Maastricht University

Is there an opportunity-performance trade-off in secondary education?

Citation for published version (APA):

Bles, P., van der Velden, R., & Ariës, R. J. (2020). *Is there an opportunity-performance trade-off in secondary education?* ROA. ROA Research Memoranda No. 009
https://doi.org/10.26481/umaror.2020009

Document status and date:
Published: 27/07/2020

DOI:
10.26481/umaror.2020009

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 12 Sep. 2026

Maastricht University

ROA

Is there an opportunity-performance trade-off in secondary education?

Per Bles
Rolf van der Velden
Roel J. Ariës

ROA Research Memorandum

ROA-RM-2020/9

Researchcentrum voor Onderwijs en Arbeidsmarkt | ROA
Research Centre for Education and the Labour Market | ROA

Is there an opportunity-performance trade-off in secondary education?

Per Bles
Rolf van der Velden
Roel J. Ariës

ROA-RM-2020/9
July 2020

This work was supported by the NWO-NRO under Grant number 405-17-305 / 2930.

Research Centre for Education and the Labour Market
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
T +31 43 3883647 F +31 43 3884914

secretary-roa-sbe@maastrichtuniversity.nl
www.roa.nl

ISSN: 2666-8823

Abstract

Is there an opportunity-performance trade-off in secondary education?*

Schools in secondary education face a dilemma. On the one hand, they would like to offer all students opportunities to develop their talent, and on the other hand they want to safeguard a minimum performance level. In tracked systems, this dilemma becomes more consequential as misallocation of students could lead to either denying access to a more optimal track or to lower performance of students that are placed too high. Based on data from the Netherlands Cohort Study on Education (NCO) from 2010 to 2017, we find that only for 55% of schools there is a trade-off between opportunity and performance. These schools show a relative preference for either opportunity or performance. However, in the other schools, opportunity and performance are optimised at the same time; this dimension is related to the quality of the school. While controlling for the school's potential student population, we show which school characteristics are associated with the relative preference and quality dimensions.

JEL classification: I21, I22
Keywords: trade-off, opportunity, performance, school quality, school relative preferences, school characteristics, potential student population

Per Bles
Maastricht University
ROA
P.O. Box 616
NL-6200 MD Maastricht
The Netherlands
p.bles@maastrichtuniversity.nl

Rolf van der Velden
Maastricht University
ROA
P.O. Box 616
NL-6200 MD Maastricht
The Netherlands
r.vandervelden@maastrichtuniversity.nl

Roel J. Ariës
Dutch Inspectorate of Education
P.O. Box 2730
NL-3500 GS Utrecht
The Netherlands
r.aries@maastrichtuniversity.nl
and Academic Workplace Education Quality

* We would like to thank the participants of the AWP conference in Utrecht (December 13, 2018), AERA conference in Toronto (April 6, 2019), ORD conference in Heerlen (June 27, 2019), and Corinna Kleinert, Eric Hanushek, Ilja Cornelisz, Chris van Klaveren, Davey Poulissen, Carla Haelermans, Katarina Weßling and Melline Somers for their valuable suggestions. All errors are ours.

Introduction

In all education systems, students get sorted, either in groups, tracks or in schools (LeTendre, Hofer, & Shimizu, 2003), but in early stratifying education systems, this sorting is more consequential. Early tracking is known to increase social inequality (e.g. Chmielewski, 2019; Van de Werfhorst, 2018) and might lead to an underdevelopment of students' talent (Borghans, Diris, Smits, & de Vries, 2020). The problem at hand is that choices on placement at the time of selection are made on the basis of imperfect information, i.e. a noisy signal (Brunello, Giannini, & Ariga, 2007). Hence, the decision might turn out wrong: in hindsight, it was wrongly assumed that either the student was able to perform in that group (false positives) or that this was not the case (false negatives). Both types of errors negatively affect a student's career: either students are denied access to a higher track (i.e. denying opportunities to develop their talent) or students underperform because they are allocated to a track that is too high for their ability level.

