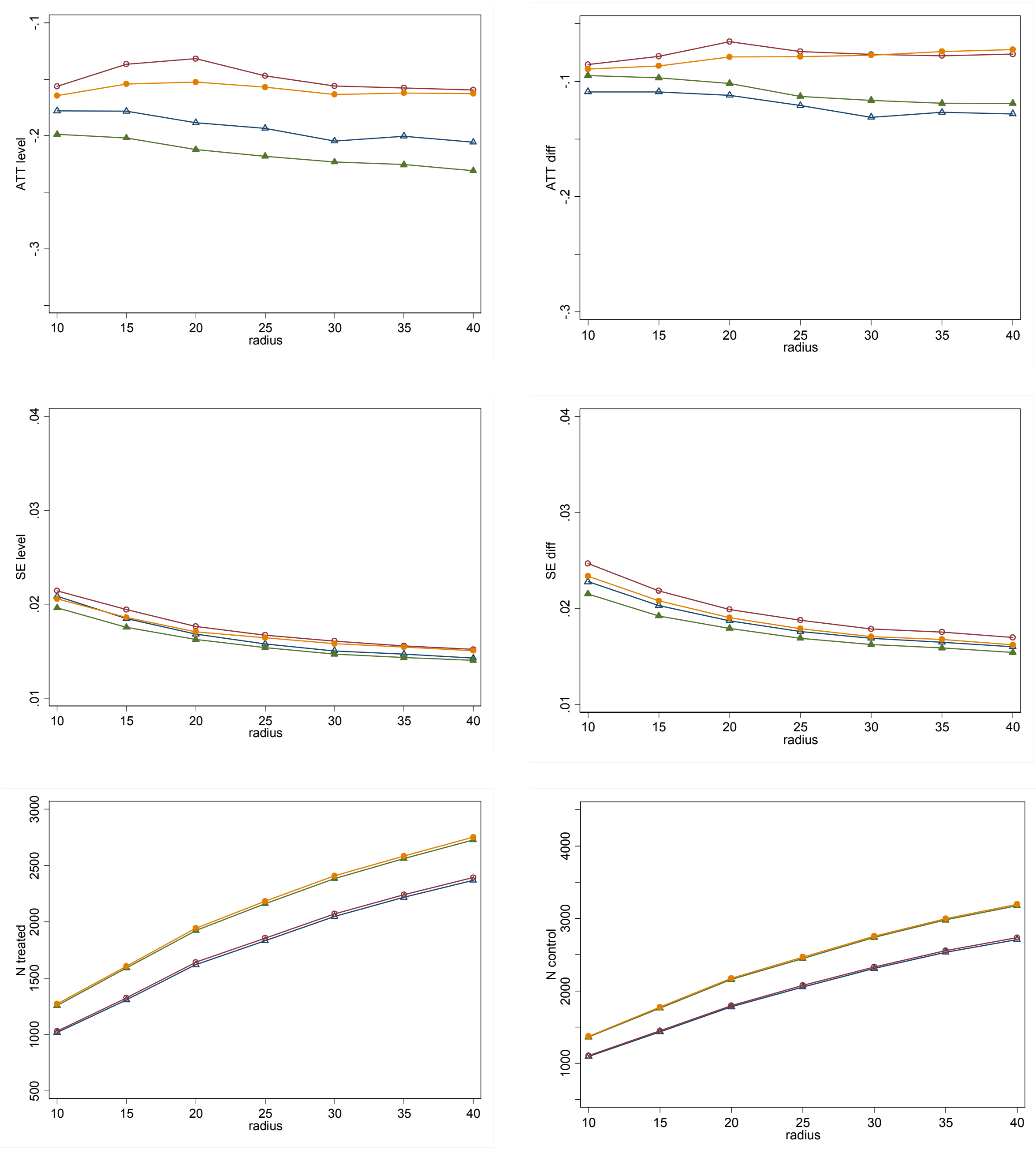
Estimated treatment effects, standard errors and number of observations with different matching parameters, preferred class: academic, actual class: mixed



