Are You on the Right Track?

The Effect of Educational Tracks and Preferred Schools on Student Achievement in Upper-secondary Education in Hungary

Zoltán Hermann

Centre for Economic and Regional Studies of the Hungarian Academy of Sciences

hermann@econ.core.hu

The paper attempts to identify causal effects of different educational tracks and better schools within tracks on student achievement in upper-secondary education in Hungary. Average treatment effects on the treated are estimated with a matching method. Rejected and admitted students with similar prior performance who applied to the same school are compared. Results indicate that both higher tracks and the preferred schools within the tracks improve student achievement significantly while the vocational track incurs disproportionate losses. At the same time the track and better school effects display substantial heterogeneity. The effect diminishes when the student is far below the peer mean of the preferred class. The overall impact of tracks and better schools on equality of opportunity is modest.

JEL No. I20, I21, I24

Keywords: education, tracking, ability sorting, equality of opportunity, matching

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

1 Introduction

When searching for institutional determinants of inequality in education, the early tracking of students is often blamed. However, the impact of tracking is still far from conclusive and causal evidence is scarce. This paper attempts to identify causal effects of enrollment 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 relatively straightforward in most cases. In a cross-country comparison the highest segregation is found in countries with extensive tracking (Jenkins, Micklewright, and Schnepf, 2008). At the same time, the second condition can not be easily evaluated due to overwhelming selection problems (Manning and 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 inherently more able and motivated and learn faster. The key question is whether differences in student achievement across tracks are explained solely by this selection or if it depends on the quality of education, 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. In this manner track effects on achievement provide indirect evidence on the overall impact of a tracking system, by representing 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 affect equity similar to tracking. Both theoretical models and empirical evidence suggest that free school choice may result in sorting of students across schools (Epple and Romano, 1998; Epple, Newlon and Romano, 2002). Ability sorting within schools (often referred as tracking in the US literature) may have similar effects. The extent to which ability sorting goes together with sorting by family background and more selective schools do indeed provide better educational quality, equality

[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.

of opportunity is endangered, as the opponents of school choice often argue (Ladd, 2002). On the other hand, a positive causal effect on achievement in the more popular schools seems indispensable for school choice to raise quality. If schools can maintain a high reputation and attract many students without higher quality, the incentives are not appropriate for raising productivity. Hence, exploring the effect of better schools provides evidence on the potential advantages of school choice for improving average quality and its side effect 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. Comparing these reveals whether track effects are something special or the tracks simply work as labels for schools of especially good or poor quality. If the latter is the case, ability sorting and stratification in a school choice regime may offset any improvement in 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 the existing evidence is ambiguous and most of the analyses are flawed by serious methodological problems and (see Betts, 2011 for a review).

Comparing tracking and non-tracking regimes tend to suggest that early tracking strengthens the impact of family background on student achievement, and thus it hinders equality of opportunity (see Ammermueller, 2005; Brunello and Checchi, 2007; Schuetz, Ursprung, and Woessmann, 2008 on cross-country analyses and Meghir and Palme, 2005; Pekkarinen, Uusitalo, and Kerr 2009 on de-tracking educational reforms). At the same time, other studies find no such effect (see Waldinger, 2006 for cross-country comparison and Manning and Pischke, 2006 analyzing de-tracking in the UK). Cross-country comparison is difficult due to the small number of observations and many possible omitted country characteristics confounding the analysis. The analysis of educational reforms offers more reliable causal evidence, though it is not always clear to what extent the estimated effects were driven by de-tracking or other elements of the reforms. Experimental evidence on tracking regimes is scarce. Duflo, Dupas, and Kremer (2011) 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, thus the results should be interpreted as partial effects.

School-level cross-sectional analyses usually struggle with the endogeneity of tracking and selection issues. The results of US studies on tracking within schools are mixed, both regarding the effect on equality of opportunity and on mean student achievement (Argys, Rees, and Brewer, 1996; Betts and Shkolnik, 2000; Figlio and Page, 2002). Evidence from European countries is scarce and mixed. Elite track attendance appear to increase student achievement (Guyon, Maurin, and McNally, 2010; Horn, 2013) but track placement seem to have no impact on long-term outcomes (Malamud and Pop-Eleches, 2010; Dustmann, Puhani and Schönberg, 2012). Altogether the results are still ambiguous and causal evidence is limited.

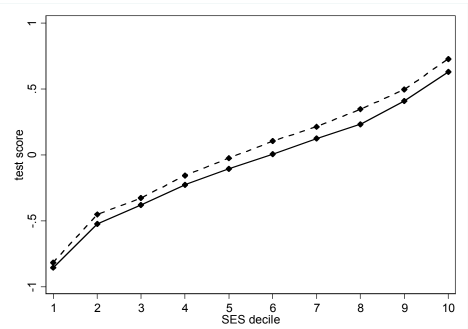
The other strand of related literature assesses the effect of selective schools. Recently several studies employed innovative strategies to identify causal effects of these schools. Some use lottery-based access to eliminate selection problems (Cullen, Jacob, and Levitt, 2006), but identification is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

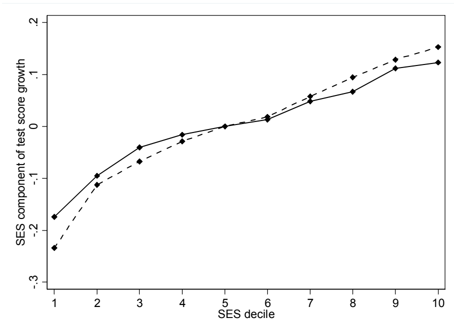
This paper attempts to provide causal evidence on the effect of educational tracks and better schools within the tracks. In order to avoid problems of selection based upon unobservable traits I use an empirical strategy similar to most papers in the better school literature. I compare rejected and admitted students who applied to the same school and have similar levels of prior achievement. Identification mostly relies on students that are on the margin, assuming that these students are also similar in unobserved characteristics like motivation, aspirations and self-confidence as they made similar application decisions. 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 achievement, essentially the same approach is employed using a different statistical method. In addition to track effects I also estimate the impact of the preferred schools of the students within tracks.