In the Netherlands, both types of indicators are monitored in the accountability framework of the Inspectorate of Education (De Wolf and Janssens, 2007; Ehren, Altrichter, McNamara, & O'Hara, 2013; Inspectorate of Education, 2017). In this framework, schools are monitored on their *performance* with indicators such as exam scores or retention rates. These might suffer from too lenient selection. Similarly, schools are also held accountable for getting the best out of the students, i.e. granting them *opportunity* at a track that is optimal for their ability, which could suffer from a myopic focus on performance indicators.

To cope with the possible consequences of the above-mentioned noisy selection instrument, schools can act differently. Expressed as a continuum, the tail-ends consist of lenient and stringent selection. At one end, schools could prefer to minimize their false negatives by granting students access to a higher track, even though they score just below the threshold. At the other end, schools could minimize their false positives by selecting more

stringently, and only select those eligible for the track. To gain more insight into this trade-off, our first aim is to describe to what extent schools balance between giving opportunity or focusing on performance:

To what extent do secondary schools face a trade-off between providing opportunities and the performance of students? (RQ 1).

The dilemma of secondary schools as formulated in the above-mentioned trade-off is implicitly placed in the context of the value-added model, used to assess school quality (Everson, 2017). In the value-added model, school outcomes such as exam grades or retention are corrected for input differences. As input differences are highly correlated with giving opportunities and output differences with performance, the value-added model does not adequately grasp this trade-off dilemma. Two schools could end up having similar value-added scores, but still be very different in how they reach this score. The one could give many opportunities at the expense of performance, while the other could aim for high performance at the cost of opportunity. Therefore, we think that we can advance the school effectiveness literature by explicitly looking at both dimensions: quality as well as relative preference. Both dimensions may be related to different school characteristics and both could subsequently have distinct implications for students' careers. So, our second research question is:

Which school characteristics are related to quality and which are related to relative preferences for opportunity or performance? (RQ 2).

Furthermore, school effectiveness research is hampered by interpreting the effects of the school's student composition. This composition can be the result of deliberate choice of the school itself (selection), the choices made by students or their parents (self-selection) or simply a reflection of the potential student population in the vicinity of the school (Cabus and Cornelisz, 2017; Leist and Perry, 2019; Weßling and Bechler, 2019). It is important to

distinguish the former two from the latter, in order to make a distinction between the selectivity of a school versus the selectivity of the potential student population. Our third research question therefore reads:

What is the effect of the social selectivity of the school on quality and relative preference for opportunity or performance? (RQ 3).

The paper proceeds as follows. The next section provides a brief overview of the Dutch education system. Section 3 reviews the theory and introduces the hypotheses. Section 4 discusses the data and methodology. Section 5 presents the results and Section 6 concludes.

1. The Dutch education system

Secondary education in the Netherlands is divided into five tracks reflecting various ability levels. Three 4-year vocational education tracks (VMBO-b, VMBO-k and VMBO-gt) qualify for various upper secondary vocational education tracks (MBO). The 5-year preparatory academic track (HAVO) gives access to the universities of applied sciences (HBO), and the 6-year preparatory academic track (VWO) prepares for a research university (WO) (see Figure 1).

At the end of primary education, a report about which track suits the student's abilities best determines the eligibility for secondary education tracks. For the cohort under consideration, this report is based on the results of a standardized national cognitive test at the end of primary education, and the recommendation of the primary school teacher, based on the student's overall performance as well as behavioural characteristics and motivation (Oomens, Scholten, & Luyten, 2017; Timmermans, de Boer, Amsing, & van der Werf, 2018).

