Track placement and the development of cognitive and non-cognitive skills

Roxanne Korthals, Trudie Schils, and Lex Borghans

School of Business and Economics, Department of Macro International and Labour Economics, Maastricht University, Maastricht, The Netherlands; Research Centre for Education and the Labour Market, Maastricht University, Maastricht, The Netherlands; Inspectorate of Education, Utrecht, The Netherlands

ABSTRACT

We investigate the effect of being in the high track for secondary school students on cognitive and non-cognitive skill outcomes. Dutch students are assigned to tracks at the end of elementary school based on a test score. We use this test score in a fuzzy regression discontinuity design to exploit the discontinuity in the probability to be assigned to the high track. Our results show that track placement influences cognitive outcomes positively, but leaves students with worse non-cognitive skills. Marginal students therefore face a tradeoff if they could choose the high track over the lower track.

1. Introduction

Tracking in education can be used to tailor education to the capabilities and the needs of students. Educational tracks are usually designed for the average student being assigned to that track and therefore work best for them in terms of learning outcomes. However, some students are at the margin between two tracks and are essentially considered suited for both tracks. In this group, parents and students tend to put in a lot of effort in getting into the high track. From their perspective, a high track is more attractive. They argue that the high track offers higher learning results (i.e. following human capital theory), and that it enables more choice options for future education and labor market prospects (i.e. following signaling theory). Teachers, however, often point to possible negative effects of pushing students into the high track, especially on non-cognitive skills. They argue that for these marginal students in particular, who have a lower ability compared to their peers in the high track, the potential high-track benefits may come at the expense of the development of non-cognitive skills (e.g. self-concept or motivation). The students might have to push harder to be successful in the high track and suffer for this.

The aim of this paper is to estimate the effects of being in the high track for students at the margin. We look at a wide set of outcomes to investigate whether tradeoffs of going to the high track exist for marginal students. We use a longitudinal dataset on cognitive and non-cognitive skill development from primary to secondary education in a Dutch region. At the time our data was collected Dutch secondary schools based the track placement of students on the score on a uniform test at the end of elementary school (6th grade). This allows the comparison of the marginal students to establish the effects of being in the high track, compared to going to the middle track. Yet, in practice there is no sharp discontinuity in the test scores used for track assignment, and
secondary schools differ somewhat in the assignment procedures they adhere to. We therefore apply a fuzzy regression discontinuity design (fRDD; Imbens and Lemieux 2007) with school-specific thresholds to exploit the discontinuity in track placement in identifying its impact on being in the high track for the marginal student. To complement our fRDD, we also use the fact that for several outcomes we have data both at the end of elementary school and in secondary school, three years after the tracking moment. Where such panel data is available, we use it to further reduce possible residual endogeneity.

The results show a positive effect of being in the high track on IQ and on track position in 9th grade, i.e. three years after the tracking decision. We find no effects on other cognitive outcomes, namely reading or mathematics test scores in 9th grade. We do observe some effects on non-cognitive outcomes. Students who are placed in the high track are more competitive in 9th grade and there seems to be some indication that these students are also less emotionally stable and have lower academic self-concept.

This paper contributes to the literature on tracking, but is also closely related to issues on streaming, ability grouping, and selective schools. The literature on the effects of tracking, streaming, and ability grouping is very extensive. Very relevant for this paper is the literature which looks at the effects of a substantial increase in the number of students entering the high track or those looking at the marginal student who moves track. Guyon, Maurin, and McNally (2012) and Van Elk, Van Der Steeg, and Webbin (2011) look at such an increased inflow of students into the high track in Northern Ireland and the Netherlands and find positive effects on outcomes of these students. Duflo, Dupas, and Kremer (2011) find, using an experiment in Kenya in which groups of students were assigned to a school with and without ability grouping, that ability grouping has positive overall effects on cognitive outcomes. The estimates in studies that look at the increased inflow of lower-ability students into the high track capture, besides the tracking effect, also the effects of a changing composition of the high track since more lower ability peers are allowed into the high track.

To isolate the treatment effect of being in the high track on individual students, this paper focuses on the marginal student who does or does not go to the high track. An example of a similar paper is Diris (2012). The results of Diris (2012) suggest that the threshold in the Netherlands for the high track is too high: Students below the threshold would benefit from being in the high track both in terms of test scores and in later earnings. Dustmann, Puhani, and Schönberg (2014) use month of birth as an instrument for track placement in Germany and show, using a reduced form, that month of birth has no effects on labor market outcomes. Pop-Eleches and Urquiola (2013) and Jackson (2010) use formal assignment rules in Romania, Trinidad, and Tobago to instrument attendance of better achieving, or more selective, schools. Both find that students in better schools have higher test scores at the end of secondary school. Jackson (2010) also finds that students in better schools pass more exams and more often earn a certificate that gives access to university, while Pop-Eleches and Urquiola (2013) also look at behavioral aspects and find that better teachers sort into better schools, parents at those schools are more involved, students do more homework, and students’ self-perception is more positive. Vardardottir (2013) looks at peer effects in Iceland due to high ability classes within schools and also uses an fRDD, showing positive effects on school grades of being in the high ability class. However, in this study the high and low ability classes only differ on peers and not on other dimensions.

There is a growing literature which analyzes the relationship between non-cognitive skills, for instance personality traits, self-concept or motivation, and both student performance and later life outcomes (e.g. Heckman and Rubinstein 2001; Ander et al. 2013; Heckman, Pinto, and Savelyev 2013; Kautz et al. 2014). Notwithstanding this growing awareness of the importance of non-cognitive skills, little is known about the effects of being in the high track on the non-cognitive skills for the marginal students. These marginal students are likely to have a lower relative ability compared to their peers in the high track, which might harm their development of non-cognitive skills such as self-concept or motivation. In educational contexts, the Big-Fish-Little-Pond Effect (BFLPE, Marsh 1987), for example, describes the mechanism that students’ academic self-concept is shaped by
the ability level of their reference group (usually their classmates). Studies show that students’ academic self-concept is reduced through social comparison within high-ability reference groups and increased through social comparison within lower-ability reference groups (e.g. Seaton, Marsh, and Craven 2009; Arens et al. 2017; Dumont et al. 2017; Schwabe, Korthals, and Schils 2019).

This paper contributes to our understanding of the effect of tracking in education, and fills particularly the gap in our knowledge on the non-cognitive skill development of marginal students that are assigned to the high track. The structure of this paper is as follows: Section 2 will elaborate on the dataset. The model and results are provided in Sections 3 and 4 respectively. Alternative model specifications are presented in Section 5, and Section 6 concludes.