Upper-secondary education in Hungary provides good 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 an array of educational programs within the three formal tracks.

In Hungary equality of educational opportunity is extremely weak in an international comparison as student achievement has an exceptionally strong correlation with family background (see the data in Schuetz, Ursprung, and 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 the student enters upper-secondary education, but poor students on average seem to suffer some additional 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, and students with more disadvantaged family backgrounds are concentrated in the lower tracks (Figure 2, left panel). Some sorting within tracks is also apparent, as poor students on average attend classes with lower peer quality (Figure 2, right panel). Taken together this strong sorting implies that if educational tracks and better schools do matter for achievement, poor students can be expected to be adversely affected.

The analysis uses a large sample covering about one third of one 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.



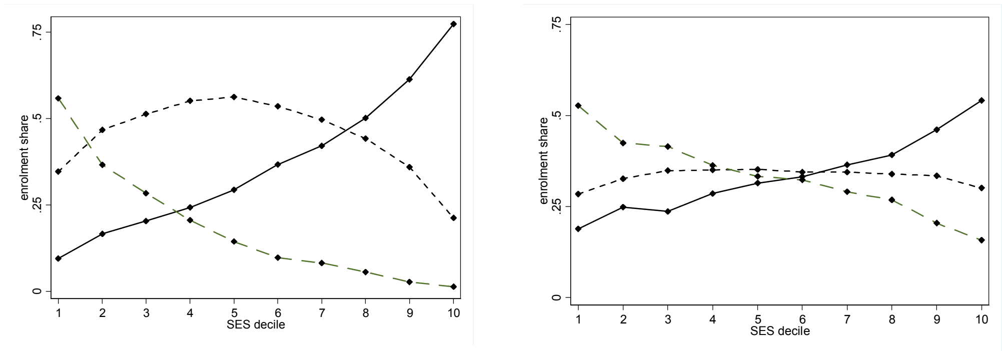


Test score in grade 10 SES component of test score growth (grade 8 – 10) [i]

Figure 1 Test scores and family background

—— : math - - - : reading

i: coefficients of SES decile dummies from regressions of test score in grade 10, controlling for prior test scores and gender



Enrollment share by track Enrollment share by peer composition within track [i]

Figure 2 Sorting of students with different family background across and within tracks

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

[i:] peer composition is measured by average prior math and reading test score of the class

The main finding is 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 (0.05-0.17 standard deviation). At the same time, attending the vocational track instead of the mixed track implies 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.

The highly 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, simulating inequality with no vocational track suggests only a minor improvement. Moreover, part of the track effect would likely be replaced by within-track ability sorting. When this is taken into account the estimated gain from de-tracking is only about half of that derived solely from the vocational track effect. Therefore tracking does not appear to be the main driver behind the overwhelmingly weak equality of opportunity in Hungary.

Finally, the results reveal substantial heterogeneity in better school and track effects. First, the less close substitute is the actual school for the preferred one in terms of peer composition, the larger is the magnitude of the effect. Second, the effect diminishes when the student is further below the peer mean of the preferred class in terms of prior achievement. Weak students may not benefit from higher standards when these are too demanding. This might explain why no consistent positive effect for elite schools is found in the literature: if the admission cut-off is set too far below the mean of admitted students, no significant effects near the cut-off can be expected.

The paper is structured as follows. In the second section the relevant features of secondary education in Hungary are reviewed. The third section introduces the data and presents descriptive analysis. The fourth section details the empirical strategy. In the fifth section the results are summarized and discussed with a focus on equality of opportunity. This is followed by robustness and sensitivity analysis. Finally conclusions are drawn.

2 Upper-secondary education in Hungary

Primary and secondary education in Hungary is organized in a two-tier system. Primary and lower-secondary education is provided in general schools, covering grade 1-8. 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.[2] 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, a prerequisite for higher education. The two tracks contain similar curriculum, with a minor share related to vocational education in the mixed track. Alternatively, general curricula in the vocational track are less demanding. Vocational training starts at grade 11, either in school workshops or at firms. Students can not take a maturity exam, thus the vocational degree they receive does not qualify for higher education. In this respect the vocational track is a dead-end. There are perceived differences in the prestige of the three tracks, and students almost exclusively rank the academic track above the mixed and that above the vocational track.

Within the tracks there is significant sorting across schools. Also, most schools offer several programs within the same track with some specialization, usually organized into separate classes. Typically in the mixed or vocational track the type of occupation defines the program, e.g. technical occupations in agriculture, health care services etc. In the academic track the distinction is often made with respect to the foreign language taught. Though an

[2] As an exception, there are few extremely 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.

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. An application refers to a given class in a given school within a given track. Second, schools rank the applicants for each class from the most to the least preferred and may strictly reject some of the applicants. While schools are relatively free to determine and weight different admittance criteria, rankings are still essentially merit-based. They mostly rely on the results from a central entrance exam and the grades received in their final year of general school. Some very popular and selective schools also require an oral entrance exam, while others do not even require the central entrance exam and instead rank the applications solely based on their grades. 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 are matched using a centralized allocation algorithm. This is a Gale-Shapely algorithm (Kóczy, 2010), providing no incentives for students to deviate from their true preferences in picking and ranking classes.

3 Data and descriptive analysis

The analysis covers a single cohort of students, who finished general school and started their upper-secondary education in 2006. The 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 grade in general school, and several variables on students' family background. The USEAA file reports the class, school and track, preference ranking, and final admission based on the admission algorithm. However, rankings of the applicants by the schools are not available in the 2006 USEAA file.

Both the NABC and USEAA files are comprised of administrative data, covering the full student population with few exceptions. However, the final sample consists only of the set of students whose data from all the three files can be linked. The final sample includes 34084 students, covering about one third of the student cohort. The major source of missing data is the failure to identify the student in one of the files properly, which occurs randomly, but students who drop out of or repeat grade 9 are not observed either. Nonetheless, a comparison of the sample and population test score distributions suggests that both the grade 8 and grade 10 student populations are represented sufficiently (Hermann, 2013).