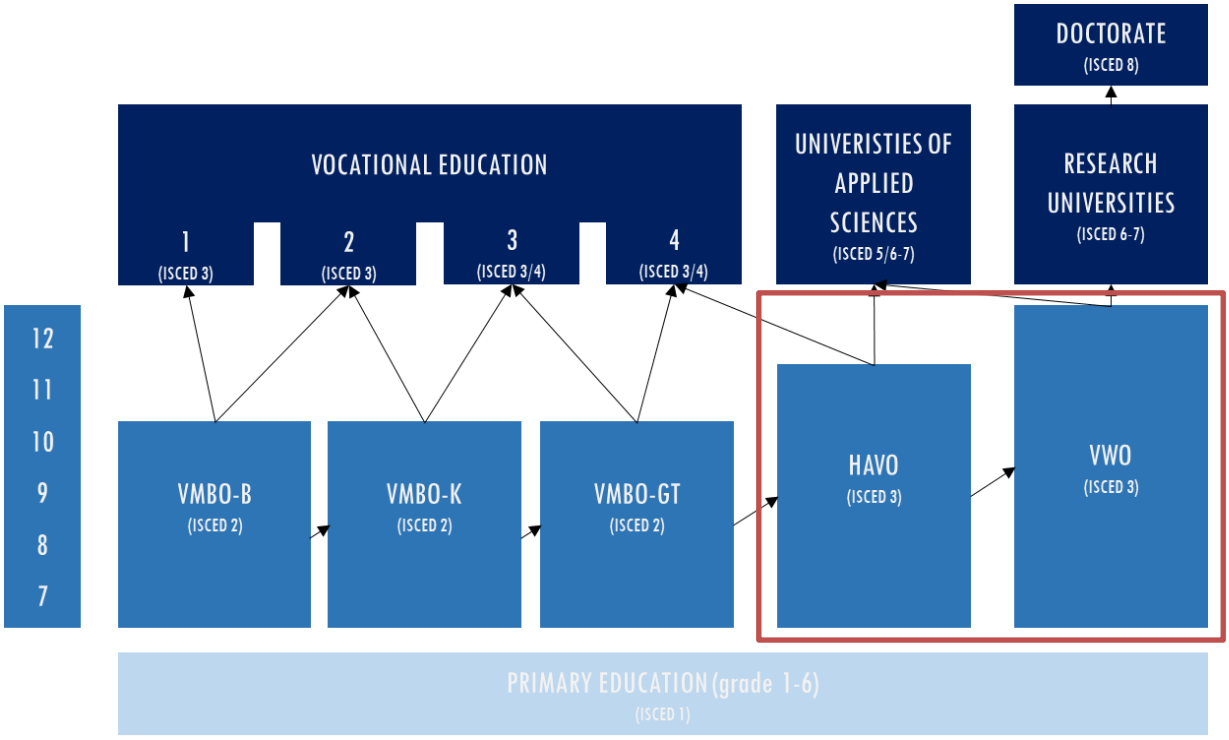

Figure 1. Education tracks in secondary education.

Note: The arrows depict pathways through the system. The red square indicates the tracks analysed in this study. Secondary education (light blue) is depicted relative to its duration (corresponding international grades are indicated at the left); upper secondary and tertiary education (dark blue) is not.

Although the Dutch system is known for its early tracking in homogeneous classes at age 12, secondary schools have the autonomy to install comprehensive ability classes in the first year (grade 7; age 12) and as late as in the third year (grade 9; age 14) (see Korpershoek, Naayer, & Bosker, 2017). In school year 2011/2012 about 63% of the students were placed in a comprehensive ability class (own calculations). A school can have just a single track (‘categorale scholen’), or multiple tracks. Multi-track schools that consists of three or more tracks are usually called school communities (‘scholengemeenschap’). Depending on the supply of schools in their vicinity, students can thus opt for placement in a single-track school or a school community.

Despite a substantial amount of autonomy, schools are also bounded by oversight (OECD, 2016). More specifically, they have to consider the criteria put forward by the

Inspectorate of Education (Ehren, et al., 2013; Inspectorate of Education, 2017). The Inspectorate formulates standards and expectations in the domain of legal requirements, the domain of the education process (e.g. the number of instruction hours), and the domain of performance and results. The performance and results indicators are fourfold: (1) the share of students placed in a higher track in the third year of secondary school compared to the primary school teacher's recommendation, (2) the share of students that have followed lower (2b) and upper (2b) secondary education without retention or dropout, and (3) the average exam results (Inspectorate of Education, 2017).