2. Data

The data used in this paper are the result of a cooperative project with almost all elementary and secondary schools in Zuid-Limburg, a region in the South of the Netherlands. The data comprise a cohort of students that were in 6th grade in 2009 (final grade of elementary school) and in the 9th grade in 2012 (third year in secondary school). Approximately 90 percent of all secondary schools in the region participated in the data collection process, implying almost full coverage of students in 9th grade, also when students move schools. The data assembled in 9th grade include a student survey, mathematics and reading tests, a parent survey, and information from the school’s administration system, including an elementary school exit (ESE) test score and the teacher recommendation for track placement in secondary school. Students take the survey in class and are questioned on their socio-economic background, their non-cognitive skills, and school satisfaction. Filled-in student surveys are available for about eighty percent of the students in the participating schools. In some cases, students do not give reasonable answers to the questions and such surveys are treated as missing values. For about eighty-five percent of the students surveyed in 9th grade, we also have information on their non-cognitive skills assessed at 6th grade, the end of elementary school. Filled-in parental surveys are available for about forty percent of the students, but these are only used to supplement the information on socio-economic background if this is missing for students.

In the Dutch system, students enter the tracked system in the 7th grade, which comprises three main tracks, with some further subdivisions in the low track. In 2011, a little more than fifty percent of the students attended the low track; another 20 percent the middle track and twenty-five percent of students were in the high track (Statistics Netherlands 2012, Figure 1.2.4). In this paper, we use students from the two upper tracks since these tracks have the strongest selection of students. This, together with our focus on the marginal student, yields a sample of 1002 students in the high track and 1242 students in the middle track. In the analyses, we also vary our RDD bandwidth to test how robust our findings are which leads to smaller samples. See more on this in Section 3.

Table 1 shows the descriptive statistics on some key variables, separated for students in the high and the middle track. All variables are described in more detail in Appendix 1. Students in the two tracks differ in some respects: compared to students in the high track, students in the middle track are slightly older, have lower educated parents, have lower cognitive skills (ESE test score, ES teacher recommendation IQ, reading and math test scores), have lower perseverance, lower academic self-concept (e.g. own assessment of school tasks), and are less open. To see whether these differences occur due to selection or due to being in the high track is the goal of this paper.

Acceptance and track placement of students in secondary school is guided by the Dutch government: Using an objective measure of skills (is most cases the ESE test score) and the track recommendation of the primary school teacher, secondary schools decide whether to accept the student and in what track the student will be placed. This shows the high level of autonomy Dutch schools have in international comparison (Hanushek, Link, and Woessmann 2013). Parents have the right to apply for multiple secondary schools and choose their preferred school when more than one school accepts the student. We discuss this last point and its consequences for our study in more detail below.
The national admission guidelines on track placement during the time the data was collected stated that each elementary school was required to send to the preferred secondary school of the student both a track recommendation and an independent and objective measure which is used for track placement (Kingdom of the Netherlands 1981). To obtain this independent and objective measure at the end of the final grade in elementary school about eighty-five percent of schools use the same exit test (the so-called CITO test; CITO 2014). This exit test is multiple-choice and is assessed by CITO, the company that developed the test, not by the teachers. The ESE test has a score which ranges from 500 to 550 and the guidelines from the test agency state that a score of 538 is needed to go to the high track and a score of 533 to go to the middle track (CITO 2014). However, as mentioned, each school sets its own threshold and most schools require higher test scores to go to the high track. The mean test score in the high track in our sample is 546 and for the middle track 541, with considerable variation. In Section 3 the thresholds are discussed more extensively.

For some outcome variables (all measured in 9th grade) we have similar information available from 6th grade, the year before students are tracked. The panel dimension of our data is illustrated in Table 1.

### Table 1. Descriptive statistics of students in 9th grade.

|                | Middle track (havo) | Highest track (vwo) | Diff in means |
|----------------|---------------------|---------------------|---------------|
| **Age**        | 1242                | 1,002               | 1.29***       |
| **Gender**     | 1221                | 992                 | 0.03          |
| **Parental education** | 1054         | 873                 | −0.33***      |
| **ESE test score** | 1242          | 1,002               | −5.10***      |
| **Recommendation** | 1237         | 986                 | −1.93***      |
| **IQ**         | 1039                | 898                 | −0.48***      |
| **Math**       | 773                 | 667                 | −0.36***      |
| **Reading**    | 795                 | 686                 | −0.45***      |
| **Track 9th grade** | 1242         | 1,002               | −0.69***      |
| **Openness**   | 970                 | 784                 | −0.04***      |
| **Conscientiousness** | 969        | 782                 | 0.00          |
| **Extraversion** | 969              | 780                 | 0.01          |
| **Agreeableness** | 970             | 780                 | 0.00          |
| **Neuroticism** | 970                | 782                 | 0.02**        |
| **Perseverance** | 965            | 778                 | −0.02**       |
| **Academic self-concept** | 959        | 776                 | 0.00          |
| **Social self-concept** | 964          | 776                 | 0.00          |
| **School motivation** | 981             | 792                 | −0.04         |

**Notes:** See Appendix A for a full description of all variables.

*Students made a math and a reading test, but not all students had the same questions. To ensure all students receive a test score on the same scale we used IRT to rescale the test scores. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels for the difference between students in the middle and higher track in 7th grade, respectively.

The national admission guidelines on track placement during the time the data was collected stated that each elementary school was required to send to the preferred secondary school of the student both a track recommendation and an independent and objective measure which is used for track placement (Kingdom of the Netherlands 1981). To obtain this independent and objective measure at the end of the final grade in elementary school about eighty-five percent of schools use the same exit test (the so-called CITO test; CITO 2014). This exit test is multiple-choice and is assessed by CITO, the company that developed the test, not by the teachers. The ESE test has a score which ranges from 500 to 550 and the guidelines from the test agency state that a score of 538 is needed to go to the high track and a score of 533 to go to the middle track (CITO 2014). However, as mentioned, each school sets its own threshold and most schools require higher test scores to go to the high track. The mean test score in the high track in our sample is 546 and for the middle track 541, with considerable variation. In Section 3 the thresholds are discussed more extensively.

For some outcome variables (all measured in 9th grade) we have similar information available from 6th grade, the year before students are tracked. The panel dimension of our data is illustrated in Table 2.

### Table 2. Descriptive statistics of students in 6th grade.

|                | Middle track (havo) | Highest track (vwo) | Diff in means |
|----------------|---------------------|---------------------|---------------|
| **IQ**         | 438                 | 531                 | −0.32***      |
| **Openness**   | 408                 | 464                 | −0.08***      |
| **Conscientiousness** | 408          | 463                 | −0.02***      |
| **Extraversion** | 408             | 461                 | 0.01          |
| **Agreeableness** | 408             | 460                 | −0.02**       |
| **Neuroticism** | 408                | 463                 | 0.02*         |
| **Competitive spirit** | 407           | 457                 | −0.03***      |
| **Perseverance** | 407             | 461                 | 0.00          |
| **Cogn. self-concept** | 404        | 459                 | −0.05***      |
| **Social self-concept** | 404          | 457                 | 0.00          |

**Note:** The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The last column shows the $p$-values of the difference between 9th grade and 6th grade for students in the middle and those in the highest track.
in Table 2, which provides descriptive statistics of the same variables as in Table 1, but now measured in elementary school in 6th grade. Again these statistics are separated by track, but at this age the students were still grouped together and the division into middle track versus high track here is therefore merely an illustrative division. In 6th grade, the students who later entered the middle track had lower IQ, were less open, conscientious, agreeable, and perseverant, had lower academic self-concept, and were more neurotic.