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 across 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.

Table 1 presents OLS estimates for track and school quality effects. In these simple specifications, test scores in grade 10 are regressed on the track dummy variables gender and the two prior test scores. The results suggest substantial track effects. The expected gain in the academic track amounts to about one third and one half of a standard deviation in math and reading compared to the vocational track. The biggest 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. In the estimation of school effects confounding unobserved student characteristics emerge as a major problem. Even when prior achievement is similar, students in different tracks can still have different aspirations, motivation and self-confidence. At the same time, these characteristics can evidently affect later achievement (see Christofides et al., 2012; Brunello and Schlotter, 2011), e.g. more motivated students can be expected to take more ambitious choices and perform better in part because of their motivation, which interferes with the estimation of the true school or track effect. For this reason controlling for prior achievement alone cannot eliminate selection bias (Manning and Pischke, 2006). Specifications 2 and 3, which include dummies for students rejected from a higher track, provide some indirect evidence indicating that selection on unobservable variables is indeed present. Rejected students tend to outperform both their track and classmates who have not applied for a higher track but demonstrate similar prior achievement, suggesting that applicants and non-applicants do indeed differ in terms of some unobserved characteristics.

	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

Table 1 OLS estimates of test scores in grade 10

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

4 Empirical strategy

4.1 Identification

The identification of track and better school effects builds on merit-based selection of students into schools. The comparison of students just below and above the admission threshold provides an estimate of the true causal impact of a certain school on student achievement. This approach has become widespread in the recent literature on elite schools effects (Abdulkadiroglu, Angrist and Pathak, 2011; Dobbie and Fryer, 2011; Clark, 2010; Jackson, 2010; Pop-Eleches and Urquiola, 2011; Janvry, Dustan, and Sadoulet, 2012). The most straightforward and often applied estimation strategy in this setting is the regression discontinuity method.

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. Regression discontinuity estimation can not be used because the rankings of students by schools and admission cut-offs are not observed. Successful applicants are matched with unsuccessful ones with similar prior achievement. 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 exclusively reflect prior achievement makes the use of the matching method even more warranted.

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

Formally, the effect of being rejected is estimated as the difference of the outcome between the treated student and the mean of its matched control pairs:

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

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, a sufficient number of control students are required who are fairly similar to the treated students according to the determinants of academic performance. Second, there should be no such differences in unobserved characteristics between the treated and the control group that might affect achievement later on.

By matching rejected and admitted students the possibility of interference from unobserved student characteristics like motivation, aspiration and self-confidence can be eliminated. It is assumed that both types of 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 a track can be quite different: those choosing an elite school 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 control students are chosen from the 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 between the treated and control groups. Since the schools rank the applicants mainly based on entrance exam scores and grades in the general school, further unobservable cognitive or non-cognitive traits are unobservable not only for the researcher but for the schools as well, and consequently the selection can hardly be based on these. Possible exceptions are few elite schools arranging also 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 performs well and is admitted. Third, schools may weigh grades in grade 8 and entrance exam scores differently while also considering non-academic factors like an enrolled sibling or a 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.[3]

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 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 secondary and tertiary options. Consequently the estimated effects are not the difference between the *average* classes of two tracks. Third, the estimated effect of educational tracks can be

[3] In 2007, 29% of student pairs who applied to the same two classes were ranked differently.

interpreted as a weighted average of individual class effects, weighed by the number of unsuccessful applicants.

4.2 Correcting for unbalanced covariates and estimating average treatment effects

Though treated and control students are similar, they are not identical with respect to prior achievement because of students that are below and above the margin. Rejected students tend to lag behind their admitted peers, and this imbalance may lead to biased estimates (Abadie and Imbens, 2002). In order to control for this bias, I employ two alternative methods.

For one, I use the bias-correction proposed by Abadie and Imbens (2002). This starts with estimating the effects of individual characteristics on the outcome parametrically:

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

where X denotes the vector of individual characteristics, β is the estimated vector of parameters and μ is the error term. Note that when treatment effect for the treated is to be estimated only the control group is used here. In the present case I use prior student achievement; math and reading test scores and the average grade score for the correction (see the calculation of the grade 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) $$(A^{10} \mid S = 0) - (A^{10} \mid S = 1) = A_T^{10} - \frac{1}{k}\sum_j (A_{Cj}^{10} - \beta X_{Cj} + \beta X_T)$$

The value of this correction depends on whether the specification of the parametric achievement model is correct. Unfortunately, this proves to be a rather strong assumption for student achievement when prior achievement is included on the left hand side (Todd and 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.

Second, 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 difference 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 approach assumes that small differences in prior achievement do not affect achievement growth. However, it may be slightly biased by regression to the mean: the tendency for students with lower prior test scores to produce larger growth. Since prior scores for the treated tend to be lower than that of the controls, the difference approach 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 cases. For the achievement level and difference models respectively:

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

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

where index i and j denotes treated cases and control students respectively, n is the number of treated observations and k_i is the number of control students belonging to treated student i. Note that the treatment group is defined as the set of rejected applications, i.e. a single student may occur more than once if she has been rejected from several classes.

4.3 Matching method

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

Matching is exact with respect to gender and the class and school that students applied to. In other words, unsuccessful applicants are matched with students enrolled in the 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 and 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 grade 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. 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 not be matched with any control students. If the 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 capable 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. As students are scattered 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. Note that, as expected, matching with respect to test scores does indeed tend to provide somewhat larger treatment effects than the preferred matching method when the test score difference model is used (see the section on sensitivity analysis below).

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

Treated and control students are matched with respect to this adjusted average grade score. This 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].

Additionally, in order to mitigate the problem of potential measurement error in the average grade *within schools*, control students with a grade 8 test score that is statistically significantly different (at the 5% level) from that of the treated student are deleted.[5] That is, even if two students are similar with respect to the average grade, if either their math or reading test scores in grade 8 is almost certainly different, they are not matched.

When using the radius matching the number of controls assigned 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 formula in Abadie et al. (2004).

