Does tracking impact short- and long-term educational outcomes? Evidence from school entry tests

Stefan Boes[a]

Dominik Hangartner[b,c]

Lukas Schmid[d,e]

March 2019

Abstract

This paper evaluates the effects of educational tracking on educational achievement in Switzerland using administrative records on secondary school entry tests linked with survey data. Regression discontinuity estimates suggest that sorting students into upper and lower level tracks does not affect secondary and tertiary education in school systems that are horizontally permeable. However, we find evidence for lock-in effects in non-permeable school systems. Here, students who barely passed the entry exam are approximately 17 percentage points more likely to achieve a university degree than students who barely failed this exam, which translates into one additional year of schooling on average. The tracking effects are most concentrated among female students from lower educational and socio-economic backgrounds.

Keywords: Early tracking, schooling achievements, quasi experimental data, regression discontinuity, educational inequality.

JEL codes: I21, I24, J10

[a] University of Lucerne, Department of Health Sciences and Health Policy and Center for Health, Policy and Economics, Frohburgstrasse 3, P.O. Box 4466, CH-6002 Lucerne, Switzerland. Phone: +41 41 229 5949, fax: +41 41 229 5635, email: stefan.boes@unilu.ch.
[b] ETH Zurich, Public Policy Group, Leonhardshalde 21, CH-8092 Zurich, Switzerland.
[c] London School of Economics and Political Science, Department of Government, Houghton Street, London WC2A 2AE, United Kingdom. Phone: +44 20 7955 6982, email: d.hangartner@lse.ac.uk
[d] Corresponding author. University of Lucerne, Department of Economics and Management, Frohburgstrasse 3, P.O. Box 4466, CH-6002 Lucerne, Switzerland. Phone: +41 41 229 5850, email: lukasdavid.schmid@unilu.ch.
[e] University of St. Gallen, School of Economics and Political Science, Bodanstrasse 8, CH-9000 St. Gallen, Switzerland.

The authors gratefully acknowledge support from the Swiss National Science Foundation (grants #100013_141333/1 and #100018_152798/1), the Jacobs foundation (grant #2012-1001), and the HSG WiPol fund.

1. Introduction

Equality of opportunity is a fundamental principle of modern education systems and at the heart of educational policy reforms (UNESCO, 2000). This principle stipulates that students from different backgrounds should have equal access to educational institutions and equal chances to complete the degree they strive to achieve. Early educational tracking[1] is an integral part of many education systems. It is designed to sort students into different tracks of secondary school according to their abilities, with the ultimate goal of providing the best education for their needs. Sorting in many cases is based on the performance at the end of primary school, or the results of a secondary school entry test. However, tracking systems run the risk of mis-tracking students whenever the selection of tracks is subject to incomplete information about the students' ability (Hanushek and Woessmann 2006).

In this paper, we investigate the effects of early tracking on educational achievements in Switzerland – a country that is characterized by a federal education system and substantive heterogeneity between local school systems. We explore the effects of two types of cantonal systems that have comparable secondary school entry tests but differ in the extent to which students can change tracks during their secondary education. We use detailed data on the long-term educational outcomes of 1,829 former students who took the test in the 1960s and 1970s to evaluate whether students just below or just above the passing threshold, who have arguably similar abilities, experience lock-in effects in their assigned school level. Switzerland offers a unique opportunity to study the effects of tracking because of its federal constitution and the regulatory differences in state-level education, within national standards that regulate the qualification at the end of secondary school and the admission to tertiary education.

We find that tracking has persistent effects on educational outcomes in a non-permeable system, where it is difficult for students to change the initial track assignment. Students who barely passed the entry exam are approximately 17 percentage points more likely to achieve a university degree than students who barely failed this exam, which translates into one additional year of schooling on average. In contrast, we find only short-term but no long-term effects on educational outcomes in a permeable system, where changing the initial track

[1] There is a somewhat different interpretation of the term 'tracking' in European education systems than in the U.S. While in the European literature tracking generally refers to the sorting of students into different tracks in secondary school (usually into a more advanced/academic track and a more vocational oriented track) based on entry examinations and/or prior achievements, in the US tracking is typically referred to as ability grouping within schools. Both concepts are of course closely related, but this paper is more about the former type.

assignment is not costly. The persistent effect of tracking in the non-permeable system is most pronounced for female students. Our results reveal that this differential impact of tracking likely stems from differences in latent populations. In particular, we document that the share of students who pursue the advanced track independent of the test result is considerably higher among male than female students. We further investigate the role of parental background and find that always-takers, those who obtain an advanced track degree independent of the entry exam, have parents with a relatively high socio-economic status in the female but not in the male sample. Our analysis also indicates that female always-takers more likely have a working mother than male always-takers, which suggests that attitudes towards the capacity of women in acquiring the same education as men is a critical factor in determining educational paths.

There are several implications of our study. First, our results help to better understand the differential impacts of tracking in permeable and non-permeable education systems. While cross-country comparisons have addressed this issue before (e.g., Hanushek and Woessmann 2006), the design of education systems and the exact mechanisms of tracking typically differ between countries, making it difficult to identify the underlying channels through which tracking affects educational achievements (see also Waldinger 2007). Within-country studies have the potential to address this heterogeneity, and exogenous variation in tracking typically comes from policy changes (altering the tracking mechanism), or from institutional characteristics of the tracking mechanism (e.g., age cut-offs to enter school). In a related discussion, Dustmann et al. (2014) do not find persistent tracking effects in Germany, which they argue is due to the permeability of the education system. We complement their study by explicitly identifying school systems within a country with different degrees of permeability, based on cantonal variation in the school regulations, and showing that the permeability of the education system can explain the long-term impacts of tracking.

Second, there is no clear consensus in the literature on the effects of tracking on student achievement (Betts 2011). While two recent studies using U.S. data find negligible effects of gaining admission to schools with high achievers (Abdulkadiroglu et al. 2014; Dobbie and Fryer 2014), Card and Giuliano (2016) show that tracking can raise educational achievements of minority students. Similarly, two studies using data on students in Romania and Trinidad and Tobago find positive effects of tracking on future student performance (Jackson 2010, Pop-Eleches and Urquiola 2013). This evidence is supportive of the claim that more homogeneous groups may perform better when educational needs of students can be better targeted and addressed by teachers. Indeed, we find coherent results of tracking on graduation

rates. When looking at students that just passed the entry test, we find that about 50 percent completed the advanced track in secondary school, and about the same proportion completed a university degree, irrespective of the permeability of the system. When looking at students just below the passing threshold, we find significant differences between the two systems. While in the permeable system, completion rates of the advanced track and of university are also around 50 percent, the fractions are substantially lower at about 35 percent in the non-permeable system. Thus, while tracking does not seem to be harmful (but also not beneficial) for students that are directly sorted into the advanced track, our results indicate that in the non-permeable system, students just below the passing threshold experience lock-in effects in the lower level track, which is not the case in the permeable system.

Third, the role of parents in supporting their children in their educational efforts in general, and in the context of tracking in particular, is a critical aspect discussed in the literature. For example, Brunello and Checci (2007) argue that family background reinforces the impact of tracking on educational attainment and labor market outcomes. Our results corroborate this finding within the non-permeable system. We show that female students are particularly harmed by the secondary school entry test if they do not pass, and that a high socio-economic status of the parents and working status of the mother can explain why female students below the passing threshold nevertheless acquire a degree from the advanced track in secondary school. The importance of working status of the mother is potentially rooted in differential views towards the role of the woman in the labor market, with higher education a pre-requisite qualification (Dryler 1998, Thornton et al. 1983).

In sum, the results of our study inform the current policy debate about designing educational tracking mechanisms. Most importantly, our findings suggest that the short-term and long-term effects of tracking on educational achievements depend on the type of education system, with a permeable system not being harmful to an individual's success in acquiring advanced level education, but with a non-permeable system imposing significant constraints on students' educational paths. These constraints seem to be particularly binding for female students from lower educational and socio-economic backgrounds, which raises concerns about how tracking might perpetuate educational inequality.

The remainder of the paper is structured as follows. Section 2 provides an overview of the related literature. Section 3 summarizes the most relevant features of the Swiss education system for our study, including the design of secondary school entry tests and the definition and degree of horizontal permeability. Section 4 describes the data and the empirical

methodology. Section 5 presents the main results of our analysis, and section 6 discusses the heterogeneity of the estimated effects by gender. Section 7 concludes the paper.

2. Related Literature

The literature on educational tracking has traditionally been concerned with the design and implementation of tracking mechanisms, and with the short- and long-term consequences of tracking on schooling achievements. Van der Hart (2006) develops a theoretical model that seeks to explain why schools engage in tracking, with school size, type and location, diversity of students' initial achievements, racial compositions and parental involvement as the main determining factors. Epple et al. (2002) show that private schools can select students' abilities indirectly through tuition fees, whereas public schools have an incentive to implement tracking mechanisms to increase their attractiveness for high ability students, resulting in a selection of low ability high income students to private schools.

The theoretical framework for our paper is based on Duflo et al. (2011) and Dustmann et al. (2014), who explain how sorting into high- and low-level tracks can affect student outcomes. The main mechanisms considered here are peer and teacher effects that differ between tracks and potentially stimulate (or discourage) learning outcomes[2], as well as switching costs that depend on the direction of switching (from low- to high-level track, or vice versa) and on the type of education system. In particular, we seek to test the hypothesis whether switching costs spatially differ in a federally-organized education system, reflected in the mobility between tracks. Our empirical model allows us to evaluate the heterogeneity in tracking effects on different stages of the educational path conditional on characteristics of the school system and characteristics of the students.

Early empirical studies report mixed results on the effect of tracking on student performance. Kerckhoff (1986) compares high, middle and low ability students in Great Britain at grouped schools to ungrouped students using lagged test scores to control for initial ability. His results suggest that students in high ability classes do better than the average student at an ungrouped school, and students in low ability classes do worse. Betts and Shkolnik (2000), however,

[2] Zimmer (2003) shows that peer effects are smaller in tracking schools, in particular for low- and average-ability students. Cortes and Goodman (2014) investigate the impacts of ability tracking in algebra classes, with the novel aspect of more intensive and better quality teaching in the lower track, which indeed leads to better short-term and longer-term academic outcomes in that track.

question studies that compare heterogeneous groups in schools with and without formal sorting, due to the difficulty in finding comparable students with the same initial achievements in both types of schools. They argue that schools without sorting may employ indirect measures, such as assigning lower class sizes to low ability students, or better educated teachers to high ability students, and they provide supportive evidence for that.

The two studies point to the empirical challenges in identifying the impacts of tracking on student achievements. Cross-country comparisons of students in school systems that employ tracking, and systems that do not, are often confounded by other differences in the school system that are difficult to control for. Within-country across-school studies typically have the problem that i) students self-select into different types of schools, and ii) schools employ latent sorting practices even though they do not have a formal tracking mechanism in place, as described above (see also Collins and Gan 2013). Finally, within-school analyses must account for the self-selection of students into tracks (Betts 2011).

For this reason, the more recent literature has explored experimental or quasi-experimental designs in which the track status of students is as good as randomly assigned. One branch of the literature looks at the timing of policy reforms within or between countries. For example, Hanushek and Woessmann (2006) compare tracking versus comprehensive education systems in a difference-in-differences (DID) design and find that early tracking increases educational inequality and decreases mean performance. The abolishment of "streaming" towards more comprehensive schooling has been evaluated in several studies. Meghir and Palme (2005) evaluate such a reform in Sweden in the 1950s and find an increase in average schooling beyond the compulsory level, in particular for students with unskilled fathers. Hall (2012) looks at a reform in Sweden in the mid 1990s that increased the academic content of the vocational track. Although the impetus behind the reform was increasing university graduation, they did not change as result of the reform. The reform did however affect dropout rates in the vocational track, which increased for low performing students. Pekkala Kerr et al. (2013) evaluate the abolishment of a two-track system in Finland and find strong effects on test scores of students with low-educated parents. Guyon et al. (2012) explore a natural experiment in Northern Ireland where elite schools were forced to accept more students. This led to an increase in average educational outcomes, indicating a beneficial effect of sorting. Exploring a reform in Romania in the 1970s that postponed tracking of students, Malamud and Pop-Elches (2010, 2011) find increases in advanced track attendance of students from a disadvantaged background, but no effects on higher education attendance or degrees.

Another branch of the literature explores mechanisms that determine tracking status, like date of birth, teacher recommendations, or regional variation in the supply of schools. For example, Juerges and Schneider (2011) look at the impact of teacher recommendations at the end of primary school on the attended track in secondary school. They find that younger students and boys less often get a recommendation for the academic track, which translates into lower enrollment rates in the advanced track. While parts of the age effect are offset during secondary school due to the permeability of the system, the gender effect persists. Schneeweis and Zweimüller (2014) estimate the relative-age effect on secondary school track choice in Austria and find evidence that younger students less likely attend the advanced track in the earlier grades. The relative-age effect persists for students from less favorable backgrounds, but it disappears for students from more favorable backgrounds, which indicates a permeability of the Austrian school system that is differentially explored by individuals from different socioeconomic backgrounds. Van Elk et al. (2011) instrument track choice using regional variation in school supply. The results show negative effects of tracking on graduation from higher education, an effect that is most pronounced for high-ability students and students with a better socio-economic background.

Overall, the evidence so far suggests that tracking has an impact on student achievements, but the direction of effects is ambiguous and depends, among other things, on the exact timing of tracking, the design of tracks, the quality of the schools, and the permeability of the school system. In addition, tracking effects are likely moderated by socio-economic and demographic background characteristics. We contribute to the literature in two dimensions. First, we explore a discontinuity in track attendance induced by a secondary school entry test. We argue that students in the close proximity to the passing threshold have similar abilities but are sorted into different ability groups. Second, we evaluate the consequences of tracking within two educational systems that differ in their horizontal permeability, i.e., in the opportunity to switch tracks, from the advanced track to the vocational track and vice versa, in higher grades, but are otherwise very similar in terms of structure and educational opportunities.