Between 6th and 9th grade we observe that personality and school-related measures change for students. This personality change over time is found more often (e.g. Roberts, Viechtbauer, and Walton 2006), but it is yet unclear whether it is due to age differences in personality or due to changing environments over time, for instance entering in a new school or school type.

3. Methodology

3.1. RDD and panel RDD

To estimate the effect of being in the high track on cognitive and non-cognitive outcomes we employ a fuzzy RDD (fRDD), using the ESE test score as our forcing variable. An fRDD assumes that, although the probability to enter in the high track does not jump to 1 after the threshold, the probability increases for larger values of the forcing variable (Imbens and Lemieux 2007). Using the observed threshold in the data for the ESE test score, we instrument track placement in 7th grade by passing the threshold for the test score (Imbens and Lemieux 2007). We estimate the following model:

\[ \text{HighTrack}_i = \alpha_0 + \alpha_1 I(\text{Test}_i \geq \text{Threshold})_i + \alpha_2 \text{Test}_i + \epsilon_i \]  

\[ Y_{i,t=9} = \beta_0 + \beta_1 \text{HighTrack}_i + \beta_2 \text{Test}_i + \epsilon_i \]  

where \( \text{HighTrack}_i \) is an indicator whether the student was placed into the high track in 7th grade and is estimated in Equation (1a). HighTrack\(_i\) is estimated using an indicator function which scores 1 if the test score of the student is above the threshold on the test score, \( I(\text{Test}_i \geq \text{Threshold})_i \). Test\(_i\) is the individual test score, our running variable, which is included linearly. The fitted values from Equation 1a (\( \hat{\text{HighTrack}}_i \)) are used as an explanatory variable in Equation 2a. \( Y_{i,t=9} \) is an outcome variable in grade 9 (for instance a reading test score or a measure of self-concept).

When we have both 6th and 9th-grade data, we make use of a panel version of an RDD by estimating the following two equations:

\[ \text{HighTrack}_i = \delta_0 + \delta_1 I(\text{Test}_i \geq \text{Threshold})_i + \delta_2 \text{Test}_i + \delta_3 Y_{i,t=6} + \nu_i \]  

\[ Y_{i,t=9} = \theta_0 + \theta_1 \text{HighTrack}_i + \theta_2 \text{Test}_i + \theta_3 Y_{i,t=6} + \mu_i \]

Here \( Y_{i,t=6} \) is a control variable in 6th grade which corresponds to the outcome variable in 9th grade, \( Y_{i,t=9} \).

Our main bandwidth is 10 score points on either side of the threshold. In this specification we have at most 2244 students in our data: 1372 before the threshold and 1020 after. As a robustness check, we also show results with bandwidth ±8, ±5, and ±3. A bandwidth of ±1 has too little observations to ensure a strong first stage.

3.2. Estimating the threshold

With an RDD most often researchers make use of a predefined threshold, whether they use a fuzzy or sharp RDD. For instance, Pop-Eleches and Urquiola (2013) and Jackson (2010) use formal assignment rules to instrument selective school attendance. As explained in Section 2, no centralized threshold is set in the Netherlands. Secondary schools are free to decide on the threshold for track placement,
which is their decision and not parents’ decision. Unfortunately, the thresholds of individual schools
are unknown, as is the procedure that schools adopt in selecting their thresholds. We therefore estimate
the thresholds. To do this, we exploit the fact that schools are obliged to base their track place-
ment decision on the ESE test and the ES teacher recommendation. The score of students on the ESE
test is therefore linked to their probability of being accepted to the high track. Following this, the
acceptance record of schools of students with certain scores gives us the option to estimate the
threshold for each school. For instance, if a certain school would have a test score threshold of
545 to allow students in the high track, we should find in the middle track only students with a
score of below 545 and in the high track only students with a score of 545 or above. In practice,
not all schools adhere strictly to their own rule and some schools might not even officially set
such a threshold, but use a more loosely formulated rule and allow themselves to divide from this
rule when they see fit. Still, the underlying concept is the same: We can use the acceptance record of
schools to estimate a school-specific threshold.

More specifically, we use as a threshold for each school the test score for which we find the stron-
gest link between track placement in 7th grade and an indicator function of having a test score
above the threshold. To obtain this threshold we estimate per school the following equation and
use each integer between 500 and 550 as a potential threshold:

\[ \text{HighTrack}_i = \rho_0 + \rho_1 I(\text{Test}_i \geq \text{Threshold})_i + \varphi_i \]  

We then use as the school-specific threshold the integer between 500 and 550 which has the cor-
responding highest F statistic when we estimate Equation (3). Referring back to the example above:
For a school which has an actual threshold of 545, we should find that when we run Equation (3) with
the potential thresholds 500–550, the threshold 545 gives us the highest F statistic. The idea behind
this approach is similar to Porter and Yu (2015), who formulate a method to estimate unknown dis-
continuity points.

Figure 1 shows the estimated school-specific thresholds in our data for schools that offer both the
middle and the high track. The thresholds range from 540 to 548. The figure also shows the schools
for which we cannot estimate a threshold, either because they only offer one type of track or because
they did not provide the data we need. Certain areas of the region are very low populated (see small
dots for the population distribution) and most people live in the Southern part of the region. There-
fore, there exist areas without schools on the map.

An advantage of having to estimate (unknown) thresholds is that, since the thresholds are
unknown, also parents do not know these thresholds and therefore parents and students cannot
target these unknown thresholds. We are therefore not concerned about parents and students

targeting a certain ESE test score. For students, it is always best to maximize their ESE test score,
since they do not know which score would be su-
ffi

dient to get them accepted to the high track. Still, to render more confidence in our results we do perform some relevant balancing tests in Section 3.5.

3.3. Sorting into schools

Although it is the secondary school that decides on track placement, students and parents decide on
the school choice. In theory, when students and parents are dissatisfied with the track assignment of
their preferred secondary school, they could register at another school and try for schools until the
student is accepted to the high track. This opens up the possibility that students with certain charac-
teristics put in more effort to find a school that accepts them to the high track and that these charac-
teristics also cause them to perform differently on our outcome measures. It would then not be track
placement that causes certain outcomes, but the outcomes could be a result of the characteristics
that got them accepted to the high track.

Looking at the available schooling options for students, the notion that students apply to multiple
schools until they are accepted is not far-fetched. The circles around each school in Figure 1
represent the 75th percentile of the distance that students of that school travel. Some schools have a very small circle, while other schools attract students from far away. All schools seem to compete with other schools for students, or said differently most students can choose between at least two schools within a reasonable distance. So, if a student wants to avoid a certain school, for instance because (s)he did not get accepted to the high track, this is often possible, although it might mean that the student has to travel further to school every day.