4.4 Results of matching and assessing the identification assumptions

In order to produce reasonable estimates of treatment effects a matching method requires a substantial overlap of the treatment and control groups. 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. The shares of successful matches are similar across and within tracks. Students in the treatment group cover 17.5% of the total sample. The matched applications were rejected in 1941 classes of 691 schools, representing 40% of the total classes in the sample.

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 join the academic track, and students who preferred another class within each of the three tracks.[6]

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

[5] The significance of the difference in the test scores is evaluated using the analytical standard errors provided for each score in the NABC file.

[6] Note that a minor share of students in the first group in fact applied to an academic track class, but for the 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.

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, on average, perform between the averages of the two tracks but closer to the track they ultimately 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.

track of preferred - actual class	rejected applicant test sc.	peer mean in pref. class	peer mean in actual class	diff. between classes	peer mean in pref. track	peer mean in actual track	diff. between tracks
math							
mixed – voc.	-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 – acad.	0.46	0.82	0.39	0.43	0.53		-
mixed - mixed	-0.15	0.14	-0.19	0.34	-0.03		-
vocational – voc.	-0.95	-0.76	-0.97	0.21	-0.86		-
reading							
mixed – voc.	-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 – acad.	0.64	0.84	0.52	0.32	0.63		-
mixed - mixed	-0.11	0.13	-0.19	0.31	-0.02		-
vocational – voc.	-0.92	-0.73	-0.97	0.24	-0.87		-

Table 2 Prior test scores of matched rejected applicants and the peer mean in the preferred and actual classes and tracks

Figure A1 of the Appendix represents rejected students' prior achievement in more detail. It illustrates the distribution of grade 8 test scores for the entire student population and rejected applicants by track. The distributions of applicants that were rejected within a track closely represent the entire student population of the tracks. 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 from their actual track. However, test scores in this group are still highly divergent, 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.

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, this would imply 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 a 0.2-0.35 standard deviation below the peer mean of the preferred class (Table 2), and half of them would be in the bottom third of the prior achievement distribution in that class if admitted (Hermann, 2013).

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 analyzed outcome. Of course, this assumption can not be tested directly. Comparing the means of prior achievement and family background variables for the treated and control groups shows no major differences, which is reassuring regarding unobserved characteristics (Appendix Table A2). However, since rejected applicants tend to be below the margin while control students are above it, prior scores of the latter group are slightly higher and family background is somewhat more favorable. Altogether, the matched sample is to some extent unbalanced in terms of student characteristics, which makes a correction for this necessary.

5 Results

5.1 Main results

The estimated average treatment effects of being enrolled in a less preferred class as opposed to a more preferred class are summarized in Table 3. The first two columns present the estimated effect on the test score levels in grade 10, the next two uses the same estimator with parametric bias correction (see in the section on empirical strategy). Column 5 and 6 present results on the test score difference 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 grade 10 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 the student had been enrolled in a mixed track class, he/she could have reached the 32 th-34th percentile for both subjects.

The estimated negative impact of the mixed track relative to the academic track is smaller, between one tenth and one fifth of a standard deviation, but statistically significant.

Note that the negative treatment effects for being rejected from a higher track are similar in size to the track effects estimated by OLS for math, and somewhat smaller for reading (Table 1).

	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 – voc.	-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
academic – acad.	-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 – voc.	-0.096	0.024	-0.115	0.022	-0.051*	0.028	911	1207
reading								
mixed – voc.	-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
academic – acad.	-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 – voc.	-0.176	0.027	-0.139	0.024	-0.126	0.029	987	1344

Table 3 Estimated effect of being rejected from the preferred class on student achievement

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

The next three 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 within the vocational track, where the preferred class effect is smaller and statistically significant only at the 10% level when estimated with the difference model.

The results of the three estimation methods are rather similar. The parametric bias correction moderates the average treatment effects estimated for test score levels, but for most cases the change is insignificant. As expected, estimates for test score differences tend to produce smaller treatment effects in accordance with the assumed bias toward zero due to regression to the mean. Thus the difference 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 the academic versus mixed track effect. This suggests that the academic track does not generate an advantage greater than a better school does 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. 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 his/her preferred one, his/her achievement may not drop too much. On the other hand, if the student must accept a far inferior class than her first choice, she can be expected to suffer a greater loss. It is 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 of substitutes for the preferred choice as the less preferred options within the same track? If so, the effect of higher tracks and preferred schools is essentially similar. Otherwise is there any track specific element represented in the large negative effect of the vocational track?

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. 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 distinguish pure track effects 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 difference from grade 8 to 10, as defined in eq. 4. The observations are the rejected applications.[7] The combinations of the track of the preferred and actual classes are represented by dummy variables. Observations with extreme values of the dependent variable; below the 1st or above the 99th percentiles are excluded[8]. Standard errors are clustered for the preferred classes.

The model is first estimated with no controls, representing the average treatment effects in this restricted sample (specification 1, Table 4). Then three controls are included in order to account for individual heterogeneity: these are 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 measure of ability was 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 the reading equation.

The equations are estimated without a constant term, thus 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 change when the control variables shift to values that are typical within-track.

---

[7] Estimates for the bias-corrected level effects are not shown, but provide similar results.

[8] Estimates for the full sample or that restricted to the 5th - 95th percentiles produce qualitatively similar results.

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

	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 – acad.	-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 – voc.	-5.005**	-5.705**	-5.663**	-13.033***	-13.545***	-12.913***
	(2.494)	(2.531)	(2.554)	(2.393)	(2.388)	(2.482)
diff. in peer mean		-0.136***	-0.135***		-0.129***	-0.132***
		(0.020)	(0.020)		(0.023)	(0.023)
distance from peer mean in pref. class		0.193***	0.195***		0.192***	0.194***
		(0.017)	(0.017)		(0.015)	(0.015)
distance from pref. 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 appl.)	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

Table 4 Regression estimates for the effect of being rejected from the preferred class
Standard errors clustered for preferred classes in parentheses. *** p<0.01, ** p<0.05, * p<0.1

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 his/her preferred choice then rejection is less harmful. Conversely, enrollment in a class that is further below the preferred class in terms of peer mean is associated with a larger negative effect. Increasing the difference between the preferred and actual classes from zero (close substitutes) to a 0.33 standard deviation of test scores (the average case within tracks) increases the effect by about a 0.04 standard deviation, regarding both math and reading.