Two papers are closely related to ours. Dustmann et al. (2014) explore a discontinuity in track attendance in Germany induced by the exact date of birth before and after the yearly age cut-off determining the school entry year (see also Muehlenweg and Puhani, 2010). The discontinuity is used as a first stage to analyze the long-term effects of tracking on labor market outcomes. The results show no impact of tracking on earnings, which is attributed to

the permeability of the German education system that allows for up- and downgrading of students at different stages, correcting a possible misallocation through early tracking.

Duflo et al. (2011) develop a theoretical model to explain why tracking may have two simultaneous effects: i) a peer effect through which high achieving students grouped in high ability classes benefit more than those grouped in low ability classes, and ii) a teacher effect through which both high and low achieving students benefit if teacher effort and instruction level are tailored to the students' needs. Duflo et al. (2011) test their model using experimental data from Kenya in which students in randomly selected schools were either grouped randomly or grouped according to prior performance. The results indicate that both peer effects and teacher effects are positive and increase academic performance. While Dustmann et al. (2014) evaluate the long-term effects of tracking, Duflo et al. (2011) look at the short-term effects of tracking on student achievements. Our work complements and goes beyond these two papers in that it i) uses a different natural experiment to identify the effects of tracking, ii) looks at both short-term *and* long-term impacts of tracking on student outcomes, and iii) compares the effects of tracking in horizontally permeable and non-permeable systems.

3. Swiss education system

3.1 General overview

The authority over the Swiss education system is divided among the federal state, the cantons, and the municipalities. The Swiss constitution (art. 62/1) guarantees every canton the right to its own education system while being compliant with the federal legislation. With the Agreement on Education Coordination, enacted in 1970, the Swiss Conference of Cantonal Ministers of Education obtained the legal authority to promote the school system, coordinate education policy at the national level, and harmonize the respective cantonal laws. Still, there exists substantial cantonal variation in school regulations, e.g., in the school entry age, the division of primary and secondary school, and the organization of secondary education.[3]

Despite this heterogeneity, the basic structure of the school system is the same for all cantons. It is given by four main education levels: primary (ISCED 1), lower secondary (ISCED 2), upper secondary (ISCED 3), and tertiary education (ISCED 5; see Figure 1). Primary and lower secondary school are compulsory and free of charge. In most cantons, the former lasts

[3] The report of the Swiss Coordination Centre for Research in Education (SKBF-CSRE 2014) provides an excellent overview of the current system.

six years, the latter three years (some cantons also know the five/four-model). Students in lower secondary school are classified into different skill levels according to their academic performance. While the medium and lower level tracks prepare for different vocational trainings, the advanced track constitutes the basis for general education and university.[4]

— Insert Figure 1 about here —

Upper secondary school consists of vocational training and general education. Vocational training takes place in companies, where students spend at least one day per week in a school attending specialized courses. It lasts two to four years depending on the complexity of the job. General education consists of the "matura "or specialized middle schools and usually takes three to four years (some matura schools also last five years).

The qualification obtained through a matura degree is nationally standardized and allows for free admission to all universities in Switzerland, with the exception of some degrees such as medicine that require an additional entry exam. Universities constitute the first branch of tertiary education, offer a wide range of curricula and are publicly funded. Universities of applied sciences build the second branch of tertiary level education. They offer more practice-oriented degrees for individuals with either a general or a vocational education background, and are a particular feature of the Swiss education system rarely seen in other countries (except for Germany and Austria); see also Herren (2008) and SKBF-CSRE (2014) for details.[5]

After compulsory schooling, about two thirds of students start vocational training, compared to about one third in general education. Since 1995, there is an increasing number of students pursuing a professional matura career (*Berufsmatura*): about 11.8% in 2007 according to the Swiss Federal Statistical Office (2010). This certificate allows entry to universities of applied sciences. There has also been an increase in regular matura degrees, which are a prerequisite to start university (from around 14.9% of total graduates in 1990 to about 19.4% in 2007). As a consequence, the total amount of university students has almost doubled since 1990 (Swiss

[4] Lower and medium level tracks are typically referred to as *Werk-, Real-* and *Sekundarschule* within the definition of ISCED 2 school levels in the Swiss education system, whereas the advanced track is typically referred to as *Vor-* and *Untergymnasium*; see SKBF-CSRE (2014) for details.

[5] Universities of applied sciences exist since the mid 1990s when in the course of a major reform parts of the higher vocational schools (*Hoehere Fachschulen*) were integrated in this new type of tertiary education. We label this level of education uniformly universities of applied sciences.

Federal Statistical Office 2014). Currently, around 18% of the working population hold a university degree which is slightly below the OECD average (SKBF-CSRE 2014).

3.2 Entry tests

Different forms of skill differentiation exist in the Swiss education system to determine the transition from primary to lower secondary school. These include teacher recommendations, grade point averages at the end of primary school, and secondary school entry tests.[6] We focus here on the entry tests in three cantons (Lucerne, Solothurn and St. Gallen) as conducted from 1957 to 1977. This selection was made for three reasons. First, the design of entry tests was very similar in these cantons during that period, as documented by the cantonal school regulations.[7] In particular, entry tests served as the only instrument to differentiate between skill levels and were required for all students who sought access to the advanced track in secondary school. No other criteria were applied, and no other cantons had regulations of similar strictness. Regulations in the three cantons changed in 1978 due to different reforms and since then include other elements than the entry tests in the skill differentiation. Second, in all three cantons the results of the entry examinations were accessible in the cantonal and/or school archives. By official regulations, these had to be conserved for at least 10-15 years and we gained access to entry examinations reaching back as early as 1957. Third, the three cantons provide an interesting case to study because despite their similarities in the cantonal school regulations regarding school structure (on all levels) and the entry examinations, the regulations differed regarding the switching of tracks in higher grades. We will discuss this latter aspect of horizontal permeability in more detail in section 3.3.

Students were usually between 12 and 13 years old (fifth or sixth grade) when taking the entry tests. In case of failure in fifth grade, they could retake the exam in the following year. Entry tests were grade-dependent, i.e., required skills were adapted to the respective grade. The tests covered topics in maths and general language skills. The entry tests were centralized and designed by the cantonal Ministry of Education, sometimes in cooperation with the heads of

[6] The entry tests are referred to as "Sekundar-" or "Gymnasialpruefung" in the Swiss education system.
[7] The discussion of the entry test in this section and the permeability of the education systems in Section 3.3 is based on the following regulations (available only in German, titles translated here into English for reference): *Cantonal School Regulation of the Canton of St. Gallen* (December 12, 1955), *Regulation of the Admission and Promotion in the Cantonal Schools of the Canton of Solothurn* (September 14, 1960), *Regulation of the Admission, Evaluation and Promotion in the Cantonal Schools of the Canton of Lucerne* (June 14, 1957 and March 23, 1972). Copies of the documents in German are available upon request.

secondary schools. Grading was done by external experts and thus could not be manipulated by class teachers, neither from primary nor secondary school. The threshold for passing the exam was determined either by the relative rank in a given school and year to account for potential capacity constraints, or by the absolute number of points. For the former case, we expect no sorting bias by students as the threshold could not be anticipated in advance. Students were typically accepted *ad interim* in the advanced track, usually for about one quarter, and then were accepted permanently given adequate academic performance.

The entry examinations created a discontinuity in the assignment to secondary school that was potentially important for both the length and the quality of education. We expect the classification to be especially sharp up until the late 1970s for three reasons. First, as outlined above, until then the entry test was the unique decision rule for entering the advanced track in secondary school. It was only in the late 1970s when many cantons decided to consider both the test result and pre-test achievements to sort students into different tracks. Second, due to economic constraints we expect that, on average, students were less likely to retake the exam (Krishna et al. 2017). Third, people at the time were less mobile than today in terms of moving between different cantonal systems, and hence the entry test result was very decisive in restricting the direction of secondary (and tertiary) education.

3.3 Definition of permeable and non-permeable systems

While the three cantons had very similar regulations regarding the entry examination at the beginning of lower secondary school (grade 1, ISCED 2), they differed with respect to their regulations for accessing the advanced track in the higher ISCED 2/3 grades. Specifically, the schools considered in the cantons of Solothurn and St. Gallen allowed for later access to the advanced track upon taking a grade-dependent entry exam. The admission to this exam was unrestricted, and our data indicate that the criteria to pass the examination in higher grades did not impose any particular constraints on the access to the advanced track.[8] We call this system a horizontally permeable education system (*System P*) because even though there was

[8] In St. Gallen we observe a passing rate in the entry test at the beginning of secondary school and in higher grades of between 80 and 90% among the participants, which is stable over time. Unfortunately, no systematic data is available for Solothurn, but the information we could access indicates a stable passing rate of almost 70% among the participants at the time. There were no restrictions in the two cantons to taking the entry test in terms of: i) prior school performance (inferred from the school regulations), and ii) teacher recommendations (inferred from the school regulations and from personal interviews with former teachers).

an initial hurdle to access advanced track education, the school system facilitated students' later mobility between tracks, when they performed well.

The advanced track schools in Lucerne also required an entry examination to the higher ISCED 2/3 grades. For the more vocational oriented schools in particular, an entry test adjusted to the student's attended grade needed to be passed, in addition to a certain grade point average that was required in order for a student to be allowed to take the exam. The cantonal school regulations and interviews with former school officials confirm that switching to the advanced track at higher grades were not supported and took place only in individual cases.[9] We call this a non-permeable system (*System NP*) as the hurdle to access advanced track education was present throughout the ISCED 2 and 3 levels, and significant constraints were placed on the mobility of students to move between tracks (at least upwards).

Because of these differences in the cantonal regulations regarding horizontal permeability in secondary school, which are potentially important regarding the impacts of tracking, we do not pool the data but instead consider the two systems separately in our analysis.

4. Data and methodology

4.1 Data sources and data collection

Data for this study stem from two main sources: i) administrative records of the secondary school entry test acquired through school and/or cantonal archives, and ii) targeted survey data with detailed retrospective information on educational paths. The first step of our data collection involved the identification of all schools that conducted the test and stored the exam results for the time period from the early 1960s to the late 1970s. Overall, we identified five school archives with relevant exam information, which were the largest in the respective cantons and therefore cover a majority of the entry examinations that were conducted in the cantons. Data access was supported by the responsible data protection officers in all relevant cantons and allowed us to extract the exam results and address information of former students.

[9] Although there is no systematic data on transition rates available for Lucerne, we could access official letters from the cantonal school administration to the rectorate of the advanced track schools, which indicate that i) students had to take a preparation course during the summer holidays to gain the necessary knowledge for the entry examination (and these courses were taken only by a few students), ii) the number participants in the entry test in general was low with about 16 students on average per year, iii) several of the participants did not pass the entry test (inferred from selectively stored originals of entry examinations and their corrections), and iv) efforts were undertaken since the mid-1970s to ease the transition to the advanced track schools from the other more vocational tracks, indicating that access to the advanced track was very much restricted before.

We also obtained macro information about the test: the passing threshold, the grade at which the exam was taken, and whether there were repeated records for the same student.

We ran several consistency checks of the data in order to maintain the highest possible data quality, and we standardized all exam results relative to the passing threshold for comparability reasons (in half-point steps). Overall, we gained access to 6,814 test results covering the time frame 1957 to 1977. Figure 2 shows three histograms of the standardized test results, in total and by school system. The test results have a slightly asymmetric bell-shaped distribution with negative skewness (-1.4). We do not observe specific test scores with an unusually high or low frequency in the distribution, which would indicate a manipulation of the results in the school entry test if present.

— Insert Figure 2 about here —

In the second step of our data collection, we traced the former students and updated their contact details via alumni networks, telephone registries, online resources and social media. We then conducted a survey based on a computer assisted telephone interview (CATI). We collected the highest completed educational degree (categorical) and asked about details of the educational path, including all steps from primary school to the highest degree.[10] This information is collected retrospectively together with background information at the time of the entry test: working status of the father and mother, educational background of the father, whether the student was living with both parents, the availability of a refrigerator, a car, a freezer, or a washing machine in the household, whether the parents were homeowners, the size of the house/apartment, and whether the household had financial problems. We also collected the individuals' gender, age, civil status, and number of children.

For the telephone survey, we focus on all individuals in the proximity of the passing threshold (+/- 2.5 points). Overall, we successfully completed 1,829 interviews out of 5,020 potential candidates between September 2011 and November 2014. The typical interview duration is 20-25 minutes. Figure 3 below shows the response rate relative to the potential candidates around the passing threshold. The graph indicates that the response rate is almost constant

[10] When contacting the former students, we informed them that the study is about education in Switzerland. To identify our target persons, we asked them whether they took the entry exam in the particular school. We do not expect that this information may have influenced the response behavior regarding their educational outcomes. The questionnaire is available upon request.

over the range of test scores and we do not observe any unusual pattern of the interview rate for particular values in the test score distribution. Thus, there is no evidence that we would have a selective sample regarding the general test score distribution.

— Insert Figure 3 about here —

4.2 Educational outcomes

Educational paths are described with nine outcome variables. The first set of outcomes characterizes the secondary school choice of students. We constructed an indicator that measures whether students went directly to the advanced track after the entry test, as opposed to a vocational track (variable *advanced track direct*), and whether they ever attended the advanced track in secondary school (*advanced track attended*). The two variables allow us to distinguish between direct and indirect or later transitions to the advanced track. To evaluate the success of students in the advanced track, we constructed an indicator that measures whether the student obtained a secondary school degree from this track (*advanced track degree*).

The other outcomes describe the educational path on the tertiary level and from a more comprehensive perspective. As for the secondary school outcomes, we distinguish between a direct transition from secondary school to university (*university direct*), whether the respondent ever attended university (*university attended*), and whether the respondent successfully completed a degree from university (*university degree*). Since higher vocational education as part of tertiary level education is very prominent in Switzerland, we also constructed an indicator whether the respondent completed any degree from a tertiary education institution (*tertiary degree*). Finally, we constructed the *years of schooling* based on the educational paths, only including completed degrees, and complement this measure with the *years of schooling without degree*, capturing the time spent in educational institutions without having completed the respective degree.