Although sorting into schools based on their thresholds is possible, we deem it not very likely for the following reasons: First, previous work shows that for secondary school children in the Netherlands the distance to school largely determines school choice (Koning and Van Der Wiel 2013). Corresponding to that, in our sample about half of the students go to the school closest to home, which is on average 3.1 km from the house of the student. Students who do not go to their closest school travel on average 4.1 km. Even though it is about a third longer in distance, the travel time of this extra distance is still quite small. With an average biking speed of 14–17 km an hour (Fietsersbond 2016), students who choose a school further away from home bike approximately 4 min longer on a bike trip of about 11–15 min. So, it could be the case that some students sort into specific schools.

Second, we find little evidence that students go to schools with very different thresholds if they choose a school further away from home suggesting sorting into schools based on the thresholds is not done often. The threshold of the school closest to home is strongly correlated (0.79) with the threshold of the own school, which corresponds with the assumption that most students attend the nearest school. Table 3 shows for the students who do not attend the closest school the threshold of their own school and their closest school. A large share of students is close to the diagonal (shaded cells). Also, schools nearby each other have very similar thresholds as can
be seen in Figure 1. Thirdly, we find no clear evidence that students who travel further to their chosen school are from specific subgroups, as Table 4 shows. Students with higher educated parents, for instance, do not travel more to their chosen school than other students. The background variables we use are the student’s gender, the highest obtained educational degree of the student’s parents, whether the father and the mother work full time or part time in grade 6 (before entry into secondary school), whether the student lived in a two-parent-household in 6th grade, and the IQ of the student in 6th grade.

Still, these arguments do not provide certainty that sorting (of specific subgroups) does not happen, and to account for this possibility we allocate to each student the available school-specific threshold of the school closest to their home. By doing this, we essentially use an intend to treat (ITT)-analysis in our fRDD: we can exclude the ‘non-compliance’ of students who decide to not go to the closest school, but travel to a school further away, potentially with the reason to be allocated to the high track. Equation (1c) shows this adjustment to the model in Equation (1a), and Equation (1d) to Equation (1b).

\[
HighTrack_i = \alpha_0 + \alpha_1 I(\text{Test}_i \geq \text{Threshold}_\text{Closest school})_i + \alpha_2 \text{Test}_i + \epsilon_i \quad (1c)
\]

\[
HighTrack_i = \delta_0 + \delta_1 I(\text{Test}_i \geq \text{Threshold}_\text{Closest school})_i + \delta_2 \text{Test}_i + \delta_3 Y_{i,t=6} + \pi_i \quad (1d)
\]

The amount of noise in our main analyses is potentially substantial. Not only do we use an fRDD, we estimate the unknown threshold for our fRDD and we use an ITT-version of the fRDD to overcome

| Table 3. Frequency table of thresholds for students not going to their closest school. |
|---|
| Threshold of closest school | 540 | 543 | 544 | 545 | 546 | 547 | 548 |
| Threshold of own school | 540 | 86 | 80 | 9 | 132 | 0 | 0 | 1 |
| 543 | 100 | 198 | 17 | 1 | 0 | 0 | 0 |
| 544 | 0 | 0 | 0 | 0 | 0 | 0 | 9 |
| 545 | 45 | 1 | 1 | 30 | 0 | 0 | 0 |
| 546 | 0 | 0 | 0 | 0 | 98 | 15 | 0 |
| 547 | 0 | 0 | 0 | 8 | 36 | 13 | 0 |
| 548 | 0 | 2 | 61 | 13 | 0 | 1 |

Note: The data used for this table consist of all students attending one of the two upper tracks in secondary school.

| Table 4. Distance to own school and background characteristics. |
|---|
| | I | II | III | IV | V |
| Parental education: secondary or vocational | −100.5 | (193.8) |
| Parental education: university of applied sciences | 65.86 | (201.7) |
| Parental education: university | −261.1 | (191.5) |
| Girl | −26.93 | (109.2) |
| Father works parttime | 44.65 | (230.5) |
| Mother works parttime | 112.3 | (147.4) |
| IQ | 30.13 | (92.92) |
| Observations | 1902 | 2186 | 1623 | 1447 | 1138 |
| R-squared | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS models where we regress distance to own school on background characteristics. To be conservative in our estimation we do not cluster the standard errors at the school level. Constant included, but excluded from table. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The reference categories are, respectively: parental education: lower secondary or lower (column I), boys (column II), father or mother working fulltime (columns III and IV). IQ (column V) is a continuous variable.
possible sorting. Luckily though, all this causes us to obtain lower bounds for the true effects: we might wrongly ‘assign’ students who sort into other schools in our first stage to the middle track as opposed to the high track they actually attended, thus lowering the difference in outcomes between students in the middle and high track. The fRDD and estimated threshold lower our power to find any significant differences, although the threshold is chosen in such a way that the discontinuity is the ‘sharpest’ we can get.

### 3.4. Graphical analyses

Figure 2 shows the probability to enter the high track for different distances to the school-specific thresholds of the ESE test score. There is a clear discontinuity in the probability to go to the high track just after the threshold. Just before the threshold the probability is around 35 percent, while for students with a test score at the threshold it is over 60 percent.

Figure 3 shows that this discontinuity can also be seen in some outcome variables in 9th grade. Figure 3(a–c) shows some of the cognitive outcomes in 9th grade. It shows that both for IQ and track level the discontinuity is clearly visible, while for math it is not apparent. All three outcomes also seem very linear around the threshold. Figure 3(d–f) shows some of the non-cognitive outcomes in our analysis. For these outcomes the discontinuity is less clear and the confidence intervals are much bigger.

Figure 4 shows the density of the test score for students entering the top two tracks relative to the threshold of the closest school. There is no large heaping around the school-specific thresholds. This is also not expected since the thresholds are unknown to the student and negotiation with the grader, who is not the own teacher, to obtain a higher grade is not possible, so it is in the best interest of the student to score as high as possible on the ESE test.

### 3.5. Balancing tests

Figure 5 shows figures of background variables over the ESE test score relative to the threshold of the closest schools. The background variables are used in Table 4: the student’s gender, the highest...
obtained educational degree of the student’s parents, whether the father and the mother work full time or parttime in grade 6 (before entry into secondary school), and the IQ of the student in 6th grade. Figure 5 shows that the students are identical around the threshold: around the threshold (‘zero’ in the Figure) no jumps can be seen. Students do seem to be slightly different with respect to the highest obtained parental degree (Figure 5(b)), although the difference is not significant at the 5 percent level. Still to correct for this we run robustness analyses where we control for parental

Figure 3. Discontinuity in outcomes. Note: Graphs of the smoothed values of an Epanechnikov-kernel-weighted local polynomial regression (using the Stata command lpoly). Grey area represents the 95% confidence interval.
education. In Table 5 we show that students above the threshold are not significantly different from those below.