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, while 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, he/she would have benefitted much less or nothing from being admitted. For math the effect is slightly non-linear, resulting in some flattening when the distance is large. This implies that beyond some point being further below the level of the preferred class makes less difference.

These results suggest that choices which are overly ambitious, even if realized, do not necessarily pay off. This is probably related to standards and the level of required effort established by teachers with respect to the capabilities of the typical students in the class (see Moen and Tjelta, 2010 for evidence on setting grading standards this way in Norway). A student who fits well into the class in terms of his/her skills, abilities, and prior knowledge may benefit from higher standards. Conversely, if these are too demanding for the student then less benefit can be expected. This is also consistent with grading standards heterogeneously affecting the performance of students, as shown by Figlio and Lucas (2004).

Gender appears to be related to the size of the treatment effect for math; girls on average lose more with being rejected. 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 a 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 within-track average values; 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 versus a preferred class within tracks, pair wise comparisons of the coefficients of the track combination variables are tested. 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).

When the peer mean and the relative student ability is controlled for, the difference between the vocational track effect and the other effects decreases somewhat. This confirms that the large negative average treatment effect estimated for the vocational track is indeed in part due to vocational classes being further apart as 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, with an added negative vocational track effect. For education policy the next question is what makes better schools better? In assessing the patterns of the effect of the difference in peer means and the distance of the student from the preferred class, it is tempting to invoke a pure peer group effect explanation (Hoxby, 2000; Hanushek et al, 2001; Lavy, Silva, and Weinhardt, 2009). Differences in teacher quality together with positive student-teacher matching are another natural candidate (Lankford, Loeb, and Wyckoff, 2002; Clotfelter, Ladd, and Vigdor, 2006; and Varga, 2009 for Hungary).

Unfortunately, the available data do not allow for the mechanisms behind the effects to be distinguished. However, peer effects and teacher-student matching can be assumed to be equally influential with regard to tracks and better schools within tracks. Consequently, the vocational track effect should be generated in part by other factors, specific to this track. Differences in curriculum and lower standards are likely candidates.

5.2 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 gains or loses by these effects. As shown above (Figure 2) poor students are sorted into the lower tracks at a disproportionately higher rate 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 make 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 while 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 randomly among the rejected applicants.[9]

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.

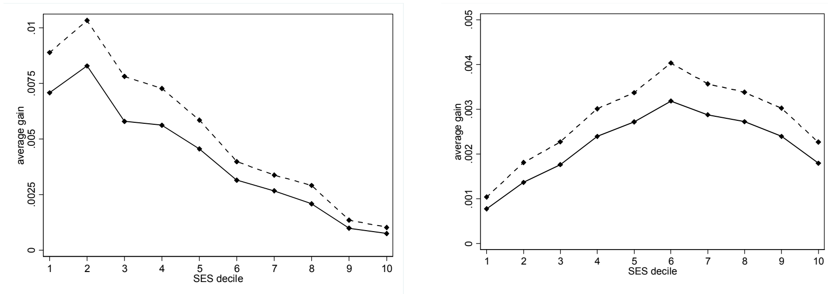
Figure 3 depicts the results of these simple back-of-the-envelope calculations about the effect of admitting additional marginal students into their preferred classes. The results reveal that decreasing the share of the vocational track has the strongest positive impact on equality of opportunity. Students from poor families would benefit most while more affluent families would remain unaffected. Still the magnitude of the gain from a marginal change is modest in absolute terms when compared to the overall level of inequality (see Figure 1).

Admitting more students to the academic track has a more ambiguous and weaker impact. Altogether, if better schools expanded enrollment within all the three tracks equality of opportunity would remain unaffected.

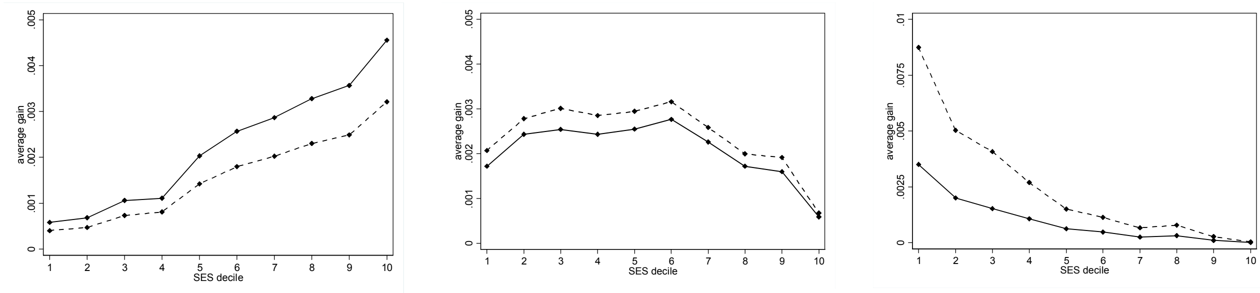
Overall, the simulated impact of marginal changes suggests that tracking, and more specifically tracking at the lower end, is more closely related to inequality of opportunity than ability sorting within tracks. The reason for this is twofold. First, the estimated vocational track effect is larger than better school or academic track effects. Second, family background is much more diverse between the vocational and the mixed track than in the other cases. However, the improvement in equality of opportunity from a marginal change in the

[9] Not limiting the choice for rejected students provides similar results (Hermann, 2013).

vocational track enrollment is still small. This is not only due to the small changes in the share of tracks, but also because within the vocational track students who applied to the mixed track tend to have a more favorable family background than those who settled for vocational education when making their application decisions.