Table 1 shows basic summary statistics of the nine educational outcomes in our sample, in total and separately for the two education systems (*P* and *NP*). Our descriptive statistics are confined to those individuals close to the passing threshold. The proportion of students that directly transfer to the advanced track after the secondary school entry test is about 73% and does not differ much between the two systems. However, the proportion of students that ever

attended the advanced track and that received a degree from that track are significantly lower in the non-permeable system, as expected. While there are little differences on the university level between the students that come from the two systems, we find a somewhat higher rate of tertiary degrees for students from the non-permeable system.

4.3 Identification of tracking effects

The objective of our study is to identify the effects of early educational tracking on different short- and long-term educational outcomes. Our identification strategy is based on a sharp regression discontinuity design (Imbens and Lemieux 2008, Lee and Lemieux 2010). The treatment variable (*T*) is whether the respondent passed the secondary school entry test or not, while the running variable (*X*) is the score in the entry test, with threshold τ that determines treatment *T*. Educational outcomes are denoted by *Y*.

The sharp regression discontinuity (RD) estimand Δ_{RD} is defined as

$$\Delta_{RD} = \lim_{x \downarrow \tau} E(Y|X = x) - \lim_{x \uparrow \tau} E(Y|X = x) \qquad (1)$$

where *E(Y|X=x)* denotes the conditional expectation of *Y* given *X*, and the limits are calculated for values of the running variable approaching the threshold from above (first term) and from below (second term). That is, the regression discontinuity estimand of the effect of early educational tracking on educational outcomes compares the average educational outcomes of former students just below and just above the passing threshold. Δ_{RD} can be estimated in a linear regression framework using the following specification

$$Y = \alpha_0 + \alpha_1 T + f^-(X - \tau) + f^+(X - \tau) + \nu \qquad (2)$$

where $f^-(\cdot)$ and $f^+(\cdot)$ are functions in the running variable below and above the threshold. The parameter α_1 measures the average effect of tracking on educational outcomes. The key identifying assumption in this framework is that students just below and just above the passing threshold are comparable in terms of their observed and unobserved background (captured in ν) and the functional form of *Y* in *X* is correctly specified through f. In this case, least squares estimation of α_1 in (2) provides an unbiased estimator of Δ_{RD} and inference on the effects of tracking can be conducted in the usual least squares regression framework.

In a first step towards estimating α_1 in equation (2), we provide a refined set of summary statistics for each educational outcome, comparing individuals who passed and who did not pass the entry test (still confining the analysis to all students in a neighborhood of +/- 2.5

points around the passing threshold). Table 2 indicates that about 93% of those students who passed the entry test directly went to the advanced track, irrespective of the school system. We observe that about 7% (10%) of the students in the non-permeable (permeable) system who did not pass the entry test still went directly to the advanced track in secondary school. This fraction can be explained by (i) students who conducted the entry test at more than one advanced track school and/or gained admission through another test, (ii) students who went to private residential schools with an advanced track, and (iii) measurement error in the retrospective information about educational paths. While we cannot entirely rule out the last explanation, our qualitative data indicate that the first two mechanisms likely explain most of the observed direct transition rates for students below the passing threshold.

— Insert Table 2 about here —

4.4 Validity of the research design

In a second step, before presenting the results of equation (2), we provide some basic checks to support the validity of the RD design (see for example Lee and Lemieux 2010 for details about RD validity checks). The distribution of test scores around the passing threshold is shown in Figure 2. As discussed above, the shape of the distribution (in total and by system) does not indicate a manipulation of test results. There are no unusual jumps or particular values in the test score distribution that would be a sign of systematically lifting students above the threshold. Following the ideas of McCrary (2008), Table A.1 in the appendix reports *p*-values of a test of the null hypothesis that the observed step in the discrete distribution of exam results at the passing threshold is the same as the step we would expect if an underlying continuous data distribution was discretized at the passing threshold (simulated based on repeated draws from a rescaled beta distribution).[11]

[11] In a first step, we bootstrapped the standard errors in the difference between the proportion of observations on the passing threshold and just below the threshold (-0.5 points). In a second step, we normalized the original distribution of exam results to the unit interval and simulated a continuous random variable from a beta distribution with the same mean and standard deviation. This variable was then re-transformed to the scale of the original exam results and values were rounded to multiples of 0.5. In a third step, we calculated the difference (and uncertainty in the difference) between the proportion of observations just below and on the passing threshold for the discretized simulated variable. In a fourth step, we performed a classical mean comparison test of the two differences. The test statistic and p-values of this test are reported in the appendix Table A.1.

As a next check, we consider a series of background variables and test whether students around the passing threshold differ in any of these variables. Figure 4 shows that all covariates are well balanced. This includes the student's gender and mean age at the time of the exam and parent characteristics, such as homeownership; the size of the house; ownership of a car; both parents having Swiss citizenship; father having a higher educational degree; and the mother working outside of the house. The 95% confidence intervals as well as parametric versions of equation (2) with the background variables as left-hand side variables and linear functions f^- and f^+ indicate statistically insignificant jumps at the passing threshold (see Table A.1 in the appendix).

— Insert Figure 4 about here —

Further background variables assessed for systematic differences include whether the parents owned a freezer, washing machine or television, the father was working, and if the respondent lived with both parents (and brothers/sisters) versus single-parent household structures at the time of the exam. None of these variables shows significant jumps at the passing threshold.[12] We also tested the balance of data over the years of observation, but did not find a significant jump at the passing threshold. Thus, overall we do not find evidence in our data that would suggest a systematic and precise manipulation of test scores around the passing threshold and we conclude that the RD design seems valid.

A very different concern for the interpretation of our findings is that regions in which schools adhere to the non-permeable system may be systematically different from regions with the permeable system. Assume, for instance, that the canton Lucerne had a very low share of students who complete an advanced track degree compared to Solothurn and St.Gallen. In this situation, it may be that the entry test in Lucerne has more explanatory power for educational outcomes because students who do not pass the exam may be less willing to invest in the relatively risky advanced track. If this were true, we would interpret the effect of the entry test as the effect of permeability, but in fact it is a consequence of the heterogeneity in the educational system. Note that, if true, this would not invalidate the internal validity of our tracking estimates, but confound the comparison across school system.

[12] These results are available upon request.

To explore the comparability of regions with non-permeable and permeable school system, we leverage census data on a variety of educational indicators in 1979, roughly the middle of our study period. Table 3 presents the results. Overall, we find that regions with permeable schools are fairly similar to the non-permeable schools across a wide range of educational indicators such as the timing of the tracking decision (6 years in both systems), the share of teenagers in vocational training (66% in permeable, 61% in non-permeable schools), the share of matura graduates (7% versus 6%), the share of university graduates (5% in both systems); and the share of teenagers aged 16 to 18 who received state-sponsored educational and professional career counseling (29% versus 35%). Note that in particular the last two outcomes are partially also a reflection of the differences in the permeability of the school system, with lower shares of matura graduates and more demand for counseling to be expected in the non-permeable system.

— Insert Table 3 about here —

Turning to external validity and comparing the non-permeable and permeable schools to the rest of Switzerland, we see that the transition to the secondary track takes place a bit earlier (on average after 5.6 years) and that the share of matura graduates is somewhat higher (10%). All other educational indicators are very similar to the regions in our study sample. In sum, these results suggest the permeable and non-permeable schools in our study are very similar across a wide range of educational indicators. Together with the fact that federal standards regulate the qualification at the end of secondary school and the admission to tertiary education, these findings lend support to the view that the comparison of the different regression discontinuity estimates in permeable and non-permeable systems yields an estimate of the causal effect of permeability.

5. Impact of tracking on educational outcomes

We now turn to the effect of early educational tracking on short- and long-term educational outcomes and the estimation results of equation (2). Figure 5 summarizes the tracking effects for six different educational outcomes by educational system using the parametric RD models

in the bandwidth +/- 2.5 points. All regressions control for school fixed effects.[13] Note that all results are robust to using smaller bandwidths of +/- 0.5 or +/- 1.0 points and non-parametric RD methods (see Table A.2 in the Appendix for a full set of results).

— Insert Figure 5 about here —

System *P* results: Within the permeable system, we find a strong impact of tracking on the direct transition of students to the advanced track in secondary school of about 78 percentage points, starting from about 17% for those just below the passing threshold to about 95% for those just above. For the attendance of the advanced track at any time during the educational path, we find a significantly lower impact of tracking of only about 34 percentage points. The reduction comes from the significantly higher proportion of students below the passing threshold who at some stage attend the advanced track, while the proportion of students above the passing threshold that attended the advanced track does only slightly increase (but from an already high level). Regarding the degree from an advanced track, or the later transition and completion of university education, we do not find an impact of early educational tracking. Table 4 shows the results for three additional outcomes. The effect of passing the entry test on obtaining a tertiary degree is only 3 percentage points, the effect on years of schooling is about 0.32 years, and the effect on years without a degree is about 0.13 years. All three estimates are far away from conventional levels of statistical significance.

— Insert Table 4 about here —

The results confirm our notion of a permeable system, where students have the opportunity to switch between tracks according to their skill levels. Switching between tracks is possible upwards and downwards, i.e., students that perform well in the vocational track after their initial sorting can still achieve university level education, and students that possibly do not perform well in the advanced track can switch to the vocational track. Figure 6 illustrates the impacts of tracking on the educational outcomes in an RD graph, where on the horizontal axis

[13] Note that the results remain unchanged when including year fixed effects. In all regressions, we use heteroscedasticity-robust standard errors, our results are also robust to clustering at the school-year level.

is the standardized test score in points from the threshold, and on the vertical axis is the average educational outcome. The scatters show averages by test score, and the lines are local linear smooths through the observed data. Overall, these graphs confirm our parametric and non-parametric regression results. The graphs also indicate that apart from educational tracking there are skill-based educational achievements as all outcomes increase on average with the score obtained in the entry test.

— Insert Figure 6 about here —

Thus, in the permeable system our results suggest that early educational tracking has a large impact on initial sorting into different tracks, but by the design of the system the differences between students just below and above the passing threshold disappear over the educational path, and there is no evidence of long-term impacts of tracking on educational achievements.

System *NP* results: Within the non-permeable education system we find that early tracking has an even stronger impact on the initial sorting of students into the vocational and advanced track. Our RD estimates indicate that the jump at the passing threshold in the proportion of students that directly transfer to the advanced track is about 81 percentage points, from about 8% below the threshold to almost 90% above. As opposed to the permeable system, early tracking in the non-permeable system has significant long-lasting impacts on educational achievements. The results indicate that the impact of tracking on advanced track attendance is about 41 percentage points (compared to 32 p.p. in the permeable system), and the impact of early educational tracking on the propensity to obtain an advanced track degree is about 18 percentage points. The latter effect is significantly different from the effect measured in the permeable system (p-value 0.016) and translates into a significant difference in university degrees between individuals just below and above the passing threshold. As in the permeable system, we do not find a significant impact of tracking on tertiary level degrees in general, the point estimate is about 6 percentage points but not statistically significant. This indicates that in the non-permeable system individuals in the vocational track still have good chances to achieve a tertiary level degree, but the pathways to such a degree are more restricted and separated (universities versus higher vocational education and training institutions). The observed differences in educational pathways, depending on the secondary school track, lead

to a significant increase in the years of schooling for those individuals just above the passing threshold. We estimate this increase as about 1.2 years of schooling due to the advanced track pathway, which is significantly different from the effect measured in the permeable system (*p*-value 0.029). The effect on passing the entry test on years without degree is close to zero and statistically insignificant.

Figure 7 illustrates the findings of the RD models in graphs. While the local linear smooths confirm our parametric and non-parametric RD estimates, we also observe some differences in the relationship between the educational outcomes and test scores in the non-permeable system compared to the permeable system. First, the skill-based educational achievements in the proximity of the threshold are less pronounced, i.e., the slope of the regression functions is smaller in the non-permeable system. Second, the jump observed at the passing threshold for the direct transition to the advanced track looks very similar in both systems, confirming our notion of comparable secondary school entry tests (Sections 2.2 and 2.3). However, the systems differ in the possibilities to switch between tracks (in particular regarding upwards mobility) generating lock-in effects for individuals below the passing threshold in the vocational secondary school track.[14] This has immediate consequences on secondary school degrees and university attendance, but does not translate in the long-run to a tertiary level degree in general due to the prominence of higher vocational education in Switzerland.

— Insert Figure 7 about here —

6. Heterogeneity in the impacts of tracking on educational outcomes

6.1 Gender-specific effects of tracking

As an extension to our baseline results, we seek to test whether the differential impacts of tracking in the two educational systems are heterogeneous by gender. The literature offers three competing explanations why we might find a gender-specific heterogeneity in the impact of tracking: (i) gender differences in preferences, (ii) different role models and egalitarian values, and (iii) differential parental investments in the education of boys and girls.[15] All three

[14] Note that tests for the difference in the RD estimates between systems *P* and *NP* indicate significant differences for advanced track degree, university degree and the years of schooling.
[15] See Alderman and King (1998) for an overview of the literature on the gender gap in education.

mechanisms suggest that male students are more likely to obtain a higher education than female students, and we briefly discuss each of them in turn.

Among the many preference differences between men and women explored in the literature (see Croson and Gneezy 2009 for an overview), gender differences in risk preferences and competition attitudes seem most relevant in the context of educational choice. Overall, it appears that women tend to be more risk-averse than men, both in risk lotteries as well as high stake portfolio decisions (Croson and Gneezy 2009, Dohmen et al. 2011). Regarding differences in reacting to competitive environments, the literature has provided evidence that men react to competition with extra effort, while women do not (e.g., Gneezy et al. 2003, Gneezy and Rustichini 2004). In addition, previous research has shown that females tend to favor less competitive pay schemes compared to men, a difference that appears to be present already at a young age (Sutter and Glaetzle-Ruetzler 2014).

The sociological literature has provided evidence that parental role models shape educational choice. Dryler (1998) shows that it is primarily fathers who influence the educational choice of sons, while the effect between mothers' occupation/education and daughters' educational choice is less pronounced. Other studies find that parents who are better educated tend to hold more egalitarian values and may strive to ensure that sons and daughters receive equal education (Thornton et al. 1983).