4. Analyses

4.1. First stage results

The first stage shown in Table 6 shows that students with higher test scores are more likely to be in the high track as is to be expected. Our instrument (having a test score of equal or above the threshold of the closest school) is highly significant in predicting track placement in 7th grade. With the different bandwidths the threshold has an $F$-statistic of between 14 and 37, well above the required $F$-statistic of 10 as proposed by Staiger and Stock (1997) and later refined by Stock and Yogo (2005). Depending on the dependent variable in the second stage the sample will change, and subsequently also the corresponding $F$-statistic of the first stage changes. For that reason, all tables will also include the $F$-statistic of the excluded instruments.

Figure 4. Density of forcing variable. Note: A McCrary test is performed using rddensity package in Stata following Cattaneo, Jansson, and Ma (2018). A possible discontinuity around the threshold is rejected using a McCrary test when using the bandwidth of 5, but not while using 10 or 3. For bandwidth 1 the test cannot be performed.

Table 5. Identical students around the threshold.

|                  | Girl | Parent obtained higher education | Father works parttime | Mother works parttime | IQ |
|------------------|------|----------------------------------|-----------------------|-----------------------|----|
| $\text{I(test score} \geq \text{threshold)}$ | 0.214 | 0.132                            | -0.044                | 0.316                  | -0.051 |
| Rel. test score  | -0.034** | 0.054***                         | 0.025                 | -0.034*               | 0.049*** |
| # of students    | 2,213 | 1,927                            | 1,645                 | 1,465                 | 1,151 |
| (Pseudo) $R^2$   | 0.002 | 0.017                            | 0.001                 | 0.001                 | 0.069 |

Notes: The table presents coefficients (robust standard errors in parentheses) from logistic (gender, parental education and employment, household formation) and OLS model (IQ). To be conservative in our estimation we do not cluster the standard errors at the school level. Constant included, but excluded from table. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.
4.2. Stage results

Table 7 shows the results for the cognitive outcomes in 9th grade using OLS and the fRDD approach using different bandwidths. The first column of each set shows the estimated effect of being in the high track in the OLS model, the second column of each set shows the effect of being in the high track in the fRDD models, and the final column per set shows the number of observations and

Figure 5. Discontinuity in the background of student. Note: Graphs of the smoothed values of an Epanechnikov-kernel-weighted local polynomial regression (using the Stata command lpoly). Grey area represents the 95% confidence interval.

4.2. Stage results

Table 7 shows the results for the cognitive outcomes in 9th grade using OLS and the fRDD approach using different bandwidths. The first column of each set shows the estimated effect of being in the high track in the OLS model, the second column of each set shows the effect of being in the high track in the fRDD models, and the final column per set shows the number of observations and
the F statistics of the excluded instruments. The rows show the results for the four different bandwidths: ±10, ±8, ±5, and ±3. Comparing the results across rows therefore gives a good indication of the robustness of our findings. The first outcome of Table 7 is the IQ score in 9th grade: There is a positive effect of being in the high track on this IQ score for the bandwidths 10, 8, and 5. When we use a bandwidth of ±3 we find no significant effect of being in the high track, but it seems to be at least partly due to lack of power due to the lower number of observations. The estimate does point in the same direction. For the mathematics test score, and also for reading test score (not shown), we find no effect of being in the high track. Lastly, Table 7 shows the effect of being in the high track in 7th grade on being in the high track in 9th grade, and as expected, we find a very strong and large effect. Between 7th and 9th grade track mobility is possible, but it seems that the track placement in 7th grade is an important determinant of the track level in 9th grade. Table 7 also shows that with a bandwidth of ±5 and ±3 the first stage becomes very weak and the F statistics drops below 12.

Table 8 presents results for the effects of placement in the high track on non-cognitive skills. We find a very robust effect of being in the high track on competitive spirit: Students who are placed in the high track score between 0.2 sd and 0.4 sd higher on being competitive (comprised of ‘I would like to get high marks’ and ‘Later I want to be good at my job’). For neuroticism and academic self-concept, we only find an effect in the specifications using a bandwidth of ±10 or ±8. For academic self-concept we find the effect also with bandwidth ±6 (not shown), while with neuroticism the significant effect disappears if the bandwidth goes under the ±8. For these two outcomes smaller bandwidths lead to very weak first stages, although the coefficients stay relatively the same. All in all, the non-cognitive outcomes show that students who are placed in the high track in 6th grade are more

### Table 6. First stage results following Equation (1d) for the IQ sample.

| Bandwidth: | 10 | 8 | 5 | 3 |
|------------|----|---|---|---|
| l(Test score ≥ threshold) | 0.256*** (0.057) | 0.208*** (0.056) | 0.173*** (0.043) | 0.187*** (0.031) |
| Rel. test score | 0.042*** (0.003) | 0.050*** (0.003) | 0.057*** (0.007) | 0.051*** (0.011) |
| 6th grade IQ | 0.021* (0.012) | 0.022* (0.012) | 0.034*** (0.011) | 0.060** (0.022) |
| # of students | 969 | 894 | 677 | 462 |
| # of schools | 11 | 11 | 11 | 11 |
| R-squared | 0.264 | 0.328 | 0.392 | 0.460*** |
| F stat instrument | 20.26 [20.26] | 13.88 [18.88] | 16.06 [30.54] | 37.30 [30.54] |

Note: The table presents first stage coefficients (robust and clustered standard errors in parentheses) from IV models using the IQ sample. Constant excluded from table. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

### Table 7. The effects of being in the high track on cognitive outcomes in 9th grade.

| Bandwidth: | OLS | IV | N/[F] | OLS | IV | N/[F] | OLS | IV | N/[F] |
|------------|-----|----|-------|-----|----|-------|-----|----|-------|
| +/-10     | 0.225*** (0.077) | 0.661** (0.327) | 969 [20.26] | 0.174*** (0.079) | 0.661** (0.265) | 1440 [18.88] | 0.460*** (0.114) | 0.326* (0.193) | 2,244 [30.54] |
| +/-8      | 0.235*** (0.077) | 0.783* (0.420) | 894 [13.88] | 0.174* (0.085) | 0.174* (0.349) | 1313 [9.30] | 0.462*** (0.111) | 0.364** (0.145) | 2,057 [16.15] |
| +/-5      | 0.188* (0.085) | 0.944* (0.547) | 677 [16.06] | 0.163 (0.099) | 0.499 1000 [2.46] | 0.468*** 1,579 [5.03] | 0.675*** (0.116) | 0.574** (0.315) | 1,066 [4.73] |
| +/-3      | 0.155 (0.094) | 1.352 (0.879) | 462 [37.30] | 0.128 (0.105) | 1.058 (2.150) | 671 [1.20] | 0.428*** 1,066 [4.73] | 1.089* (0.113) | 0.659 [4.73] |

Grade 6 control ✓ ✓ na na

Notes: The table presents the coefficients (robust and clustered standard errors in parentheses) for being in the high track from OLS and IV models estimating the effect of being in the high track on cognitive outcomes. Each cell represents a single regression. Constant and running variable excluded from table. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.
competitive in 9th grade and there seem to be some indications that these students are also more neurotic and have lower academic self-concept.16

In many cases the OLS estimates in Tables 7 and 8 are lower than the IV estimates, while we would expect that the direction of the endogeneity implies that the OLS estimates are larger than the IV estimates. However, the OLS estimates are within the confidence intervals of the IV estimates, which are themselves quite large due to the noise added in our first stage. Still, even though the estimates are not significantly different from each other, it remains counterintuitive. The reason that the OLS and IV estimates differ is because the IV-estimates capture the local average treatment effect for those students affected by our instrument while the OLS estimates depict average differences, controlling for the forcing variable (Lee and Lemieux 2010). The downward bias of the OLS estimates for school motivation and agreeableness could be explained by the role the forcing variable takes: In the IV models we see that removing the endogeneity of being in the high track shifts part of the effect of the forcing variable (as seen in the OLS models) to the dummy for being in the high track. We correct this using an IV-approach.