more mixed, less vocational   more academic, less mixed

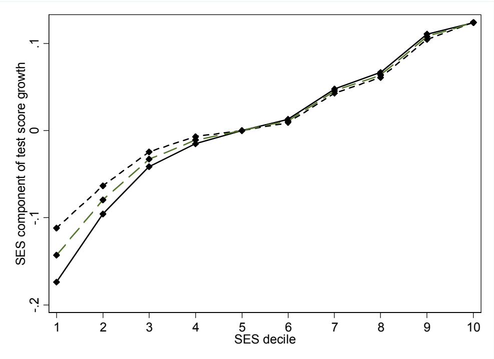


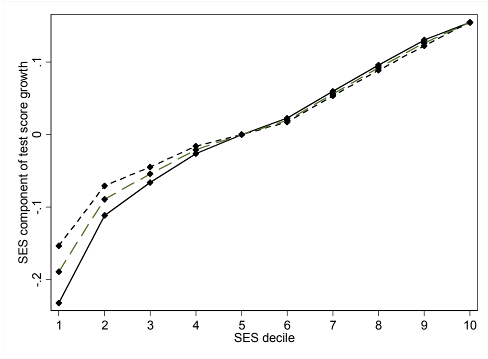
more better schools: academic   more better schools: mixed   more better schools: vocational

Figure 3   The estimated average gain in grade 10 test scores from a 2% expansion of higher tracks and better schools within tracks for family background deciles

—— : math, – – – : reading

Beside the effect of marginal changes in enrollment there is a more fundamental question as to what would happen if tracking was eliminated altogether. The large detrimental effect of the vocational track, together with a relative concentration of poor students in it, suggests that eliminating this track 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 those 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, equality of opportunity would indeed be improved (dashed line). However, the minor magnitude of this improvement suggests that tracking is not the main factor behind the outstanding inequality of opportunity that is evident in Hungary. Moreover, the track effect would probably be replaced by within-track ability sorting. When this is taken into account and the estimated gains are decreased by within track better school effects (long dashed line) about half of the improvement disappears.





math reading

Figure 4 The SES component of test score growth with the vocational track eliminated

Coefficients of SES decile dummies from regressions of test scores 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

6 Robustness and sensitivity analysis

6.1 Robustness to alternative explanations

A possible argument against the causal interpretation of the track and preferred class effects could refer to the peculiar effect of rejection on motivation. Being rejected can cause frustration and hence diminish student effort. One would expect frustration to grow with the number of rejections; being admitted to the class ranked as a 4^{th} or 6^{th} choice is probably more frustrating than getting a place in the 2^{nd} one. This hypothesis can be tested by including dummy variables representing the preference rank given by the student to his/her actual class in the regression models for the treatment effect. 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 A3 of the Appendix do not appear to support the frustration hypothesis as lower ranked classes do not consistently increase the magnitude of the negative treatment effect.

When comparing track and better school effects, some omitted school characteristics may bias the results. It is possible that student preferences also depend on other factors beyond academic quality, and that these preferences play a larger role in within-track choices. Two possibilities are the distance from the school and the owner of the school. Some students may rank a weaker school that is closer higher than a 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, first I repeat the regression estimates excluding applications when the preferred and actual class is located in distinct towns and when the type of the owner (government, church, non-church private) is different. Second, the estimation sample is restricted to government-run schools. Results are essentially unchanged, suggesting that these preferences do not account for the within-track effects being smaller than the vocational track effect (Hermann, 2013).

6.2 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.

Regarding alternative matching methods 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). Third, track effects are estimated using a restricted set of control students with a more similar application profile, i.e. applying to the lower track as well.[10] The first and second alternatives produce virtually identical results to the baseline method (Hermann, 2013). 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.

Matching by other covariates also provide similar results (Hermann, 2013). First, matching by the prior score of the subject (math or reading) provides similar treatment effects, which tend to exceed the baseline estimates, as expected, due the bias related to the measurement error in prior test scores. In the second case both the average grade and the two prior test scores are used for the matching. A matched pair is constructed only 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

[10] E.g. a treated student who applied for a mixed class but enrolled in a vocational class is matched to control students who in the mixed track who also applied to a vocational class.

the same radius value, 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 do not appear to be sensitive to the radius value (between 0.1 and 0.4) and the set of covariates used for the parametric bias correction (Hermann, 2013).

7 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 were ultimately 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 significant. 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.

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 suggests that if the academic and mixed tracks were replaced by a single general track, the average achievement and equality of opportunity would probably not change - school choice accompanied by ability sorting would essentially reproduce the present stratified school system and produce similar outcomes. In other words, tracking does not necessarily matter when school choice is present. At the same time the vocational track does have a specific negative effect in addition to that of less preferred schools in general. This implies that in some cases tracking may have a greater impact than school choice, especially if one of the tracks provides much more inferior opportunities and prospects.

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 improve equality of opportunity to some extent. However, the estimated gains of de-tracking are modest at best.

An important question for policy is whether the better school or track effects are specific to a particular group of students; those on the margin. In other words, can the benefits of higher tracks or better schools be easily extended to other students? First, remember that rejected students are not exceptional in terms of prior achievement, but represent the greater part of the achievement distribution. However, regarding other traits 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 of adolescents are more malleable than cognitive skills, as evidence on remediation programs targeting disadvantaged students suggests (Brunello and Schlotter, 2011 and Cunha et al., 2006). This implies that there is some room for education policy to make less motivated, ambitious, or confident students benefit from higher tracks and better schools.

References

Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido W. Imbens. 2004. “Implementing Matching Estimators for Average Treatment Effects in STATA.” Stata Journal 4(3): 290–311.

Abadie, Alberto, and Guido W. Imbens. 2002. “Simple and Bias-Corrected Matching Estimators for Average Treatment Effects.” NBER Technical Working Paper No. 283.

Abdulkadiroglu, Atila, Joshua D. Angrist, and Parag A. Pathak. 2011. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." NBER Working Paper No. 17264.

Ammermueller, Andreas. 2005. “Educational Opportunities and the Role of Institutions.” ZEW Discussion Paper No. 05–44.

Argys, Laura M., Daniel I. Rees, and Dominic J. Brewer. 1996. “Detracking America’s Schools: Equity at Zero Cost?” Journal of Policy Analysis and Management 15(4): 623–645.

Betts, Julian 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, 341–381.

Betts, Julian R., and Jeff Grogger. 2003. “The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings.” Economics of Education Review 22(4): 343–352.