2. Theory

Early allocation into different tracks is based on the assumption that selection in homogeneous classrooms "permit[s] a focused curriculum and appropriately paced instruction" (Hanushek and Wößmann, 2006, p. c63) in order to improve every student's performance. Therefore, a valid and accurate selection is key for a well-functioning tracked system (Korthals and Dronkers, 2016; Smolkowski and Cummings, 2015), as it determines the degree of focus and pace allowed in class.

However, the selection instrument on the placement contains noise. These relate to imperfect information on the ability of the student at the time of decision-making (Brunello, et al., 2007): what seemed a correct decision at the time of selection, might in the end not be the best fit for the student's ability. One reason for the noise is the differential dynamic pace of the development of children's brain (Knudsen, Heckman, Cameron, & Shonkoff, 2006). Also, during their secondary education career students are exposed to other teachers, curricula and classroom environments than in primary education (Muijs et al., 2014), which might affect their cognitive development. Furthermore, changes in motivation, preferences, expectations and access to information on the school system over and above ability measures

(Büchner and Van der Velden, 2013; Forster and van de Werfhorst, 2019; Timmermans, et al., 2018), might affect the predictive power of the selection instrument.

On average, suboptimal track placements can be expressed in language that mimics hypothesis testing in statistics: first, those students who were wrongly assumed to be able to perform well in a certain track (false positives or Type I error) or, second, those who were wrongly assumed not to be able to perform in a certain track (false negatives or Type II error) (Smolkowski and Cummings, 2015). Thus, although one cannot speak of an erroneous placement decision at the time of decision per se (after all, there was no other information available yet), we use the statistical terms of Type I and Type II errors to express that the hypothesised ability tracks turned out to be a misfit.

The Trade-Off between Opportunity and Performance

In coping with the consequences of imperfect information when students enter secondary education, schools can put forward different alternatives, latently placed on a continuum ranging for selecting very strict to very lenient. This selection applies to both the selection when students enter the school and the school's policy with regard to upward and downward mobility between tracks in later years. Schools might be very strict, thus lowering the fraction of Type I errors, but simultaneously increasing the fraction of Type II errors. The assumption of those schools is that this will result in accomplishing the highest possible performance rates, such as low retention rates and high exam grades. They will thus score higher on the second and third criterion of the Inspectorate, but lower on the first criterion as the actual track placement or upward mobility might not result in students who are tracked higher than the recommendation of the primary school teacher indicated. On the other hand, schools might focus on giving opportunities to students at the start of secondary education or by being upward mobile and thus increase the number of students placed higher than the primary school teacher recommended. This will decrease the Type II errors but might increase the

Type I errors. The school thus risks decreased performance rates. Hence, a trade-off mechanism comes to the fore, which is also underpinned by the statistical insight that Type I and Type II errors are by definition negatively correlated (Sheskin, 2004). Thus, schools face a trade-off between giving opportunities to students and their average performance indicators *(Trade-off hypothesis).*

Quality and Relative Preferences

School accountability research shows that quality indicators are an integral part of the Inspectorate's school oversight and thus in the choices schools make in their policy (De Wolf and Janssens, 2007; Ehren, et al., 2013). Plotting performance and opportunity indicators of schools in a two-dimensional graph sheds light on the matter in a visual way (see Figure 2). The quality of schools is reflected by a combination of "the degree to which [it] scores better than other schools, corrected for student intake characteristics" and the extent to which it is good at "compensating for input characteristics" (Reynolds et al., 2014, p. 205).