5. Alternative model specifications

To come to our main model specification, we have chosen to not use some, at first sight, simpler methods. For instance, (1) we could have not used the panel dimension of our data, (2) we could have used one common threshold for all schools instead of school specific thresholds, and (3) we could have used the threshold of the own school instead of the threshold of the closest school. In this section we show how these alternative model specifications would have changed our results. In general, the results here are similar to our main results, but less precise. Comparing results when using the threshold of the own school with the threshold of the closest school suggests that some sorting indeed took place and the main model specification removed more endogeneity.

In the main models presented in Section 4 we removed residual endogeneity by making use of the panel dimension which is available for some of our data. Table 9 shows what happens to our estimates when we do not use this panel dimension. We find very similar estimates as before and also the F-statistics of the excluded instrument are very similar. For neuroticism the coefficients are also significant for the ±10 and ±8 bandwidths, just like before. However, the coefficients in Table 9 do not reach the same level of significance as before for IQ and academic self-concept.

In our main model we used school-specific thresholds. We could have also used one general threshold. As a robustness check, we estimated this overall threshold using data from all schools combined. Tables 10 and 11 show the results when we use this threshold for our analyses. In these models sorting of students due to school choice is excluded since we do not let the threshold vary by school, but the model is less precise than the main model since it uses less information.

| bandwidth | Competitive spirit | Neuroticism | Cognitive self-concept |
|-----------|--------------------|-------------|------------------------|
|           | OLS | IV | N/F | OLS | IV | N/F | OLS | IV | N/F |
| +/−10     | 0.003 | 0.189* | 864 | 0.023 | 0.148* | 871 | 0.011 | −0.072* | 863 |
|           | (0.013) | (0.114) | [14.31] | (0.021) | (0.088) | [16.53] | (0.007) | (0.040) | [14.34] |
| +/−8      | −0.000 | 0.262* | 796 | 0.017 | 0.191* | 802 | 0.010 | −0.101** | 795 |
|           | (0.012) | (0.152) | [9.88] | (0.022) | (0.108) | [10.47] | (0.007) | (0.047) | [9.62] |
| +/−5      | −0.006 | 0.386* | 603 | 0.015 | 0.138 | 609 | 0.010 | −0.090 | 604 |
|           | (0.011) | (0.231) | [7.12] | (0.021) | (0.143) | [8.62] | (0.006) | (0.061) | [7.09] |
| +/−3      | −0.010 | 0.344** | 404 | 0.032 | 0.144 | 410 | 0.015 | 0.017 | 407 |
|           | (0.014) | (0.171) | [23.24] | (0.020) | (0.124) | [27.98] | (0.009) | (0.065) | [21.70] |

Notes: The table presents the coefficients (robust and clustered standard errors in parentheses) for being in the high track from OLS and IV models estimating the effect of being in the high track on non-cognitive outcomes. Each cell represents a single regression. Constant and running variable excluded from table. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.
Tables 10 and 11 suggest that, except for the track in 9th grade, being in the high track does not affect the displayed outcomes using this model.

As a third robustness check we perform our analyses using the threshold of the own school, instead of the closest school. Tables 12 and 13 show results using the threshold of the own school.
school, rather than using the threshold of the closest school. As discussed in Section 3.3 this threshold is likely endogenous since school choice is free and thus students could sort into schools based upon the threshold. Tables 12 and 13 show that when we use the threshold of the own school, we find no significant effects of being in the high track on cognitive and non-cognitive outcomes. The estimated coefficients are most often in between the OLS estimates and the RDD estimates, suggesting that using the threshold of the own school is to some extent endogenous.

### Table 12. The effects of being in the high track on cognitive outcomes in 9th grade – threshold of own school.

| Bandwidth: | IV N/[F] | IV N/[F] | IV N/[F] |
|------------|----------|----------|----------|
| +/-10      | 0.386*   | 969      | 0.099    | 1440     | 0.394**  | 2244     |
|           | (0.222)  | [51.98]  | (0.232)  | [81.38]  | (0.177)  | [140.10] |
| +/-8       | 0.327    | 894      | 0.043    | 1313     | 0.441*** | 2057     |
|           | (0.254)  | [36.83]  | (0.282)  | [57.56]  | (0.164)  | [104.90] |
| +/-5       | 0.539**  | 677      | 0.063    | 1,000    | 0.635*** | 1579     |
|           | (0.266)  | [49.28]  | (0.368)  | [47.06]  | (0.184)  | [59.47]  |
| +/-3       | 0.685**  | 462      | 0.017    | 671      | 0.626**  | 1066     |
|           | (0.341)  | [65.04]  | (0.482)  | [31.91]  | (0.271)  | [58.41]  |

Grade 6 control ✓

Notes: The table presents the coefficients (robust and clustered standard errors in parentheses) for being in the high track from OLS and IV models estimating the effect of being in the high track on cognitive outcomes. Each cell represents a single regression. Constant and running variable excluded from table. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

### Table 13. The effects of being in the high track on non-cognitive outcomes in 9th grade – Threshold of own school.

| Bandwidth: | IV N/[F] | IV N/[F] | IV N/[F] |
|------------|----------|----------|----------|
| +/-10      | 0.091    | 864      | 0.050    | 871      | 0.004    | 863      |
|           | (0.070)  | [65.45]  | (0.073)  | [58.30]  | (0.029)  | [62.56]  |
| +/-8       | 0.107    | 796      | 0.063    | 802      | 0.004    | 795      |
|           | (0.076)  | [47.75]  | (0.086)  | [41.02]  | (0.025)  | [41.80]  |
| +/-5       | 0.102    | 603      | 0.089    | 609      | −0.007   | 604      |
|           | (0.091)  | [89.30]  | (0.070)  | [62.92]  | (0.036)  | [82.03]  |
| +/-3       | 0.128    | 404      | 0.167**  | 410      | 0.020    | 407      |
|           | (0.097)  | [88.91]  | (0.077)  | [90.85]  | (0.045)  | [83.83]  |

Grade 6 control ✓✓✓

Notes: The table presents the coefficients (robust and clustered standard errors in parentheses) for being in the high track from OLS and IV models estimating the effect of being in the high track on cognitive outcomes. Each cell represents a single regression. Constant and running variable excluded from table. The superscripts *, **, and *** indicate significance at the 10%, 5, and 1% levels, respectively.