Betts, Julian R., and Jamie L. Shkolnik. 2000. “The Effects of Ability-Grouping on Student Math Achievement and Resource Allocation in Secondary Schools”, Economics of Education Review 19(1): 1–15.

Bonesronning, Hans. 2008. “The Effect of Grading Practices on Gender Differences in Academic Performance.” Bulletin of Economic Research 60(3): 245-264.

Brunello, Giorgio, and Daniele Checchi. 2007. "Does School Tracking Affect Equality of Opportunity? New International Evidence." Economic Policy 22(52): 781–861.

Brunello, Giorgo, and Martin Schlotter. 2011. "Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education and Training Systems." IZA Discussion Paper No. 5743.

Clark, Damon. 2010. "Selective Schools and Academic Achievement", The B.E. Journal of Economic Analysis & Policy, 10(1)

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." Journal of Human Resources, 41(4): 778–820.

Christofides, Louis N., Michael Hoy, Joniada Milla, and Thanasis Stengos. 2012. "Grades, Aspirations and Post-Secondary Education Outcomes." IZA Discussion Paper No. 6867.

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

Cullen, Julie, Brian Jacob, and Steven Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." Econometrica 74(5): 1191– 1230.

Dee, Thomas S. (2005): "A Teacher like Me: Does Race, Ethnicity, or Gender Matter?" American Economic Review 95(2): 158-165.

Dobbie, Will, and Roland G. Fryer, Jr. 2011. "Exam High Schools and Academic Achievement: Evidence from New York City." NBER Working Paper 17286.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." American Economic Review 101(5): 1739–1774.

Dustmann, Christian, Patrick A. Puhani, and Uta Schönberg. 2012. "The Long-term Effects of School Quality on Labor Market Outcomes and Educational Attainment." CReAM Discussion Paper No. 08/12.

Epple, Dennis, and Richard Romano. 1998. "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects." American Economic Review 62(1): 33–62.

Epple, Dennis, Elizabeth Newlon, and Richard Romano. 2002. "Ability Tracking, School Competition, and the Distribution of Educational Benefits." Journal of Public Economics 83(1): 1-48.

Figlio, David N., and Maurice E. Lucas. 2004. "Do High Grading Standards Affect Student Performance?" Journal of Public Economics 88(9-10): 1815– 1834.

Figlio, David N., and Marianne E. Page. 2002. "School Choice and the Distributional Effects of Achievement Tracking Does Separation Increase Inequality?" Journal of Urban Economics 51(3): 497–514.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2001. "Does Peer Ability Affect Student Achievement?" NBER Working Paper No. 8502.

Hermann, Zoltán. 2013. “Are You on the Right Track? The Effect of Educational Tracks on Student Achievement in Upper-Secondary Education in Hungary.” Budapest Working Papers On The Labour Market, 2013/16, http://www.econ.core.hu/file/download/bwp/bwp1316.pdf

Horn, Dániel. 2013. “Diverging Performances: the Detrimental Effects of Early Selection on Equality of Opportunity in Hungary.” Research in Social Stratification and Mobility 32: 25-43.

Hoxby, Caroline. 2000. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” NBER Working Paper No. 7867.

Guyon, Nina, Eric Maurin, and Sandra McNally. 2010. “The Effect of Tracking Students by Ability into Different Schools: A Natural Experiment.” Fondazione Eni Enrico Mattei, Working Paper No. 152.2010

Jackson, Clement Kirabo. 2010. “Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago.”The Economic Journal120(549): 1399–1429.

Janvry, Alain de, Andrew Dustan, and Elisabeth Sadoulet. 2012. “The Benefits and Hazards of Elite High School Admission. Academic Opportunity and Dropout Risk in Mexico City.” Univeristiy of California at Berkeley. Unpublished.

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

Kóczy, Á. László. 2010. “A magyarországi felvételi rendszerek sajátosságai.” Közgazdasági Szemle 57(2): 142-164.

Ladd, Helen F. 2002. “Market-Based Reforms in Urban Education.” Washington: Economic Policy Institute

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. “Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis” Educational Evaluation and Policy Analysis 24(1): 37-62.

Lavy, Victor, Olmo Silva, and Felix Weinhardt. 2009. “The Good, The Bad and The Average: Evidence on the Scale and Nature of Ability Peer Effects in School.” NBER Working Paper No. 15600.

Leahey, Erin, and Guang Guo. 2001. “Gender Differences in Mathematical Trajectories” Social Forces 80(2): 713–732.

Malamud, Ofer, and Cristian Pop-Eleches. 2010. “General Education versus Vocational Training: Evidence from an Economy in Transition.”The 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, Costas, and Marten Palme. 2005. ”Educational Reform, Ability and Family Background.”American Economic Review 95(1): 414–424.

Moen, Jarle, and Martin Tjelta. 2010. “Grading Standards, Student Ability and Errors in College Admission” Scandinavian Journal of Educational Research 54(3): 221-237.

Pekkarinen, Tuomas, Roope Uusitalo, and Sari 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, Cristian, and Miguel Urquiola. 2011. Going to a Better School: Effects and Behavioral Responses. NBER Workin Paper No. 16886.

Schuetz, Gabriela, Heinrich W. Ursprung, and Ludger Woessmann. 2008. “Education Policy and Equality of Opportunity.” Kyklos 61(2): 279-308.

Todd, Petra E., and Kenneth I. Wolpin. 2003. “On the specification and estimation of the production function for cognitive achievement” Economic Journal 113(485): F3–F33.

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

Waldinger, Fabian. 2006. “Does tracking affect the importance of family background on students’test scores?” London School of Economics. Unpublished.

Appendix

Appendix A Calculating the grading standards adjusted average grade score

Following Betts and Grogger (2003) and Figlio and 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 grade 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 and school s. The estimated fixed effects represent the grading standard of schools. The control variables are gender, special education needs status, classes specialized for advanced education of foreign languages, or other subjects and the share of special education needs students in the class. The control variables may modify the grading standard within the school or that faced by 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 achievement nor in the estimated fixed effects. The predicted values from this model provide a prior score measured by the average grade 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 grade of the student and the grading standard in his/her school. 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.