Finally, sociological and economic studies argue that in the presence of gender-based labor market discrimination, it is rational for utility-maximizing families to invest more in sons compared to daughters (Becker 1991, Buchmann and DiPrete 2006). If men are expected to earn substantially more than women, the returns to education are likely to be higher for men, and thus a rational family would invest more in a son than in a daughter with similar ability.

Figure 8 shows the RD estimates of the effect of early educational tracking on educational outcomes by gender.[16] The results indicate that in the permeable system there are no significant differences between males and females. In the non-permeable system, we find larger impacts of tracking for females than males. The gender differences are particularly relevant for the degrees from the advanced track in secondary school (*p*-value 0.015) and from university (*p*-value 0.094), and they translate into a 33% larger increase in the years of schooling for females just above the passing threshold (+1.2 years versus +0.9 years for males), which is however not statistically different between male and female students.

[16] Table A.3 in the Appendix shows that these results are robust to using alternative bandwidths.

— Insert Figure 8 about here —

A striking result in the male and female RD estimates for the non-permeable system is that the proportion of students attending and obtaining a degree from the advanced track in secondary school above the passing threshold is almost the same between both genders. Significant differences between the male and female samples are observed in the proportions of students below the passing threshold and in the impacts of tracking. This result indicates that female students more likely experienced lock-in effects in the vocational track. And only if they passed the secondary school entry test, they had almost equal chances than male students to complete an advanced track secondary school degree. This mechanism is less pronounced at the university level, but still persists, and even though statistically insignificant it can also be observed for degrees on the tertiary level more generally.

6.2 Potential mechanisms of the gender-specific effects of tracking

To better understand the reasons why female students were affected stronger by the results of the secondary school entry test, we examine three background characteristics that may influence the students' secondary school pathway. We look at the father's highest educational degree (tertiary vs. non-tertiary), whether the mother was working at the time of the entry test, and the socio-economic status of the parents. Socio-economic status was constructed based on the sum of seven indicators describing the household at the time of the entry test: whether the parents owned a house, had a car, television, washing machine, freezer, whether at least one parent had a tertiary degree, and whether the family did not have financial problems. A value of at least five (the median) was defined as high socio-economic status.

We then calculated the proportions of these background characteristics in the total sample and separately for females and males, and for compliers, always-, and never-takers. Following the framework of Angrist, Imbens, and Rubin (1996), we define compliance status with respect to the advanced track degree. Hence, we call students that obtained a degree from the advanced track because they passed the entry test and that would not have obtained the degree otherwise compliers. Students that passed the entry test and did not obtain an advanced track degree are part of the never-takers, and students that obtained an advanced track degree but did not pass

the entry test are part of the always-takers. Although we cannot individually identify compliance status, their proportions are known from Table 4 (assuming that there are no defiers, which is a very reasonable assumption in this context). The size of these three latent groups is also displayed in Figure 9, which shows that the female compared to the male subsample features a much higher proportion of compliers (34.8% versus 14.8%), a much lower proportion of always-takers (14.1% versus 38.4%) and a slightly higher proportion of never-takers (51.1% versus 48.8%). Even more informative is to compare the background characteristics of different subgroups of the female and male populations. Using methods proposed in Angrist and Fernandez-Val (2013) we compare the background characteristics of compliers, never-takers and always-takers defined with respect to their advanced track degree.

— Insert Figure 9 about here —

Figure 10 shows the background variables means for all subgroups. Important differences between male and female subpopulations can be observed for compliers and always-takers, i.e., those subgroups that explain the proportion of advanced track degrees just below the passing threshold (always-takers) and the RD estimate (compliers). When comparing male compliers to male always-takers, we observe that father's education (55.8% versus 39.7%), and mother's employment level (40.5% versus 34.3%) is somewhat higher for compliers, while there are no differences in terms of socio-economic status (53.5% versus 55.4%). Among all three background characteristics, male compliers and always-takers score much higher compared to male never-takers. For the female subsample, we find a very different pattern: When comparing female compliers to female never-takers, the former have only slightly higher levels of father's education (33.8% versus 33.1%), mother's employment (44.2% versus 36.1%) and socio-economic status (59.3% versus 54.1%). For the female always-takers, however, we observe a significantly higher probability that the father had a high education level (66.7%), that the mother has been working (69.2%), and that the parents had a high socio-economic status (83.3%).

— Insert Figure 10 about here —

Based on these results, we conjecture that parental background is a decisive factor in supporting students in pursuing their educational career, but has a differential impact for males and females. While male always-takers, which account for more than a third of the male sample, have a background that is fairly similar to male compliers, we find that female always-takers, who account for less than 15% of the female sample, can be characterized by a parental background that is significantly more conducive to educational investment compared to female compliers and never-takers.

Keeping in mind the limitations of our sample size, we conduct three further tests of gender differences in the impact of tracking. First, we test whether the tracking difference between male and female students depends on parental socio-economic status. If the effect of tracking mainly operates through the optimal allocation of family resources, then we should find a larger effect for financially constrained families, possibly related to the quantity-quality trade-off (Becker 1991). Second, we test whether the tracking difference between both genders declines over time. This may be due to a shift towards more egalitarian values or due to decreasing labor market discrimination of women. Third, and related to the previous point, we test whether gender differences depend on the mother's working status at the time of the entry test. If the mother was working, this might be indicative of different attitudes and values towards the role of women in the household and life in general.

Table A.4 in the appendix shows the results of the three subgroup analyses based on the above arguments. We indeed find slightly stronger effects for females from a low socio-economic background than from a high socio-economic background, as hypothesized. Regarding the temporal changes, tracking effects for females seem to be even stronger in the later period (year of the entry test between 1971 and 1977) compared to the earlier period (before 1971). While this result contradicts our hypothesis of a declining discrimination of women in the labor market, it might be explained by the general business cycle in Switzerland and the economy entering a recession in the mid-1970s. As Switzerland historically has been a country with a very low unemployment rate, and effectively a zero unemployment during the period between 1960 and 1975 (Sheldon 2010), one explanation of the slightly stronger tracking effects in the 1971-77 subsample for women could be the uncertainty about the economy after the oil price shocks in the late 1970s and early 1980s, with young women in the vocational track having a bigger incentive to stay in this track and not taking the risk of an uncertain outcome and potentially longer education in the advanced track. Although generally less

strong, the result is confirmed by the male RD estimates in the non-permeable system (see Table A.6 in the appendix).

Finally, we find slightly stronger effects of tracking for female students with a working mother at the time of the entry test. This is in line with the hypothesis that family attitudes towards the capacity of women in acquiring the same education than men, and ultimately being active in the labor market, is a critical factor at all stages of the educational career, for the completion of an advanced track secondary school degree as well as acquiring a university degree. However, it should be noted that none of the subsample differences is statistically significant, and our sample size does not allow us to make stronger conclusions. We therefore interpret these results as indicative of possible mechanisms underlying the impacts of educational tracking, but further research is needed to evaluate them more thoroughly.

7. Discussion and conclusion

Tracking of students into different ability groups is a controversial topic. Proponents of tracking argue that more homogenous groups of students with about the same abilities in one track can be provided with needs-based education, ultimately improving the outcomes of all, students in the lower and in the upper level tracks. Opponents of tracking maintain that equal access to opportunities in educational institutions is a fundamental principle of modern education systems. While previous evidence on the consequences of tracking on educational outcomes is inconclusive, we add to the literature by investigating tracking effects within a country characterized by a federal education system, in which states have authority over the design of their system. This allows us to compare schools that employed a tracking mechanism at the beginning of secondary school, which are comparable in terms of educational indicators, but differed with respect to the possibility to switch between a more vocational-oriented and an advanced track at later grades during secondary school. To this end, we identify two types of systems in Switzerland, a horizontally permeable and a non-permeable education system.

Our results suggest that tracking has a strong effect on immediate sorting of individuals, as expected, and this sorting is independent of the type of education system, confirming our notion of comparable secondary school entry tests in both systems. Using registry data on entrance exams merged with detailed information about educational paths, we are able to compare the long-term effects of tracking in the two types of systems in a regression discontinuity framework. While we find a closing gap between students that just passed or did

not pass the entry test in the permeable system (for advanced track degrees and at the tertiary level), the impacts of tracking are persistent and significant in the non-permeable system. On average in the latter system, the chance of obtaining an advanced track degree in secondary school is about 23 percentage points higher for students that just passed the entry test, and the chance of completing a university degree increases by about 17 percentage points. This persistence of tracking effects translates into about one additional year spent in formal schooling on average. Our results also suggest that female students are affected most by the secondary school entry test. In particular, the share of students who decides to pursue an advanced track degree only after successfully passing the entry test is significantly higher among female than among male students. An important mechanism underlying these gender effects are family attitudes towards the equality of education between men and women that appear to be a major determinant of female students' educational decision.

The results of our study should inform policy-makers in the current debate about implementing and/or reforming educational tracking mechanisms. Most importantly, our results suggest that the short-term and long-term effects of tracking on educational achievements depend on the type of education system, with a permeable system not being harmful to an individual's success in acquiring advanced level education, but with a non-permeable education system imposing significant constraints on students' educational paths. These constraints seem to be particularly binding for students in the lower level track, calling for action to facilitate upward mobility, in particular because family background has been found a decisive factor in supporting students in acquiring higher level educational degrees, bearing the danger of increasing educational inequalities.

References

Abulkadiroglu, A, Angrist J, and Pathak P. (2014): The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1), 137–196.

Alderman H, King EM (1998): Gender differences in parental investment in education. *Structural Change and Economic Dynamics* 9(4), 453-468.

Angrist JD, Fernández-Val I (2013): ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework, In: Acemoglu D, Arellano M, Dekel E (eds), *Advances in Economics and Econometrics.* 1st ed. Vol. 3. Cambridge: Cambridge University Press, pp. 401-434.

Angrist, JD, Imbens G, and Rubin DB (1996). "Identification of causal effects using instrumental variables." *Journal of the American Statistical Association* 91 (434), 444-455.

Becker G (1991): A Treatise on the Family. Cambridge, MA: Harvard University Press.

Betts JR (2011): The economics of tracking in education. In: Hanushek EA, Machin S, Woessmann L (eds), *Handbook of the economics of education Vol. 3*, North-Holland, pp. 341-381.

Betts JR, Shkolnik JL (2000): The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review* 19, 1-15.

Brunello G, Checchi D (2007): Does school tracking affect equality of opportunity? New international evidence. *Economic Policy* 52, 781-844.

Buchmann C, DiPrete TA (2006): The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement. *American Sociological Review* 71(4), 515-541.

Card, D., & Giuliano L. (2016): Can Tracking Raise the Test Scores of High-Ability Minority Students? *American Economic Review*, 106(10), 2783–2816.

Collins CA, Gan L (2013): Does sorting students improve scores? An analysis of class composition. NBER Working Papers No. 18848.

Cortes KE, Goodman JS (2014): Ability-tracking, instructional time, and better pedagogy: The effect of double-dose algebra on student achievement. *American Economic Review* 104(5), 400-405.

Croson R, Gneezy U (2009): Gender Differences in Preferences. *Journal of Economic Literature* 47(2), 448-474.

Dobbie, W, & Fryer, RG (2014): The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools. *American Economic Journal: Applied Economics*, 6(3), 58–75.

Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG (2011): Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *Journal of the European Economic Association* 9(3), 522-550.

Dryler H (1998): Parental role models, gender and educational choice. *The British Journal of Sociology* 49(3), 375-398.

Duflo E, Dupas P, Kremer M (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 C, Puhani PA, Schönberg U (2014): The long-term effects of early track choice. IZA Discussion Paper No. 789; forthcoming in: *Economic Journal*.

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

Gneezy U, Niederle M, Rustichini A (2003): Performance in competitive environments: Gender differences. *Quarterly Journal of Economics* 118(3), 1049-1074.

Gneezy U, Rustichini A (2004): Gender and Competiton at a Young Age. *American Economic Review* 94(2), 377-381.

Guyon N, Maurin E, McNally S (2012): The effect of tracking students by ability into different schools: a natural experiment. *Journal of Human Resources* 47(3), 684-721.

Hall C (2012): The effects of reducing tracking in upper secondary school: Evidence from a large-scale pilot scheme. *Journal of Human Resources* 47(1), 237-269.

Hanushek EA, Woessmann L (2006): Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal* 116, C63-C76.

Herren M (2008): Die nationale Hochschul- und Forschungspolitik in den 1960er- und 1970er-Jahren. Haupt, Bern.

Imbens GW, Lemieux T (2008): Regression discontinuity designs: a guide to practice. *Journal of Econometrics* 142(2), 615-35.

Jackson, CK (2010): Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago. *Economic Journal*, 120(549), 1399–1429.

Juerges H, Schneider K (2011): Why young boys stumble: early tracking, age and gender bias in the German school system. *German Economic Review* 12(4), 371-394.

Kerckhoff AC (1986): Effects of ability grouping in British secondary schools. *American Sociological Review* 51(6), 842-858.

Krishna K, Lychagin S, Robles V (2017): Retaking in High Stakes Exams: Is Less More? *International Economic Review*, forthcoming.

Lee D, Lemieux T (2010): Regression discontinuity designs in economics. *Journal of Economic Literature* 48, 281-355.

Malamud O, Pop-Eleches C (2010): General education versus vocational training: evidence from an economy in transition. *Review of Economics and Statistics* 92(1), 43-60.

Malamud O, Pop-Eleches C (2011): School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics* 95, 1538-1549.

McCrary J (2008): Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics* 142(2), 698-714.

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

Meylan JP (1996): Die Erneuerung des Gymnasiums und die Anerkennung der Maturitaeten. Stationen der Debatte 1968-1995. In: EDK (ed.) Von der "Mittelschule von morgen" zur Maturitaetsreform 1995. EDK, Bern.

Muehlenweg AM, Puhani PA (2010): The evolution of the school-entry age effect in a school tracking system. *Journal of Human Resources* 45(2), 407-438.