6. Conclusion and discussion

In many countries, children are placed into tracks after the transition from primary to secondary education. Students that are on the margin between two tracks, as well as their parents, often tend to focus on getting into the high track, due to expected positive learning outcomes and signaling effects. Teachers, however, are sometimes more hesitant to do so because they believe it might hamper the students’ non-cognitive development if put in a too high track. So far there is little evidence for this presumed tradeoff between cognitive and non-cognitive skill development. In this paper we look at exactly this tradeoff and include both cognitive and non-cognitive outcomes in the analysis, since being in the high track might affect these outcomes differently.

We use a fuzzy regression discontinuity design to investigate the skill development of placement in a high track for the marginal student using Dutch data. In the Dutch case, the score of a uniform elementary school exit test is used by secondary schools to decide on track placement for students in 7th grade. We use these test scores to estimate thresholds to look at students around the
discontinuity in track assignment. We complement this fRDD with panel data to further exclude any remaining endogeneity due to our fuzzy design. Students around the threshold for going into the high track are likely to be similar, but they are assigned to different tracks.

Having removed the endogeneity of track placement, we find that most of the significant differences between students in the middle and high track are due to selection and not due to track placement. However, there are some positive and some negative effects indicating that a tradeoff does seem to exist for the marginal students. Track placement in the high track positively influences a number of cognitive outcomes: When marginal students are placed in the high track in 7th grade, they have higher IQ scores in 9th grade. We also show that when students are placed in the high track, they are more likely to (still) be in the high track in 9th grade. Since these marginal students at the beginning of tracking are more likely to still be in the high track in 9th grade, these students do seem to be able to cognitively cope with the higher demands from school. In addition to these positive effects on cognitive outcomes, we observe effects on some non-cognitive outcomes. We find that neuroticism, and competitive spirit are increased by being in the high track, while academic self-concept is lowered. Marginal students who are placed in the high track are the worst-performing students in the class. This can contribute to the competitive nature of these students, while also making them insecure about their cognitive abilities and raising their anxiety levels. The lower academic self-concept for those students at the lower end of the ability distribution in a group is referred to as the Big-Fish-Little-Pond Effect (BFLPE, Marsh 1987) and commonly found in studies (e.g. Seaton, Marsh, and Craven 2009; Arens et al. 2017; Dumont et al. 2017; Schwabe, Korthals, and Schils 2019).

We conclude that striving for being in the high track is not necessarily a good thing. There is a tradeoff between positive cognitive effects and some negative non-cognitive effects. Students and parents should be aware of this when they strive for the high track, and so should teachers when they consider track placement of students. One limitation is of course that we did not look at the longer-term effects of being in the high track, nor which of the two sets of skills are more important for later life outcomes. If the cognitive skills which are positively affected by track placement are more important, it might be worth it to accept the negatively affected non-cognitive skill development. If they are not however, placement in the high track could also have negative consequences for students. These could be interesting avenues for further research.

Our results could also help improve track allocation outcomes in another way. Teachers could still place the marginal students in the high track, but help these students prevent the in this paper observed negative non-cognitive skill development with targeted interventions. For instance, previous research has shown, although not specifically in a tracking context, that positive teacher relations could be a path to lower the Big-Fish-Little-Pond Effect and thus avoid the lower academic self-concept (Schwabe, Korthals, and Schils 2019).

In this paper we only focus on the marginal students around the placement threshold for whom the tradeoff might be most pronounced. However, it is not unconceivable that a similar tradeoff also exists for other students. For instance, students who already pose lower non-cognitive skills, or student who are particularly susceptible for the pressure of being in the high track. Furthermore, there could be costs at the school level (inefficient allocation of resources), and even at the societal level (over- or under-investments in human capital development). If this is the case, increased knowledge on also these consequences might improve the quality of track allocation and mitigate those negative effects.

Notes

1. For data availability reasons we exclude the low track in our analyses.
2. Selective schools can be considered as the high track, for instance when they prepare students for university entrance exams (i.e. the so-called preparatory schools). These so-called preparatory schools are quite common in France, but exist also in the United States, United Kingdom and Canada.
3. Some studies on tracking, like Hanushek and Woessmann (2006) and Ariga and Brunello (2007), focus on the overall effects of tracking which compare different tracking policies across countries. A number of other papers on tracking make use of different tracking policies within one country, often due to policy changes, to look at the effect of tracking. E.g. Pekkarinen (2008), Pekkarinen, Kerr, and Uusitalo (2013), and Hall (2012) who all find little effect. For ability grouping, Betts and Shkolnik (2000) find that only the grouped classes with average ability suffer from grouping, while there is no effect for the lower ability groups and a small positive effect for the high ability grouped classes. Figlio and Page (2002) find no negative effect of ability grouping for low-ability students and find some evidence they might even benefit from ability grouping.

4. See Hoekstra (2009) for similar analyses for entry into selective colleges.

5. See http://educatieveagendalimburg.nl/onderwijsmonitor-p/english for more information.

6. The main reason for schools not to participate was that they argued that it was too much workload for the teachers (participation takes two hours per class).

7. Not all children received the complete student questionnaire, resulting in a smaller sample for school well-being questions. Also, not all children took all tests or all test questions. Using IRT test scores are put on the same scale for all children who saw 13 or more test questions on each of the tests. We use the expected posterior estimates using a 2 parameter Bayesian Markov Chain Monte Carlo model.

8. Opt-out from the study is possible following a formal request from parents, yet only two percent of them do so. Further student non-response is mostly related to illness or other reasons for students not being in class.

9. Reasons for not having students at both survey moments are that the elementary school did not participate, the student was at an elementary school outside the region, or the student retained or skipped a grade in either elementary or secondary school. Since the survey in secondary school was held only every two years, for students who were retained or skipped a grade we do not have grade 6 data.

10. The data collection is approved by the Maastricht University inner city ethical committee (ERCIC_092_12_07_2018).

11. The three tracks are vmbo, havo, and vwo.Vmbo is preparatory middle-level vocational education which lasts 4 years, and consists of the sub-tracks vmbo-basic profession-oriented, vmbo-middle management-oriented, vmbo-mixed, and vmbo theoretical. Havo is higher general continued education and lasts 5 years. Vwo is preparatory scholarly education and lasts 6 years. Secondary schools with only students of a single track and schools with multiple tracks exist alongside each other, although the tracks could be separated across different school buildings. In the first year of secondary school, or sometimes in the first two years, so-called bridge classes exist in which students of multiple tracks are grouped together.

12. We restrict our data to those with an ES exit test score (on which our running variable is based).

13. Rejection of students is only allowed when the school faces capacity constraints, when students are registered for a different track than the recommendation or test suggests, or when parents do not accept the religious or educational philosophy of the school.

14. From the academic year 2014/2015 onwards, the government limited the use of the ESE test for secondary schools to deciding the track choice.

15. Given our fuzzy RD and the lack of a (predetermined) cutoff we do not to worry about students trying to manipulate their score to be above the cutoff. It is in all students’ best interest to have the highest possible exit test score.