▲ math, p=0.05, Δ math, p=0.1, ● reading, p=0.05, o reading, p=0.1

Figure A5

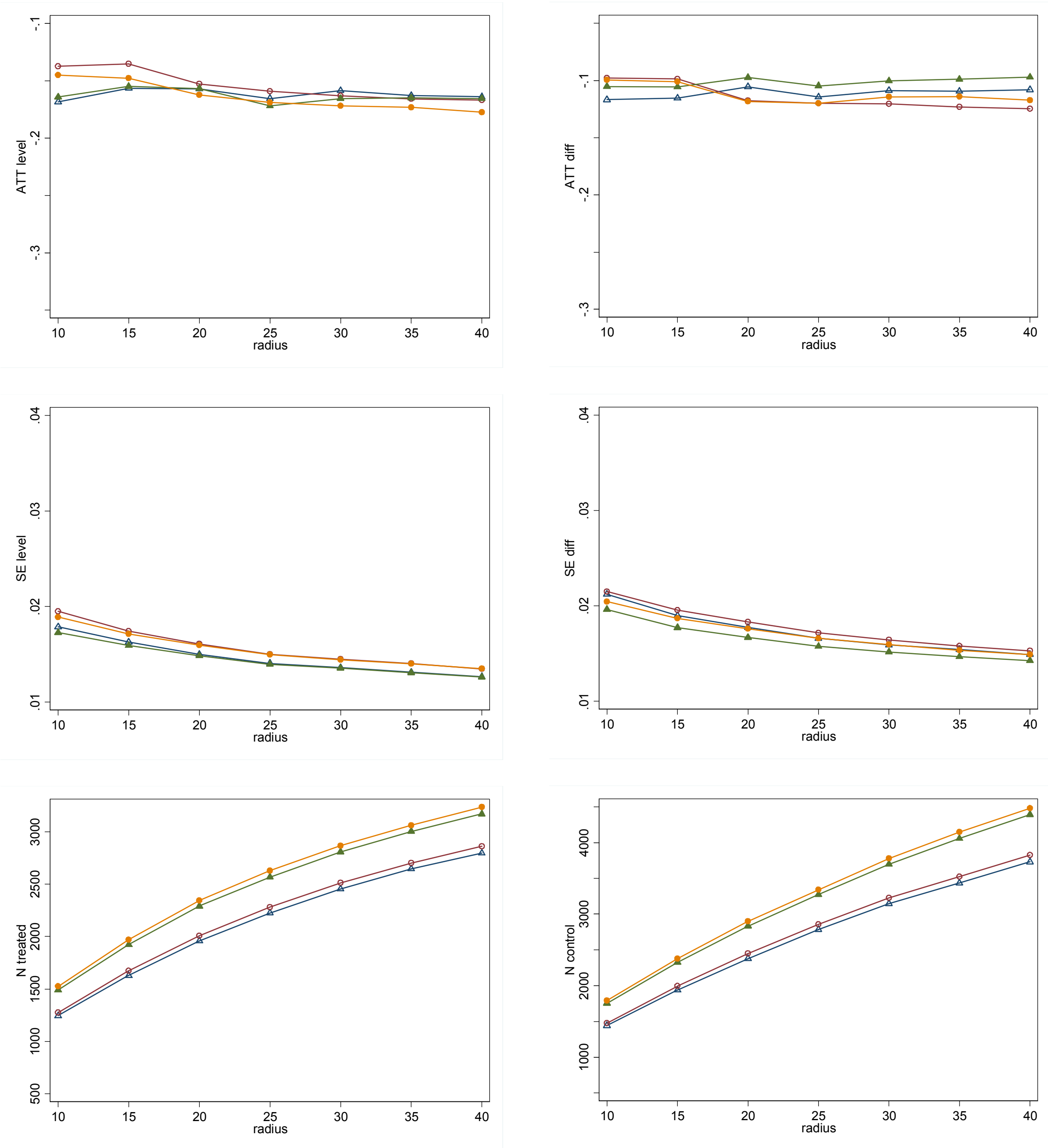
Estimated treatment effects, standard errors and number of observations with different matching parameters, preferred class: academic, actual class: academic