In our framework, we follow the Inspectorate and mimic their quality indicators. Both the placement of the student in the third year compared to the primary school teacher's recommendation (mimicking the input compensation mechanism), and the pace at which students move through secondary school (mirroring the quality outcome score), are part of the quality indicators. Therefore, we regard high scores on both performance and opportunity as qualitatively good schools. These are the schools in Quadrant 1. Reversely, if schools score low on both dimensions, they can be labelled as low-quality schools (Quadrant 2). In Figure 2, diagonal line A depicts this dimension of school quality, which is simply defined as the sum of the standardised score of the two underlying variables (opportunity and performance).

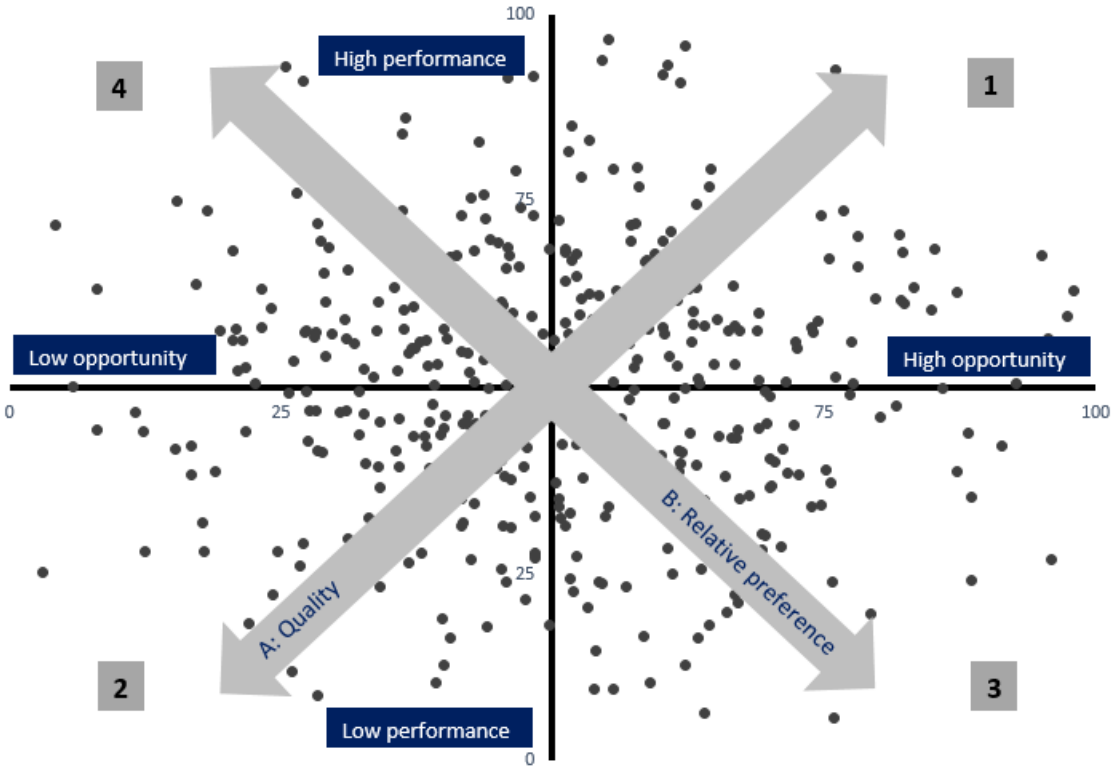

Figure 2: Quality and Relative Preference Outcomes in Opportunity and Performance.

Note: the dots represent hypothetical schools. Line A depicts the quality dimension; Line B depicts the relative preference dimension. 1 = high-quality schools; 2 = low-quality schools; 3 = relative preference for opportunity; 4 = relative preference for performance.

However, two schools with the same score on the quality indicator might still differ in how they reach this quality score. Certain schools might combine a high opportunity score with a relatively lower performance score (schools in Quadrant 3), while other schools might combine a relatively lower opportunity score with a higher performance score (Quadrant 4). Both will end up with an 'average' quality score, but they are nevertheless different in their approach. This is depicted by the relative preference diagonal B. This dimension of relative preference reflects the consequences of choices that schools might make in their policy. This does not mean that schools always make these policy choices deliberately. The relative preference is defined as the subtraction of the standardised score of the two underlying variables (opportunity and performance). Given this definition, quality and relative preference are, by definition, orthogonal (see also the operationalisation section).