Pop-Eleches, C, Urquiola, M (2013): Going to a better school: Effects and behavioral responses. *American Economic Review*, *103*(4), 1289–1324.

Pekkala Kerr S, Pekkarinen T, Uusitalo R (2013): School tracking and development of cognitive skills. *Journal of Labor Economics* 31(3), 577-602.

Schneeweis N, Zweimueller M (2014): Early tracking and the misfortune of being young. *Scandinavian Journal of Economics* 116(2), 394-428.

Sheldon G (2010): Der Schweizer Arbeitsmarkt seit 1920: Langfristige Tendenzen, *Die Volkswirtschaft - Das Magazin fuer Wirtschaftspolitik* 1/2-2010, 15-19.

SKBF-CSRE (2014): *Swiss Education Report 2014*. Swiss Coordination Centre for Research in Education. Aarau, Switzerland.

Sutter M, Glaetzle-Ruetzler D (2014): Gender Differences in the Willingness to Compete Emerge Early in Life and Persist. *Management Science, forthcoming*.

Swiss Federal Statistical Office (2010): Statistisches Jahrbuch der Schweiz 2010. Verlag Neue Zuercher Zeitung, Zuerich.

Swiss Federal Statistical Office (2014): Studierende an den universitaeren Hochschulen: Basistabellen. Neuchatel.

Thornton A, Alwin DF, Camburn D (1983): Causes and Consequences of Sex-Role Attitudes and Attitude Change. *American Sociological Review* 48(2), 211-227.

UNESCO (2000): *The Dakar Framework for Action, Education for All: Meeting our Collective Commitments.* UNESCO, France.

Van Elk R, Van der Steeg M, Webbink D (2011): Does the timing of tracking affect higher education completion? *Economics of Education Review* 30(5), 1009-1021.

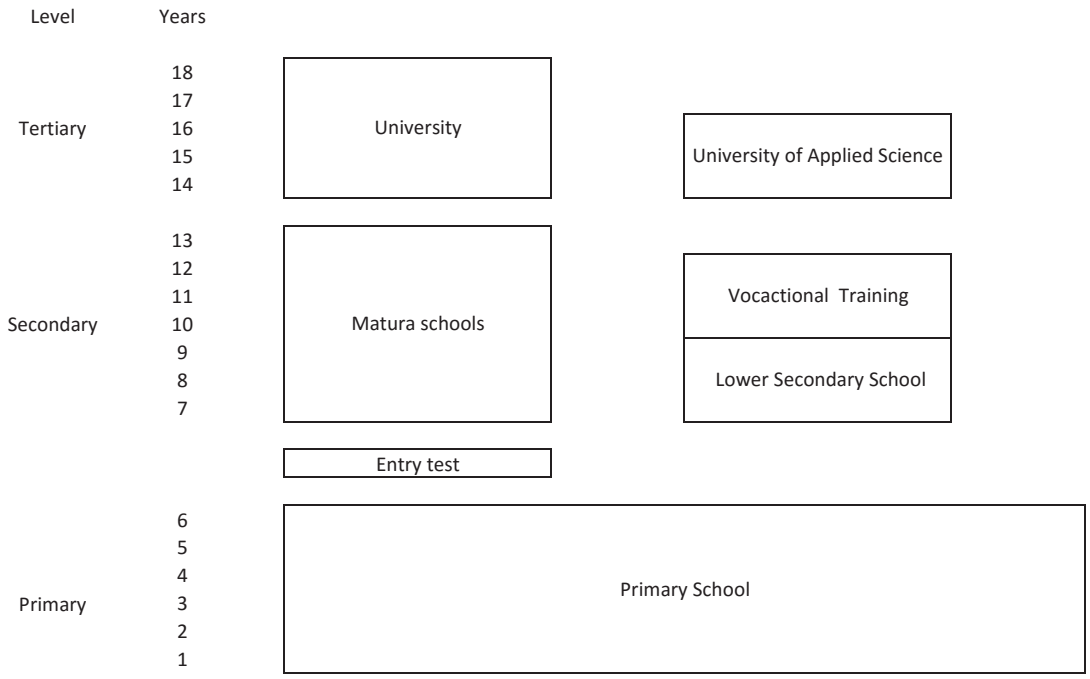
VanderHart, Peter G (2006): Why do some schools group by ability? Some evidence from the NAEP. *American Journal of Economics and Sociology* 65(2), 435-462.

Waldinger F (2007): Does Ability Tracking Exacerbate the Role of Family Background for Student's Test Scores? Unpublished Manuscript. LSE and CEP. http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/waldinger/research/ability_tracking.pdf [last accessed: March 31, 2018].

Zimmer R (2003): A new twist in the educational tracking debate. *Economics of Education Review* 22(3), 307-315.

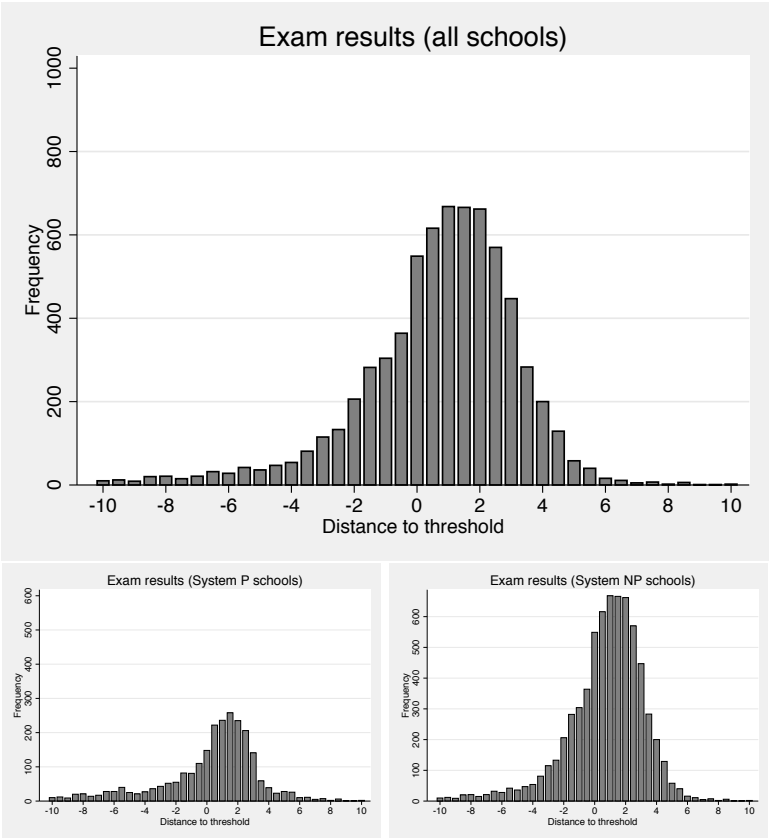
Figures and Tables

Figure 1: Swiss education system



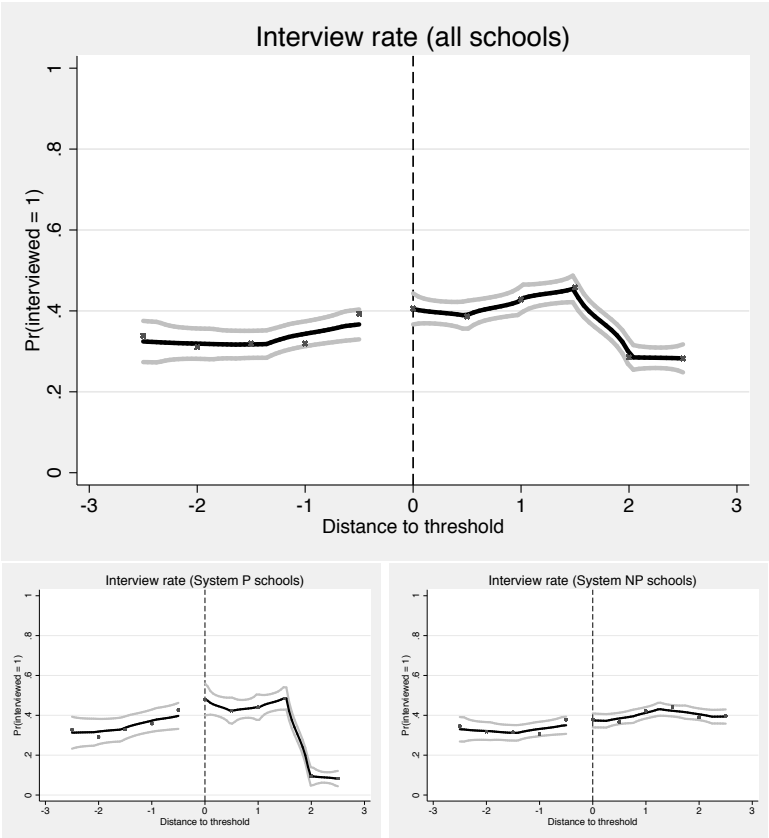
Notes: The figure provides a graphical illustration of the education system in Switzerland. It is a modified version of the figure shown in the yearbook of the Swiss Federal Statistical Office (2010: 355), adapted to the system in the 1970s.

Figure 2: Histogram of exam results



Notes: The top figure shows the distribution of exam results based on the full sample of administrative exam results (6,814 observations). The bottom-left figure shows the distribution of exam results for the subsample of schools with a permeable education system (2,440 observations); the bottom-right figure shows the distribution of exam results for the subsample of schools with a non-permeable education system (4,374 observations).

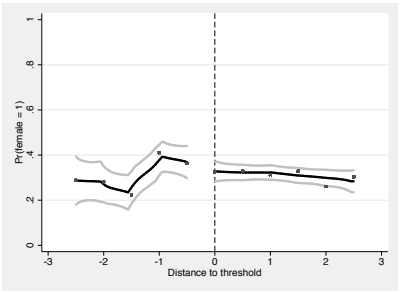
Figure 3: Interview rate



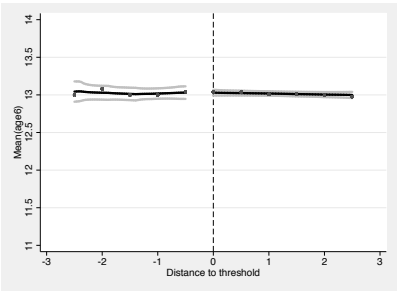
Notes: The top figure shows a non-parametric estimation of the relationship between the distance to the threshold of passing the entry exam and the probability of being interviewed (0/1) with a corresponding 95% confidence interval. The bottom figures show this relationship split by schools in the permeable (bottom-left) and the non-permeable (bottom-right) education systems. These figures are based on the same sample as Figure 2, but restricted to the bandwidth ±2.5 points around the passing threshold (top figure 5,020 observations, bottom left-figure 1,685 observations, bottom-right figure 3,335 observations).

Figure 4: Background variables

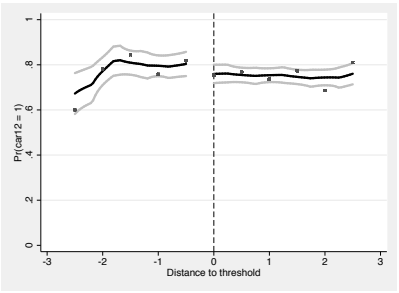
(a) Female



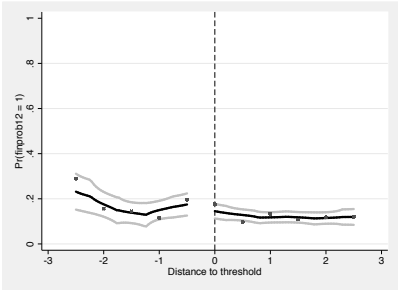
(b) Age at exam



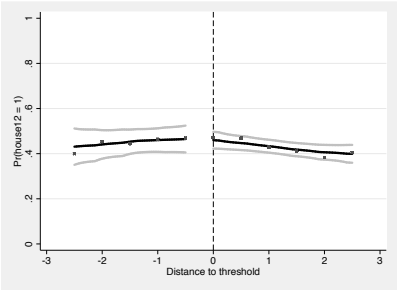
(c) Car



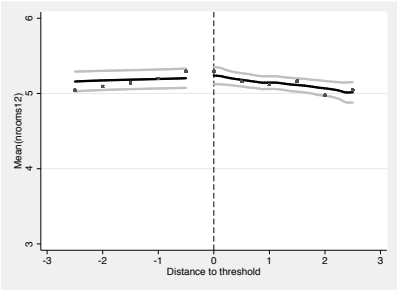
(d) Financial problems



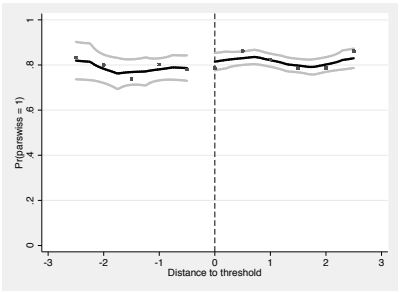
(e) House



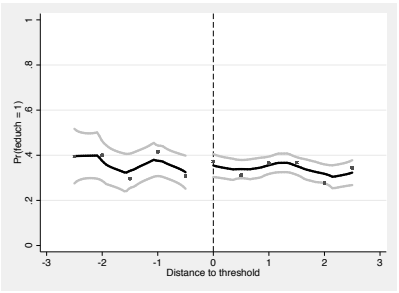
(f) Number of rooms



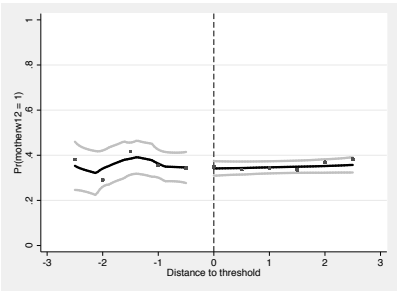
(g) Parents Swiss citizenship



(h) Father university degree

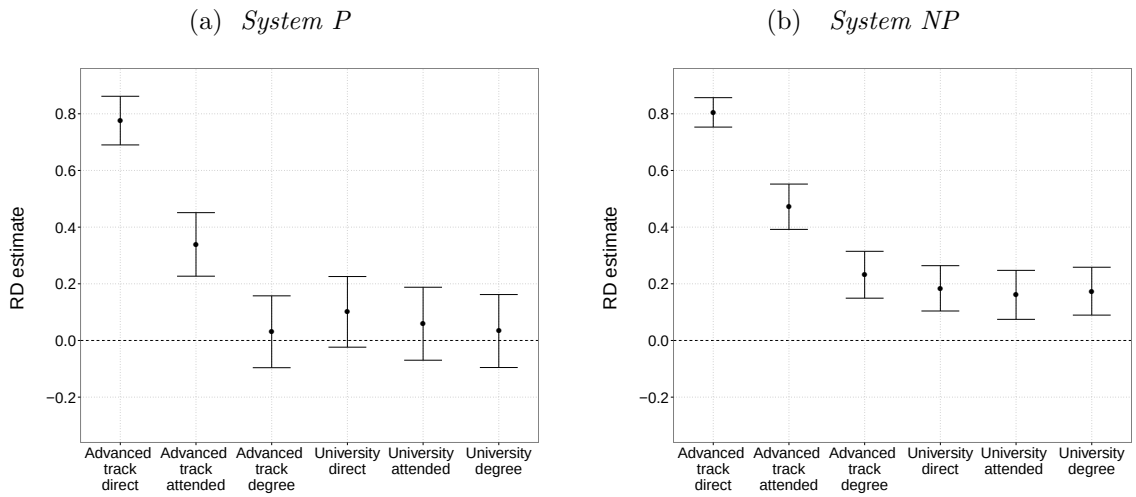


(i) Mother working



Notes: The figure shows a non-parametric estimation of the relationship between the distance to the threshold of passing the entry exam and nine background variables based on retrospective questions about the family situation at age 12. The dots are means calculated separately for each test score value; the black line shows a non-parametric estimate with the corresponding 95% confidence interval. The background variables include an indicator for female gender, age at the time of the exam, car ownership, family financial problems, house ownership, number of rooms in the family home, Swiss citizenship status of the parents, father's university degree, and mother's employment status. The figures are based on the same sample as Figure 2, but restricted to survey participants in the bandwidth ±2.5 points around the passing threshold (1,829 observations).