Appendix B Tables and figures

		Grade 8			Grade 10	
		math	reading	grade 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 A1 Summary statistics of student achievement by educational track

	math		reading		father's years of education	
	treated	control	treated	treated	treated	control
track of preferred - actual class						
mixed – vocational	-0.56	-0.48	10.32	10.32	-0.47	-0.42
academic – mixed	0.20	0.30	11.62	11.62	0.31	0.35
academic – academic	0.46	0.57	12.47	12.47	0.55	0.58
mixed - mixed	-0.15	-0.08	10.93	10.93	-0.13	-0.09
vocational – vocational	-0.95	-0.91	9.79	9.79	-0.79	-0.77

Table A2 Average prior achievement and father's educational attainment of rejected applicants and control students

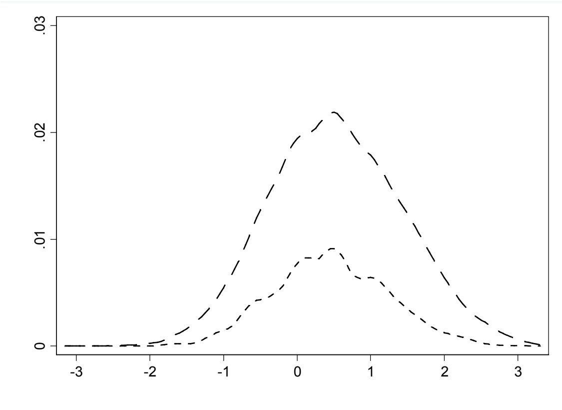
Averages calculated for rejected applications.

	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)

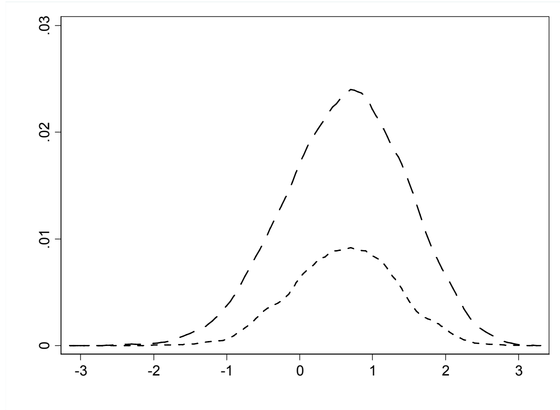
Table A3 Regression estimates of the effect of the rank of the actual class in rejected student's preference list on the effect of being rejected from the preferred class

Standard errors clustered for preferred classes of rejected applicants are in parentheses.

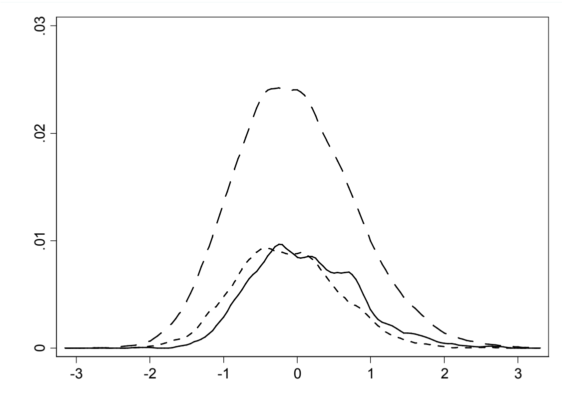
*** p<0.01, ** p<0.05, * p<0.1 Control variables as in Table 4.



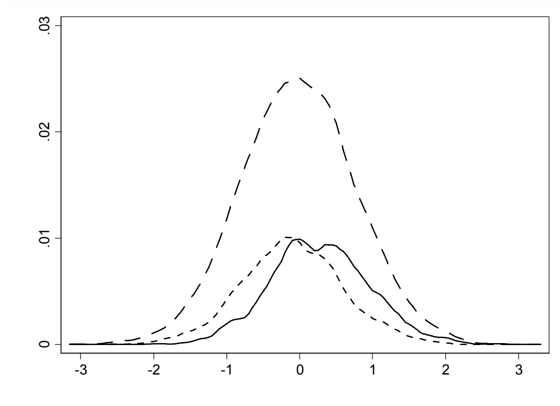
math, academic track



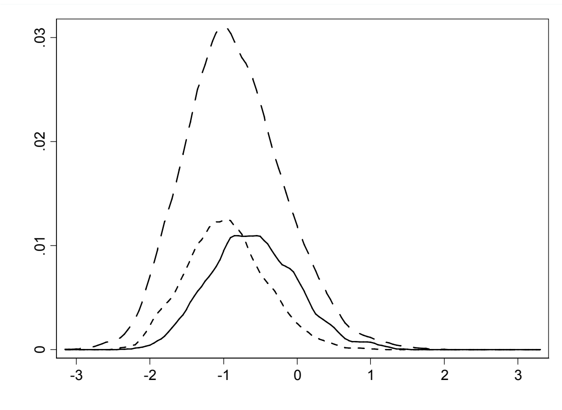
reading, academic track



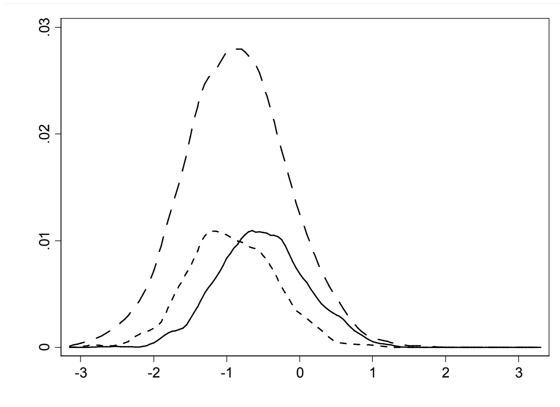
math, mixed track



reading, mixed track



math, vocational track



reading, vocational track

Figure A1 The distribution of test scores in grade 8 for the sample population of students and for rejected applicants by educational 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 reduced by a factor of three.