Type of Schools and the Quality and Relative Preference Dimensions

As schools have considerable autonomy (OECD, 2016), we will discuss and hypothesize different characteristics of schools and their association with quality and relative preferences. First, there are characteristics that schools have no influence on: urbanisation level and the social composition of the potential student population in the school's so-called student supply radius. We will control for those factors. Second, there are school characteristics that schools can influence in the medium to long-term: their track size, denomination and instalment of comprehensive classes. We will especially focus on and derive hypotheses about the structure of the tracks within the school (Korpershoek, et al., 2017) that enables schools to modify track placement decisions (Van der Velden and Weßling, forthcoming). Finally, the third type of factor that determines a school's relative preferences and quality relates to the characteristics of the student population, net of the composition of the potential student population in the neighbouring area of the school. This is related to either the specific selection of schools or the self-selection of students (or their parents) in choosing a specific school (e.g. Leist and Perry, 2019).

Flexibility Mechanisms within the Institutional Structure

The selection regimes in education can broadly be described according to three models: a sponsored mobility, an open contest mobility, and a tournament model (Rosenbaum, 1979; Turner, 1960). If there is no permeability between different tracks after the entry into secondary education, it aligns with the 'sponsored mobility regime' of Turner (1960). In this model schools have strict entry selection requirements, but at the same time they make sure that those who are selected will get a diploma and will thus perform well (see also Inspectorate of Education, 2016, pp. 22-23). In this regime Type I errors are reduced at the expense of Type II errors.

However, when there are multiple moments to decide upon the placement of the student after the start of secondary education, the selection regime rather resembles the 'contest mobility regime' (Turner, 1960) or the 'tournament mobility regime' Rosenbaum (1979). The distinction between the two models is that in Turner's model those who do not make it to a track still have an opportunity to reach for it another time ('new round, new chances'), while in the tournament model students have to perform continuously in order to get a diploma (Jennings, 2010, p. 228; Rosenbaum, 1979, p. 223; Turner, 1960, p. 861). In both models, the amount of Type II errors is decreased. However, the 'contest mobility regime' also reduces the number of Type I errors. Not only are able students who in first instance were not selected given opportunities (Type II), but also unable students who were selected are now on the right (lower) track (Type I) (Van der Velden, 2011).

The different selection regimes are mirrored in how Dutch schools use the flexibility possibilities in track placements. Dutch schools have these flexibility options because of three structural characteristics: single-track schools, school communities and comprehensive tracks in the first or second year (Korpershoek, et al., 2017).

First, in a single-track school there are no options to move to a different track within the same school. Although students always can transfer to another school, transferring within the same school lowers the bar to do so. In single-track schools, the permeability between tracks and thus the option to reduce the total amount of misallocated students is lower. Following our theoretical model, we assume that both the amount of opportunities given to students is low in a single-track school and the performance of students who are wrongly selected in the track is low, therefore the single-track school is likely to score low on our quality indicator. The single-track schools most closely resemble the sponsored mobility regime, with a relative preference for performance over opportunity (Inspectorate of

Education, 2016, p. 22). We therefore hypothesise that single-track schools score low on the quality dimension and have a relative preference for performance (*Single-track hypothesis*).

Second, unlike single-track schools, school communities allow students more mobility within the same school and, consequently, to be better allocated (Inspectorate of Education, 2019, p. 90). The school environment within a school community is more similar across tracks, transferring information about students is easier, and teachers overlap across tracks, all making mobility between tracks easier (Bronneman-Helmers, Herweijer, & Vogels, 2002, pp. 102-115). Within a school community, there are more possibilities to select more lenient, thus lowering the Type II error, because the mobility between tracks is easier and the placement is thus less final, which in turn also reduces the Type I error. This type of school could enable a contest mobility regime. We therefore hypothesise that schools that are part of a school community score high on the quality dimension and show a relative preference for opportunity (*School community hypothesis*).