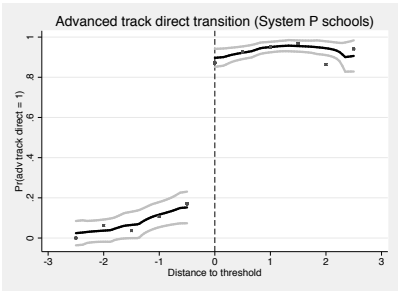
Figure 5: Long-term effects of ecuational tracking



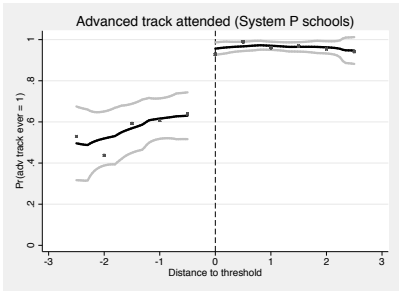
Notes: Both subfigures show the results of six different regression discontinuity estimations using educational outcomes as the dependent variable and a dummy indicating whether a student passed the entry exam as the independent variable. The dots are point estimates with a 95% confidence interval. Figure (a) shows the results for the sample of students in the *system P*; figure (b) uses the sample of students in the non-permeable *system NP*. All estimations use a bandwidth of ±2.5 points. Table A.2 in the Appendix shows the results for all bandwidths, namely ±0.5, ±1.0, and ±2.5. The figures are based on the same sample as Figure 2, but restricted to survey participants in the bandwidth ±2.5 points around the passing threshold (565 observations for system *P*, 1,264 observations for system *NP*).

Figure 6: Discontinuities in educational outcomes at passing threshold – *System P*

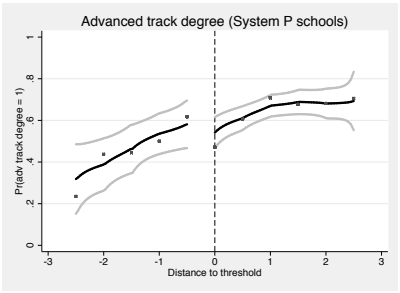
(a) Advanced track direct



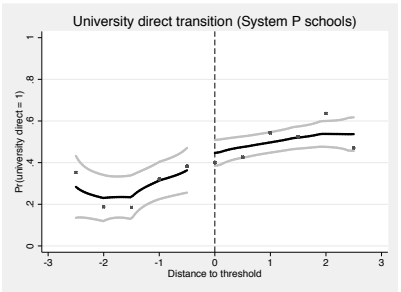
(b) Advanced track attended



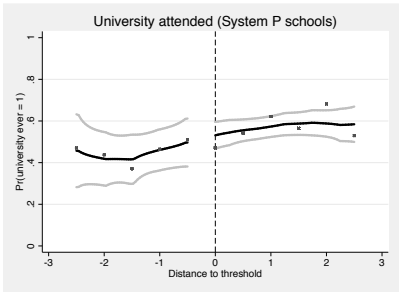
(c) Advanced track degree



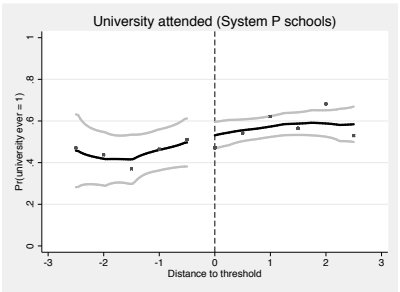
(d) University track direct



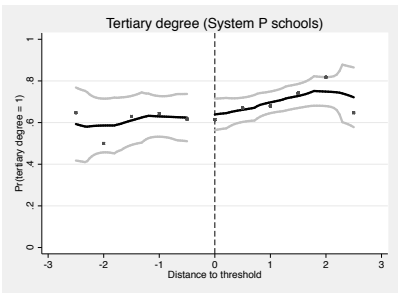
(e) University attended



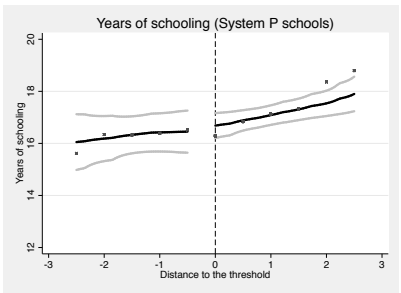
(f) University degree



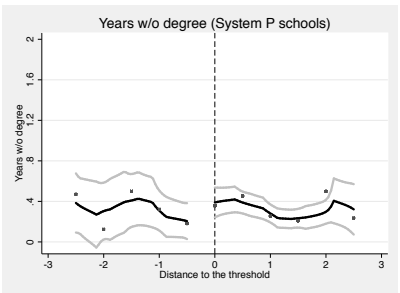
(g) Tertiary degree



(h) Years of schooling



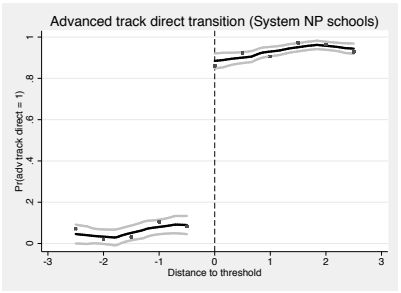
(i) Years without degree



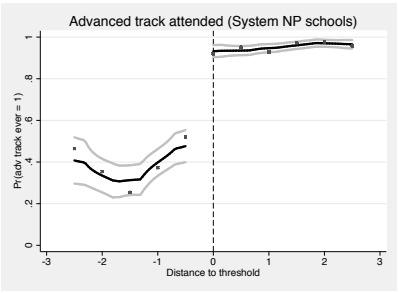
Notes: The figure shows a non-parametric estimation of the relationship between the distance to the threshold of passing the entry exam and nine educational outcomes using data for students in the *permeable* system. The dots are means calculated separately for each test score value; the black line shows a non-parametric estimate with the corresponding 95% confidence interval. Educational outcomes in Panels (a) to (c) include an indicator for a student's direct transition to the advanced track, an indicator for advanced track attendance, and an indicator for advanced track degree. Educational outcomes in Panels (d) to (f) include an indicator for a student's direct transition to university, an indicator for university attendance, and an indicator for university degree. We further depict the results of tracking on an indicator for tertiary degree (g), the years of schooling (h), and the number of years without degree (i). The figures are based on the same sample as Figure 2, but restricted to survey participants in the bandwidth ± 2.5 points around the passing threshold in system P schools (565 observations).

Figure 7: Discontinuities in educational outcomes at passing threshold – *System NP*