16. The non-cognitive skills for which no effect with RDD is found are conscientiousness, openness, agreeableness, perseverance, and school satisfaction. Extraversion only shows significant effects with the ten score point bandwidth.

Acknowledgements

We would like to thank Bas ter Weel, participants of the Economics of Education group in Maastricht University, the International Workshop on Applied Economics of Education (IWAEE) 2014, the CEPA PhD workshop (Stanford University) and the AMCIS conference on Educational Systems in 2014 for useful comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Trudie Schils http://orcid.org/0000-0003-3533-9031
Lex Borghans http://orcid.org/0000-0003-3953-1370
References

Ander, R., S. Heller, H. Pollack, and J. Ludwigu. 2013. “Preventing Youth Violence and Dropout: A Randomized Field Experiment.” NBER Working Paper Series no. 19014, Cambridge: NBER.

Arens, A. K., H. W. Marsh, R. Pekrun, S. Lichtenfeld, K. Murayama, and R. vom Hofe. 2017. “Math Self-Concept, Grades, and Achievement Test Scores: Long-Term Reciprocal Effects Across Five Waves and Three Achievement Tracks.” Journal of Educational Psychology 109 (6): 621–634.

Ariga, K., and G. Brunello. 2007. “Does Secondary School Tracking Affect Performance? Evidence from IALS.” IZA Discussion Paper, 2643.

Betts, J., and J. Shkolnik. 2000. “The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools.” Economics of Education Review 19: 1–15.

Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2018. “Manipulation Testing Based on Density Discontinuity.” Stata Journal 18 (1): 234–261.

CITO. 2014. Cito Dossier Eindtoets 2014. Arnhem: CITO.

Diris, R. 2012. “The Economics of the School Curriculum.” PhD diss., University Maastricht.

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

Dumont, H., P. Protsch, M. Jansen, and M. Becker. 2017. “Fish Swimming Into the Ocean: How Tracking Relates to Students’ Self-Beliefs and School Disengagement at the end of Schooling.” Journal of Educational Psychology 109 (6): 855–870.

Dustmann, C., P. Puhani, and U. Schönberg. 2014. “The Long-Term Effects of Early Track Choice.” The Economic Journal 127 (603): 1348–1380.

Fietsersbond. 2016. “Resultaten fietsstelweek.” https://www.fietsersbond.nl/ons- werk/mobiliteit/fietsen-cijfers/. Last accessed: April 24 2017.

Figlio, D., and M. Page. 2002. “School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?” Journal of Urban Economics 51: 497–514.

Guyon, N., E. Maurin, and S. McNally. 2012. “The Effect of Tracking Students by Ability Into Different Schools.” Journal of Human Resources 47 (3): 684–721.

Hall, C. 2012. “The Effects of Reducing Tracking in Upper Secondary School.” Journal of Human Resources 47 (1): 238–269.

Hanushek, E. A., S. Link, and L. Woessmann. 2013. “Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA.” Journal of Development Economics 104: 212–232.

Hanushek, E., and L. Woessmann. 2006. “Does Educational Tracking Affect Performance and Inequality? Difference-in-Difference Evidence Across Countries.” Economic Journal 116: C63–C76.

Heckman, J. J., R. Pinto, and P. A. Savelyev. 2013. “Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes.” American Economic Review 103 (6): 2052–2086.

Heckman, J. J., and Y. Rubinstein. 2001. “The Importance of Noncognitive Skills: Lessons from the GED Testing Program.” American Economic Review 91 (2): 145–149.

Hoekstra, M. 2009. “The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach.” The Review of Economics and Statistics 91 (2): 717–724.

Imbens, G., and T. Lemieux. 2007. “Regression Discontinuity Designs: A Guide to Practice.” Journal of Econometrics 142 (2): 615–635.

Jackson, K. 2010. “Do Better Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago.” The Economic Journal 120 (December): 1399–1429.

Kautz, T., J. J. Heckman, R. Diris, B. Ter Weel, and L. Borghans. 2014. “Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success.” NBER Working paper no.20749 (revised in 2017). Cambridge: NBER.

Kingdom of the Netherlands. 1981. Wet op het Primair Onderwijs. Hoofdstuk I. Basisonderwijs. Artikel 42. Onderwijskundig rapport. The Netherlands: Law book.

Koning, P., and K. Van Der Wiel. 2013. “Ranking the Schools: How School-Quality Information Affects School Choice in the Netherlands.” Journal of the European Economic Association 11 (2): 466–493.

Lee, D., and T. Lemieux. 2010. “Regression Discontinuity Designs in Economics.” Journal of Economic Literature 48 (June): 281–355.

Marsh, H. W. 1987. “The Big Fish Little Pond Effect on Academic Self-Concept.” Journal of Educational Psychology 79: 280–295.

Pekkarinen, T. 2008. “Gender Differences in Educational Attainment: Evidence on the Role of Tracking from a Finnish Quasi-Experiment.” Scandinavian Journal of Economics 110 (4): 807–825.

Pekkarinen, T., S. Kerr, and R. Uusitalo. 2013. “School Tracking and Development of Cognitive Skills.” Journal of Labor Economics 31 (3): 577–602.

Pop-Eleches, C., and M. Urquiola. 2013. “Going to a Better School: Effects and Behavioural Responses.” American Economic Review 103 (4): 1289–1324.
Porter, J., and P. Yu. 2015. “Regression Discontinuity Designs with Unknown Discontinuity Points: Testing and Estimation.” *Journal of Econometrics* 189: 132–147.

Roberts, B., W. Viechtbauer, and K. Walton. 2006. “Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies.” *Psychological Bulletin* 132 (1): 1–25.

Schwabe, F., R. Korthals, and T. Schils. 2019. “Positive Social Relationships with Peers and Teachers as Moderators of the Big-Fish-Little-Pond Effect.” *Learning and Individual Differences* 70: 21–29.

Seaton, M., H. W. Marsh, and R. G. Craven. 2009. “Earning its Place as a Pan-Human Theory: Universality of the Big-Fish-Little-Pond Effect Across 41 Culturally and Economically Diverse Countries.” *Journal of Educational Psychology* 101: 403–419.

Staiger, D., and J. Stock. 1997. “Instrumental Variables Regression with Weak Instruments.” *Econometrica* 65 (3): 557–586.

Statistics Netherlands. 2012. *Jaarboek Onderwijs in Cijfers 2012*. Den Haag/Heerlen: Centraal Bureau voor de Statistiek.

Stock, J., and M. Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by D. Andrews, J. Stock, and T. Rothenberg, 80–108. New York: Cambridge University Press.

Van Elk, R., M. Van Der Steeg, and D. Webbink. 2011. “Does the Timing of Tracking Affect Higher Education Completion?” *Economics of Education Review* 30 (5): 1009–1021.

Vardardottir, A. 2013. “Peer Effects and Academic Achievement: A Regression Discontinuity Approach.” *Economics of Education Review* 36: 108–121.