Third, literature on the placement in comprehensive ability classes shows that they provide an opportunity for students at the margin to catch up and reach their full potential (Borghans, et al., 2020; Van Elk, Van der Steeg, & Webbink, 2011). Moreover, it is an opportunity for the school to gather more information about the student to base their later tracking decision on (Mühlenweg, 2007), since students' true ability is more uncertain at an earlier age (Knudsen, et al., 2006). It is a feature that is helpful for the late bloomers and thus compensates for (minimal) differential developmental trajectories at the time of selection (see also Dustmann, Puhani, & Schönberg, 2017; Inspectorate of Education, 2016, p. 23). As such, this feature enables a contest mobility regime in the first one or two years. We therefore expect that schools with one- or two-year comprehensive classes score high on the quality dimension and have a relative preference for opportunity (*Comprehensive ability classes hypothesis*).

The selection of students into schools

The value-added models that try to describe school quality do control for the student intake and make a comparison between schools fairer, as research shows that certain student compositions are associated with higher performance scores (Reynolds, et al., 2014; Wenger, Gärtner, & Brunner, 2020). Girls and students from high SES families are performing better on average and non-western students and students from divorced parents perform worse (Armor, Marks, & Malatinszky, 2018; Havermans, Vanassche, & Matthijs, 2017; Luyten, Schildkamp, & Folmer, 2009; Van Hek, Kraaykamp, & Pelzer, 2018).

However, the composition of the school population is the outcome of both the potential group of students in the vicinity of the school, their preferences and the selection policy of the school itself. Having two schools with the same composition could be either ascribed to the same composition of students in the vicinity of the schools or due to similar selection processes (deliberate either by the school or by the students and their parents), i.e. we also have to control for the *potential student population* in the surrounding neighbourhoods. Some schools are located in areas that are more densely populated with high or low performing students (e.g. Nieuwenhuis and Hooimeijer, 2016).

We therefore disentangle between the potential student composition and actual student composition, by looking at characteristics of the eligible student population in a 5 km radius of the school. Apart from performance constraints, students and their parents are free to choose which school they attend. There is no strong evidence that parents move to a different neighbourhood to be able to attend a higher quality school (Borghans, et al., 2020). Apart from the official performance related entry rules, students and parents can self-select or schools can select on other characteristics. Targeting policies (e.g. advertisements; open days; philosophical denomination; prior performance scores) can attract a social selective group of students (Allen, Bijlsma, Borghans, & Poulissen, 2016; Cabus and Cornelisz, 2017; Jennings,

2010; Leist and Perry, 2019). As schools have no direct effect on the potential student composition, but do have influence over their (self-)selection, we assume that the quality score will be higher for schools that attract relatively more girls and high SES students and less one-parent household students compared to the potential student population in the school radius (*Quality selection hypothesis*).

The preferences of schools for either opportunity or performance stem from incentives to differentiate in policies to compete with other schools. On the one hand, the preferences of schools are demand-driven and depend on which school features students, parents and primary school class mates consider important (Cabus and Cornelisz, 2017; Glazerman and Dotter, 2017; Jennings, 2010; Ruijs and Oosterbeek, 2019). We can assume that certain types of students seek a certain type of school and that parents want to prevent status decline for their children (Breen and Goldthorpe, 1997), which makes high SES parents more inclined to seek a school with a preference for performance. Moreover, research shows that high SES parents are better informed about school characteristics than low SES parents (e.g. Forster and van de Werfhorst, 2019). Hence, schools that select more high SES students than expected on the basis of the potential student composition in the school's supply radius are more likely to have a relative preference for performance (*High SES preference selection hypothesis).*

The preferences may, on the other hand, also relate to a certain philosophy that is propagated a certain school and thus cause a differential school supply. Schools with a philosophical or religious orientation are also concerned with teaching the students certain norms and values (Bronneman-Helmers, et al., 2002). They might find it important to provide more opportunities to students that adhere to such principles, even at the cost of performance, and thus opt for opportunity over performance (*Denominational preference hypothesis*).