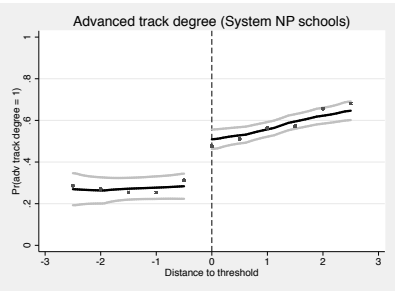
(a) Advanced track direct



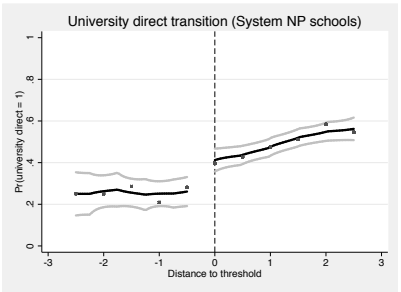
(b) Advanced track attended



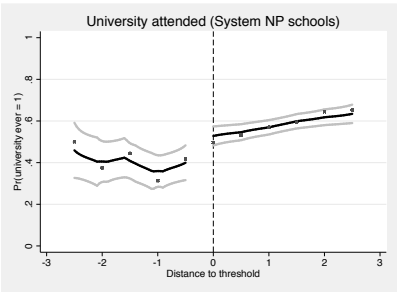
(c) Advanced track degree



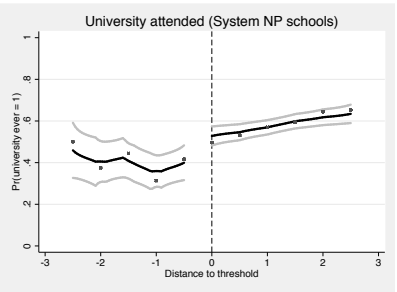
(d) University track direct



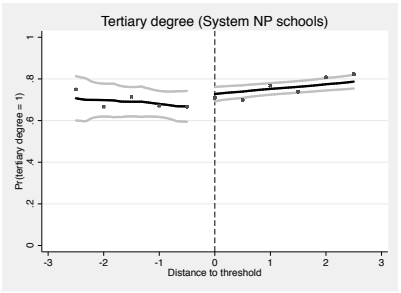
(e) University attended



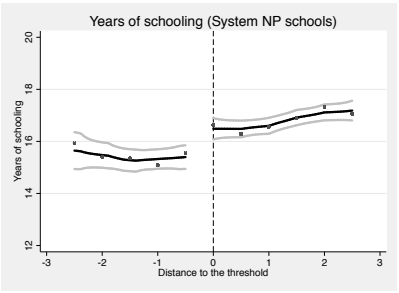
(f) University degree



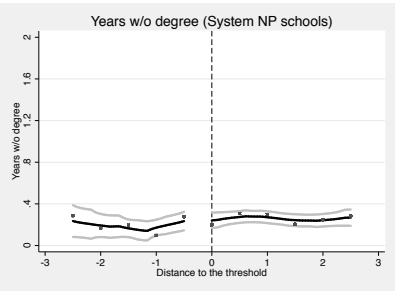
(g) Tertiary degree



(h) Years of schooling

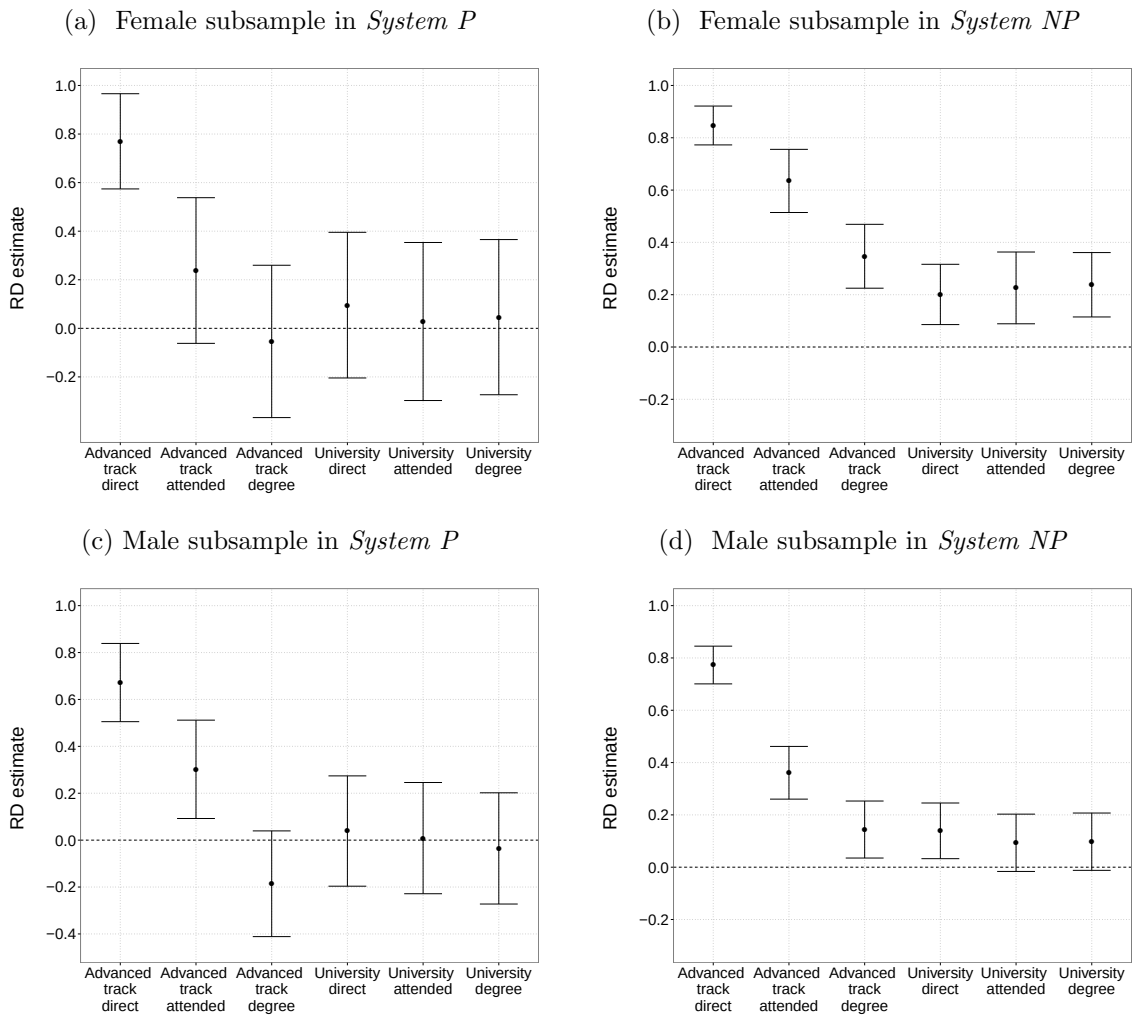


(i) Years without degree



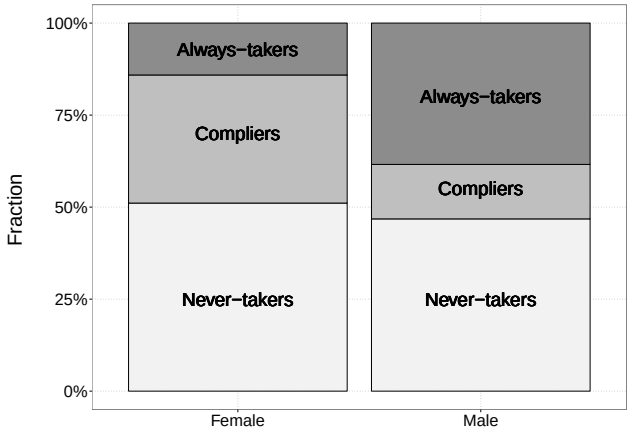
Note: The figure shows a non-parametric estimation of the relationship between the distance to the threshold of passing the entry exam and nine educational outcomes using data for students in the *non-permeable* system. The dots are means calculated separately for each test score value; the black line shows a non-parametric estimate with the corresponding 95% confidence interval. For a description of outcomes, see Figure 6. The figures are based on the same sample as Figure 2, but restricted to survey participants in the bandwidth ±2.5 points around the passing threshold in system *NP* schools (1,264 observations).

Figure 8: Long-term effects of educational tracking by gender



Notes: All four subfigures show the results of six different regression discontinuity estimations using the educational outcomes as dependent variable and a dummy indicating whether a student passed the entry exam as independent variable. The dots are point estimates with a 95% confidence interval. Figure (a) shows the results for the sample of female students in *System P*; figure (b) uses the sample of female students in *System NP*. Figure (c) depicts the results for male students in *System P*; figure (d) depicts the results for male students in *System NP*. All estimations use a bandwidth of ±2.5 points. Table A.3 in the Appendix shows the results for all bandwidths, namely ±0.5, ±1.0, and ±2.5. Figure (a) is based on 204 observations, figure (b) is based on 374 observation, figure (c) is based on 361 observations, and figure (d) is based on 890 observations.

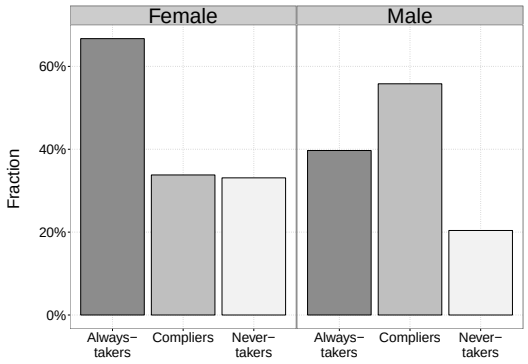
Figure 9: Share of always-takers, compliers, and never-takers in System *NP*



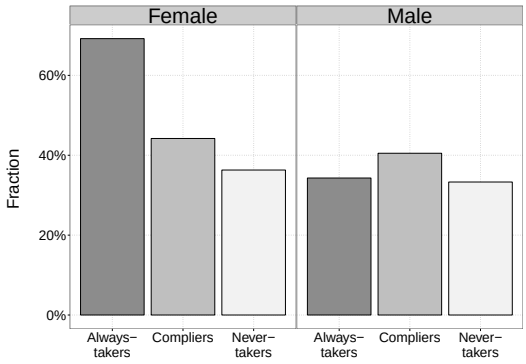
Notes: The figure shows the share of always-takers, compliers, and never-takers for both the female and the male subsample in *System NP*. "Compliers" are calculated as in Angrist and Fernandez-Val (2013) for individuals who passed (did not pass) the entry test and obtained (did not obtain) an advanced track degree, in proximity to threshold (bw±1). "Always-takers" ("Never-takers") based on characteristics for individuals who did not (did) pass the entry test but obtained (did not obtain) an advanced track degree. Table A.5 shows these shares for *System P*. The left figure is based on 197 observations; the right figure is based on 441 observations.

Figure 10: Background characteristics in *System NP*

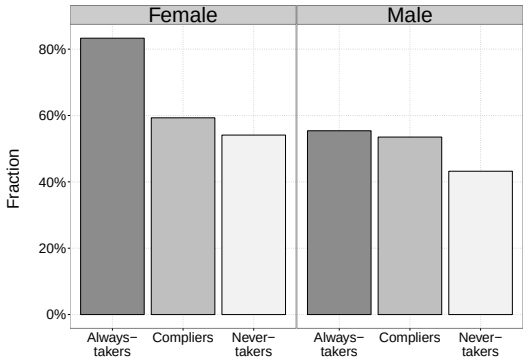
(a) High education level of father



(b) Mother working when student is age 12



(c) High socioeconomic status of parents



Notes: The figure shows background characteristics of always-takers, compliers, and never-takers for the full sample (figure (a)), the female subsample (figure (b)), and the male subsample (figure (c)) in System *NP*. "Compliers" are calculated as in Angrist and Fernandez-Val (2013) for individuals who passed (did not pass) the entry test and obtained (did not obtain) an advanced track degree, in proximity to threshold (bw±1). "Always-takers" ("Never-takers") are based on characteristics for individuals who did not (did) pass the entry test but obtained (did not obtain) an advanced track degree. Background characteristics include father's education level (high vs. low), mother's working status when the student was age 12 (working vs. not working), and parents' socio-economic status when the student was 12 (high vs. low, split at the median). The sample includes 375 women and 889 men.

Table 1: Educational outcomes by school system

	All		*System P*		*System NP*	
	Mean	StdDev	Mean	StdDev	Mean	StdDev
Advanced track direct	72.5%		73.3%		72.2%	
Advanced track attended	83.6%		87.3%		82.0%	
Advanced track degree	53.5%		60.2%		50.6%	
University direct	43.9%		44.6%		43.6%	
University attended	53.9%		53.8%		53.9%	
University degree	50.5%		50.3%		50.6%	
Tertiary degree	72.0%		67.3%		74.1%	
Years of schooling	16.6	3.5	16.9	3.8	16.5	3.4
Years w/o degree	0.26	0.68	0.31	0.78	0.24	0.63
Number of observations	1829		565		1264	

Notes: The table shows the means for nine different educational outcomes for the full sample (column "All"), the permeable system (column "*System P*"), and the non-permeable system (column "*System NP*"). Educational outcomes include the following variables: direct transition from primary school to advanced track as opposed to basic/vocational track, respondent attended advanced track, respondent obtained advanced track degree, direct transition from advanced track to university, respondent attended university, respondent has a university degree, respondent has tertiary degree (ISCED 5A+5B), years of schooling, and the years without degree. The variable "Years of schooling" is constructed from educational paths, only including students who completed their degrees. The variable "Years without degree" is constructed from educational paths, only including students who did not complete their degrees.

Table 2: Educational outcomes by school system and result in entry test

	All		*System P*		*System NP*	
	failed	passed	failed	passed	failed	passed
Advanced track direct	7.6%	92.9%	9.6%	93.3%	6.6%	92.7%
Advanced track attended	45.8%	95.5%	58.5%	96.3%	40.1%	95.1%
Advanced track degree	34.3%	59.6%	48.9%	63.7%	27.8%	57.7%
University direct	27.2%	49.1%	30.4%	49.1%	25.8%	49.2%
University attended	41.9%	57.6%	45.9%	56.3%	40.1%	58.2%
University degree	38.7%	54.2%	42.2%	52.8%	37.1%	54.9%
Tertiary degree	66.4%	73.7%	61.5%	69.1%	68.5%	75.8%
Years of schooling	15.7	16.9	16.3	17.1	15.4	16.8
Years w/o degree	0.23	0.27	0.30	0.31	0.20	0.26
Number of observations	437	1392	135	430	302	962

Notes: The table shows the means for nine different educational outcomes for the full sample (column "All"), the permeable system (column "*System P*"), and the non-permeable system (column "*System NP*"). The column "failed" reports the means for students who failed the entry exam, the column "passed" reports the means for students who passed the exam. For a description of educational outcomes, see notes Table 1.

Table 3: Permeable and non-permeable schools in comparison

	NP Schools	P Schools	Swiss Average
School years at transition to secondary track	6.00	6.00	5.56
Share of teenagers in vocational training	0.61	0.66	0.61
Share of matura graduates	0.06	0.07	0.10
Share of university students	0.05	0.05	0.06
Share of teenagers that received career counseling	0.35	0.29	0.30
Average pupil and student fellowship (in CHF)	19.65	23.30	27.08

Notes: The share of teenagers in vocational training in 1979 uses the resident population aged 16 to 18 as denominator; the share of matura graduates in 1979 uses the resident population aged 19 as denominator; the share of university students in 1979 uses the resident population aged 19 to 27 as denominator. The share of teenagers that received state-sponsored counselling on educational and professional career choices uses the resident population aged 16 to 18 as denominator. The average fellowship is measured in Swiss francs and divided by the resident population. Note that both pupils and university students benefitted from these state-sponsored fellowships.

Table 4: RD estimates of passing the entry test on years of schooling and tertiary degree

	System P	*System NP*
Tertiary degree	*0.617*	*0.667*
	0.0309	0.0613
	(0.0635)	(0.0420)
Years of schooling	*16.52*	*15.55*
	0.321	1.164**
	(0.499)	(0.277)
Years w/o degree	*0.181*	*0.276*
	0.131	0.0714
	(0.0967)	(0.0508)
Number of observations	342	638

Notes: The table shows the results of regression discontinuity estimations using an indicator for tertiary degree (ISCED 5A+5B), the years of schooling, and the years of schooling without degree as the dependent variable, and a dummy indicating whether a student passed the entry exam as independent variable. All estimations use a bandwidth of ±2.5 points, include linear functions in distance below and above the passing threshold, and control for school fixed effects. Mean values (in italics) are calculated for closest observed distance below the passing threshold (-0.5 points). Robust standard errors are in parentheses. Table A.2 in the Appendix shows the results for all bandwidths, namely ±0.5, ±1.0, and ±2.5. Significance levels: * $p < 0.05$, ** $p < 0.01$

Appendix

Table A.1: RD validity checks

	All		*System P*		*System NP*	
	Est.	p-value	Est.	p-value	Est.	p-value
Female	-0.076	0.152	0.071	0.447	-0.109	0.115
Age at exam	0.009	0.902	-0.089	0.476	0.062	0.516
Car	-0.028	0.598	-0.117	0.245	0.016	0.803
Financial problems	-0.003	0.944	-0.031	0.682	0.010	0.848
House	-0.041	0.438	-0.103	0.286	-0.015	0.814
Number of rooms	-0.113	0.435	-0.109	0.647	-0.120	0.511
Parents Swiss citizenship	0.041	0.374	0.056	0.500	0.034	0.542
Father university degree	0.047	0.402	0.067	0.513	0.033	0.628
Mother working	-0.013	0.817	0.050	0.602	-0.039	0.569
McCrary-type test	t =1.38	0.273	t =1.71	0.145	t =1.01	0.381

Notes: The table shows the RD effects of passing the entry exam on nine background variables, based on retrospective questions about the family situation at age 12 (see Figure 4 for the non-parametric estimates and the notes of the figure for a description of the variables). The first column shows the RD estimate; the second column shows the p-value of the test for the null hypothesis that the RD effect equals zero against a two-sided alternative. Results are shown for the overall sample, and split by *System P* vs *System NP*. RD regressions are based on a bandwidth of ± 2.5 points and control for linear regression functions in distance below and above the passing threshold. All estimates control for school fixed effects. For the McCrary-type test, we bootstrapped the standard errors in the difference between the proportion of observations right at the passing threshold (0 points) and just below the threshold (-0.5 points). We then normalized the original distribution of exam results to the unit interval and simulated a continuous random variable from a beta distribution with the same mean and standard deviation. This variable was re-transformed to the scale of the orginal exam results and values were rounded to multiples of 0.5. Then we calculated the difference (and uncertainty in the difference) between the proportion of observations just below and on the passing threshold for the discretized simulated variable. Reported numbers are t-statistics and p-values for a classical mean comparison test of the two differences.