▲ math, p=0.05, Δ math, p=0.1, ● reading, p=0.05, o reading, p=0.1

Figure A6

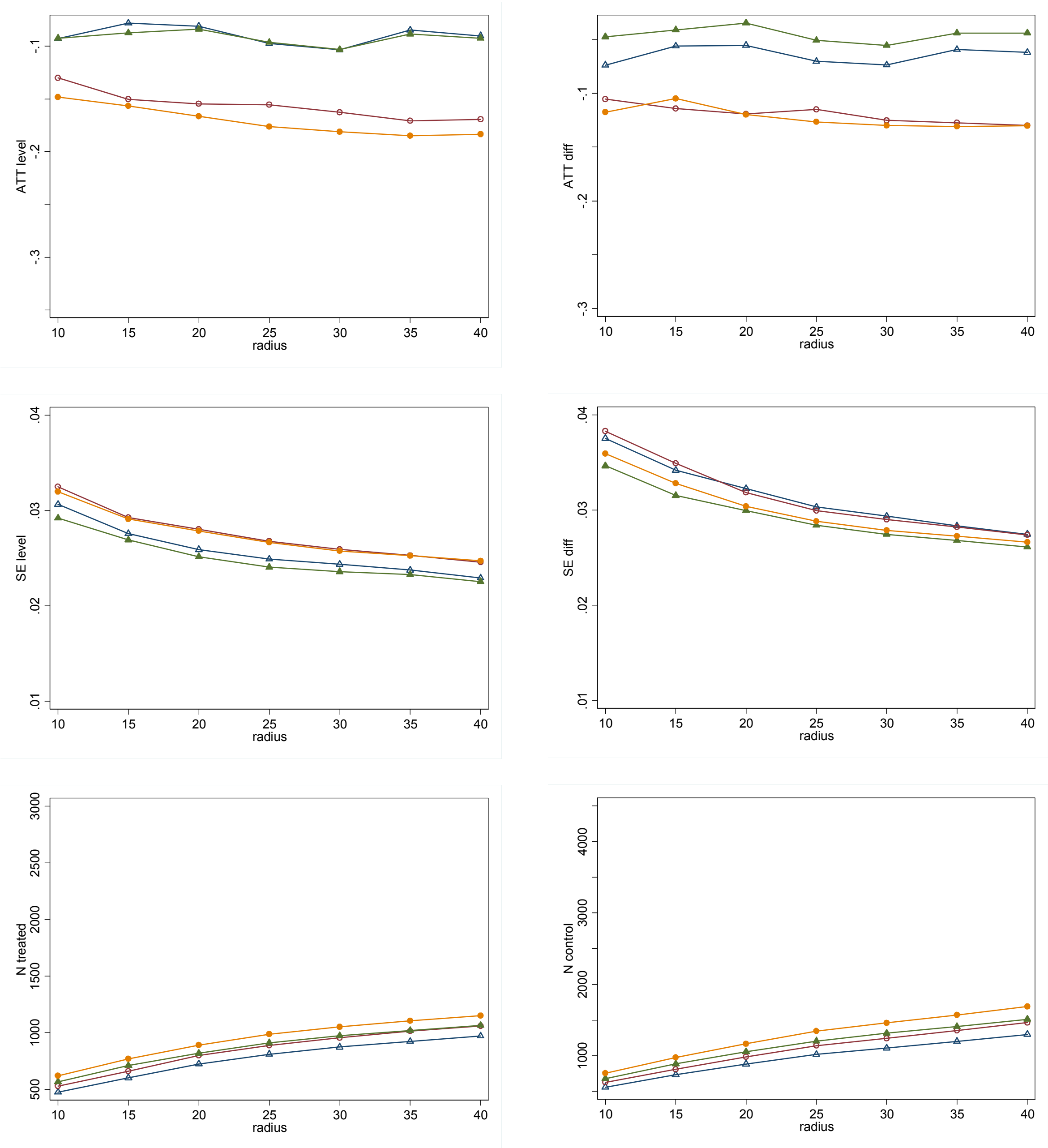
Estimated treatment effects, standard errors and number of observations with different matching parameters, preferred class: mixed, actual class: mixed



▲ math, p=0.05, Δ math, p=0.1, ● reading, p=0.05, o reading, p=0.1

Figure A7

Estimated treatment effects, standard errors and number of observations with different matching parameters, preferred class: vocational, actual class: vocational



▲ math, p=0.05, Δ math, p=0.1, ● reading, p=0.05, o reading, p=0.1