Table A.2: RD estimates of passing the entry test on educational outcomes

	System P			*System NP*		
	bw±0.5	bw±1.0	bw±2.5	bw±0.5	bw±1.0	bw±2.5
Advanced track direct	*0.170*			*0.083*		
	0.726**	0.776**	0.704**	0.807**	0.805**	0.786**
	(0.0601)	(0.0438)	(0.0653)	(0.0336)	(0.0265)	(0.0380)
Advanced track attended	*0.638*			*0.521*		
	0.321**	0.339**	0.276**	0.413**	0.472**	0.434**
	(0.0717)	(0.0572)	(0.0872)	(0.0531)	(0.0408)	(0.0633)
Advanced track degree	*0.617*			*0.313*		
	-0.0577	0.0307	-0.148	0.180**	0.232**	0.172**
	(0.0816)	(0.0647)	(0.0948)	(0.0557)	(0.0422)	(0.0631)
University direct	*0.383*			*0.281*		
	0.0372	0.101	0.0343	0.128*	0.184**	0.131*
	(0.0814)	(0.0636)	(0.0956)	(0.0542)	(0.0408)	(0.0612)
University attended	*0.511*			*0.417*		
	0.00631	0.0591	0.00600	0.0956	0.161**	0.133*
	(0.0835)	(0.0657)	(0.0985)	(0.0583)	(0.0441)	(0.0668)
University degree	*0.489*			*0.354*		
	-0.0199	0.0333	-0.0192	0.131*	0.174**	0.165*
	(0.0835)	(0.0657)	(0.0986)	(0.0569)	(0.0431)	(0.0654)
Tertiary degree	*0.617*			*0.667*		
	0.0322	0.0309	-0.000450	0.0424	0.0613	0.0481
	(0.0809)	(0.0635)	(0.0952)	(0.0546)	(0.0420)	(0.0623)
Years of schooling	*16.52*			*15.55*		
	0.157	0.321	-0.508	0.950*	1.164**	1.132**
	(0.649)	(0.499)	(0.743)	(0.377)	(0.277)	(0.435)
Years w/o degree	*0.181*			*0.276*		
	0.206	0.131	0.199	-0.0173	0.0714	0.0336
	(0.120)	(0.0967)	(0.148)	(0.0735)	(0.0508)	(0.0814)
Number of observations	211	342	565	390	638	1264

Notes: The table shows the RD effects of passing the entry exam on nine educational outcomes split by *System P* vs *System NP*. Columns bw±0.5 and bw±1 show non-parametric RD estimates using mean comparison of outcomes in proximity to the passing threshold (bandwidth ±0.5 or ±1.0 points). Columns bw±2.5 show parametric RD estimates controlling for linear regression functions in distance below and above the passing threshold (bandwidth ±2.5 points). All estimates control for school fixed effects. Mean values (in italics) are calculated for closest observed distance below the passing threshold (-0.5 points). Robust standard errors are in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$

Table A.3: Heterogeneity in RD estimates by gender

	System P				*System NP*			
	Female		Male		Female		Male	
	bw±1.0	bw±2.5	bw±1.0	bw±2.5	bw±1.0	bw±2.5	bw±1.0	bw±2.5
Advanced track direct	*0.111*		*0.167*		*0.047*		*0.121*	
	0.829**	0.770**	0.752**	0.672**	0.847**	0.801***	0.773**	0.765**
	(0.0577)	(0.100)	(0.0584)	(0.0851)	(0.0379)	(0.0569)	(0.0368)	(0.507)
Advanced track attended	*0.630*		*0.625*		*0.297*		*0.566*	
	0.314**	0.238	0.352**	0.302**	0.635**	0.530***	0.361**	0.355**
	(0.0968)	(0.153)	(0.0705)	(0.107)	(0.0616)	(0.0971)	(0.0514)	(0.0796)
Advanced track degree	*0.519*		*0.604*		*0.141*		*0.384*	
	-0.0195	-0.0539	0.0492	-0.186	0.347**	0.292***	0.144*	0.0736
	(0.104)	(0.160)	(0.0803)	(0.115)	(0.0623)	(0.0919)	(0.0556)	(0.0825)
University direct	*0.296*		*0.396*		*0.125*		*0.333*	
	-0.0246	0.0955	0.177*	0.0388	0.201**	0.153*	0.139*	0.0793
	(0.0987)	(0.153)	(0.0813)	(0.120)	(0.0588)	(0.0883)	(0.0543)	(0.0806)
University attended	*0.444*		*0.521*		*0.234*		*0.465*	
	-0.0491	0.0281	0.126	0.0086	0.226**	0.137	0.0933	0.0890
	(0.109)	(0.166)	(0.0819)	(0.121)	(0.0700)	(0.109)	(0.0559)	(0.0832)
University degree	*0.407*		*0.500*		*0.156*		*0.434*	
	-0.0621	0.0460	0.0911	-0.0353	0.238**	0.188**	0.0975	0.100
	(0.107)	(0.163)	(0.0824)	(0.121)	(0.0628)	(0.0978)	(0.0559)	(0.0833)
Tertiary degree	*0.556*		*0.667*		*0.516*		*0.768*	
	-0.0088	0.163	0.0442	-0.0718	0.122	0.0824	0.0008	-0.0123
	(0.109)	(0.164)	(0.0768)	(0.112)	(0.0747)	(0.112)	(0.0480)	(0.0721)
Years of schooling	*15.73*		*16.89*		*14.49*		*15.91*	
	0.188	0.462	0.401	-0.919	1.232**	0.963	0.874*	0.918
	(0.639)	(0.984)	(0.679)	(0.989)	(0.425)	(0.641)	(0.359)	(0.570)
Years w/o degree	*0.500*		*0.083*		*0.219*		*0.192*	
	-0.123	-0.284	0.277**	0.461**	0.120	0.0181	0.0555	0.0441
	(0.218)	(0.367)	(0.0839)	(0.123)	(0.0920)	(0.147)	(0.0632)	(0.100)
Number of observations	133	204	209	361	197	374	441	890

Notes: The table shows the RD effects of passing the entry exam on nine educational outcomes split by *System P* vs *System NP* and by gender. Columns bw±0.5 and bw±1 show non-parametric RD estimates using mean comparison of outcomes in proximity to the passing threshold (bandwidth ±0.5 or ±1.0 points). Columns bw±2.5 show parametric RD estimates controlling for linear regression functions in distance below and above the passing threshold (bandwidth ±2.5 points). All estimates control for school fixed effects. Mean values (in italics) are calculated for closest observed distance below the passing threshold (-0.5 points). Robust standard errors are in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$

Table A.4: Heterogeneity in female RD estimates in *System NP*

	Parental SES at age 12		Before/After 1971		Mother working at age 12	
	High	Low	Before	After	Yes	No
Advanced track direct	*0.000*	*0.030*	*0.080*	*0.026*	*0.000*	*0.056*
	0.873**	0.868**	0.750**	0.889**	0.889**	0.861**
	(0.0490)	(0.0592)	(0.0875)	(0.0392)	(0.0449)	(0.0509)
Advanced track attended	*0.118*	*0.303*	*0.360*	*0.256*	*0.320*	*0.278*
	0.607**	0.785**	0.608**	0.659**	0.580**	0.669**
	(0.0864)	(0.0941)	(0.102)	(0.0778)	(0.104)	(0.0808)
Advanced track degree	*0.118*	*0.182*	*0.120*	*0.154*	*0.200*	*0.111*
	0.310**	0.348**	0.294**	0.361**	0.392**	0.315**
	(0.0904)	(0.118)	(0.104)	(0.0800)	(0.107)	(0.0814)
University direct	*0.059*	*0.182*	*0.080*	*0.154*	*0.120*	*0.139*
	0.115	0.251*	0.252**	0.173**	0.276**	0.144
	(0.0887)	(0.0976)	(0.0903)	(0.0770)	(0.0968)	(0.0807)
University attended	*0.235*	*0.242*	*0.240*	*0.231*	*0.280*	*0.222*
	0.226*	0.194	0.179	0.256**	0.266*	0.183
	(0.0990)	(0.136)	(0.120)	(0.0868)	(0.120)	(0.0925)
University degree	*0.059*	*0.182*	*0.120*	*0.179*	*0.160*	*0.167*
	0.216*	0.292**	0.248*	0.233**	0.329**	0.176*
	(0.0910)	(0.100)	(0.0988)	(0.0813)	(0.103)	(0.0860)
Tertiary degree	*0.647*	*0.455*	*0.520*	*0.513*	*0.480*	*0.556*
	0.200	-0.0505	0.0713	0.138	0.223	0.0348
	(0.106)	(0.145)	(0.133)	(0.0945)	(0.118)	(0.104)
Years of schooling	*14.236*	*14.707*	*14.424*	*14.538*	*14.473*	*14.578*
	1.257*	1.309	0.446	1.607**	2.036***	0.773
	(0.623)	(0.760)	(0.661)	(0.547)	(0.740)	(0.556)
Years w/o degree	*0.235*	*0.288*	*0.340*	*0.141*	*0.220*	*0.236*
	0.0243	0.144	0.0913	0.171	0.0463	0.109
	(0.131)	(0.195)	(0.210)	(0.0930)	(0.128)	(0.135)
Number of observations	98	65	64	133	79	106

Notes: The table shows the RD effects of passing the entry exam on nine educational outcomes for female students split by three background variables that include whether parents had a high socioeconomic status at the time of the exam (columns 2 and 3, split at the median value), whether students took the exam before or after 1971 (columns 4 and 5), and whether a student's mother was working at age 12 (columns 6 and 7). All estimations use a bandwidth of ±2.5 and control for school fixed effects. Mean values (in italics) are calculated for closest observed distance below the passing threshold (-0.5 points). Robust standard errors are in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$

Table A.5: Background characteristics *System P*

	Total	Compliers	Always-takers	Never-takers
A. Overall sample ($N = 565$)	*100.0%*	*3.7%*	*57.3%*	*39.0%*
Father's education level high	40.5%	-	39.7%	31.1%
Mother working when student was age 12	31.1%	-	20.7%	33.1%
Parents' socio-economic status high	45.1%	-	50.0%	42.3%
B. Female sub-sample ($N = 204$)	*100.0%*	*2.9%*	*51.9%*	*45.3%*
Father's education level high	47.5%	-	50.0%	31.7%
Mother working when student was age 12	34.5%	-	23.5%	30.0%
Parents' socio-economic status high	53.0%	-	62.5%	50.9%
C. Male sub-sample ($N = 361$)	*100.0%*	*4.8%*	*60.4%*	*34.8%*
Father's education level high	36.2%	-	35.0%	30.6%
Mother working when student was age 12	29.2%	-	19.5%	35.6%
Parents' socio-economic status high	40.4%	-	44.7%	34.8%

Notes: The table shows the share of always-takers, compliers, and never-takers (in italics) as well as background characteristics of these groups for the full sample (Panel A), the female subsample (Panel B), and the male subsample (Panel C) in *System P*. Column "Total" displays proportions in *System NP* sample (overall and by gender). Column "Compliers" calculated as in Angrist and Fernandez-Val (2013) for individuals who passed (did not pass) the entry test and obtained (did not obtain) an advanced track degree, in proximity to threshold (bw±1). "Always-takers" ("Never-takers") based on characteristics for individuals who did not (did) pass the entry test but obtained (did not obtain) an advanced track degree. Background characteristics include father's education level (high vs. low), mother's working status at age 12 (working vs. not working), and parents' socio-economic status at age 12 (high vs. low, split at the median).

Table A.6: Heterogeneity in male RD estimates in *System NP*

	Parental SES at age 12		Before/After 1971		Mother working at age 12	
	High	Low	Before	After	Yes	No
Advanced track direct	*0.106*	*0.073*	*0.130*	*0.113*	*0.143*	*0.069*
	0.820**	0.803**	0.713**	0.830**	0.779**	0.821**
	(0.0505)	(0.0496)	(0.0565)	(0.0477)	(0.0650)	(0.0402)
Advanced track attended	*0.574*	*0.536*	*0.522*	*0.604*	*0.600*	*0.534*
	0.375**	0.380**	0.382**	0.340**	0.350**	0.389**
	(0.0753)	(0.0808)	(0.0726)	(0.0694)	(0.0871)	(0.0666)
Advanced track degree	*0.425*	*0.366*	*0.260*	*0.491*	*0.400*	*0.414*
	0.131	0.108	0.0816	0.229**	0.130	0.105
	(0.0842)	(0.0843)	(0.0785)	(0.0761)	(0.0981)	(0.0728)
University direct	*0.383*	*0.366*	*0.217*	*0.434*	*0.314*	*0.379*
	0.114	0.0495	0.125	0.179*	0.118	0.0984
	(0.0823)	(0.0857)	(0.0780)	(0.0720)	(0.0940)	(0.0718)
University attended	*0.531*	*0.439*	*0.391*	*0.528*	*0.514*	*0.465*
	0.0362	0.101	0.0775	0.130	-0.00824	0.111
	(0.0825)	(0.0863)	(0.0812)	(0.0763)	(0.0969)	(0.0731)
University degree	*0.489*	*0.439*	*0.391*	*0.472*	*0.485*	*0.431*
	0.0507	0.0667	0.0491	0.162*	0.0191	0.103
	(0.0832)	(0.0863)	(0.0808)	(0.0770)	(0.0964)	(0.0732)
Tertiary degree	*0.681*	*0.804*	*0.739*	*0.792*	*0.771*	*0.741*
	0.0902	-0.0519	-0.0225	0.0392	-0.0541	0.0416
	(0.0775)	(0.0702)	(0.0715)	(0.0612)	(0.0787)	(0.0649)
Years of schooling	*15.965*	*15.830*	*15.719*	*16.086*	*15.674*	*16.057*
	1.005	0.536	0.663	1.192*	1.024	0.625
	(0.596)	(0.458)	(0.489)	(0.520)	(0.577)	(0.494)
Years w/o degree	*0.191*	*0.171*	*0.120*	*0.255*	*0.143*	*0.206*
	0.0417	0.0974	0.147	-0.0280	0.0499	0.0621
	(0.0975)	(0.0878)	(0.0890)	(0.0892)	(0.0891)	(0.0855)
Number of observations	187	189	227	214	135	270

Notes: The table shows the RD effects of passing the entry exam on nine educational outcomes for the male sample split by three background variables that include whether parents had a high socioeconomic status at the time of the exam (columns 2 and 3, split at the median value), whether a student took the exam before or after 1971 (columns 4 and 5), and whether the mother was working when the student was age 12 (columns 6 and 7). All estimations use a bandwidth of ±2.5 and control for school fixed effects. Mean values (in italics) are calculated for closest observed distance below the passing threshold (-0.5 points). Robust standard errors are in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$