3. Data, methodology and operationalisation

Data

We make use of the data from the Netherlands Cohort Study on Education (NCO) (Haelermans et al., 2020) and observe schools and their students within the publicly funded secondary school system from school year 2011/2012 to 2017/2018. Next to background information of the student and the schools, the records contain information at the individual level about the students' school careers. An important advantage of using register data is that it covers almost all students in secondary education; less than 2% of the Dutch schools are privately funded schools (Haelermans, et al., 2020). However, we have to keep in mind that the register data only contain 'objective' variables and not the kind of subjective variables that may also play a role in the track placement of students.

Operationalisation of key variables

First, the outcome variables of interest in this paper are discussed: (1) opportunity; (2) performance; (3) relative preference and (4) quality. Moreover, in Appendix 1, we will describe the independent (control) variables that are used in our analyses. Second, the selection into schools will be discussed.

Opportunity. Opportunity is defined as students whose actual track placement in the third year of secondary school is higher than the primary school teacher's recommendation. For instance, if the student resides in the VWO track and has a HAVO/VWO recommendation or lower, we define this as giving this student an opportunity. Similarly, if the student is allocated to a HAVO class on a VMBO-gt/HAVO recommendation or lower, this is defined as an opportunity. The resulting scale has a potential range from 0% to 100%.

Performance. Performance is defined as the percentage of students in the track who graduated without delay in the secondary education track for which they enrolled in the third year of secondary education. All years spent in secondary education are considered, so also those before the third year, when calculating the percentage of students who receive a diploma without delay. The resulting scale has a potential range from 0% to 100%.

Quality and relative preferences. As described in the theory section, we define quality (Q) as the sum of the scores on performance (P) and opportunity (O) of a school (j). To ensure an equal weight, both variables are standardised (z).

$$zQ_j = zP_j + zO_j$$

Next, we define the relative preference (RP) as the difference between these two variables, by subtracting the standardised opportunity measure from the standardised performance measure.

$$zRP_j = zP_j - zO_j$$

Note that, by definition, this means that opportunity and performance are orthogonal and thus have a correlation of zero.

Selection into schools. We define the selectivity of the school as the deviation of the actual school composition from the potential composition based on the eligible student composition in a 5 km radius around the school. A radius of 5 km is considered appropriate in the Netherlands, as most students live within this range and because in densely populated areas bigger radii would yield meaningless composition effects (Allen, et al., 2016). The eligibility

is determined on the basis of the primary school teacher's recommendation.[1] We calculate composition difference scores on the percentage of girls, non-western migration status, high paternal income and one-parent families. The selectivity of a school is then expressed in the form of a segregation index:

$$I = \sum_{n=1}^{N} \frac{n_{group}}{n_{school}} - \frac{n_{group}}{n_{radius}} \times 100$$

where the n_{group}, n_{school} and n_{radius} represent the number of students with a certain characteristic, the number of students in the school and in the radius respectively. The differential score has a potential range of -100 to +100. The interpretation of the – or + is the deviation in percentage points from a situation where the actual composition of the school equals the potential composition of students in the school's radius. A negative number indicates that there are more students with that characteristic in the radius than at the school and vice versa for a positive number.

Methodology and analysing strategy

We estimate opportunity, performance, quality and relative preferences for different characteristics of schools using ordinary least squares regression (OLS). We run separate analyses for HAVO and VWO, since the tracks differ in time, content and the occurrence of ceiling-effects. We exclude VMBO tracks due to floor-effects and limited variation in our operationalised independent and dependent variables.

[1] For eligibility for HAVO, all students with a straight HAVO recommendation in that radius are considered, as well as the proportion of a recommendation consisting of more than one track. For HAVO/VWO and VMBO-gt/HAVO, half of the students are eligible and for recommendations consisting of HAVO and two more tracks, a third of the students are eligible. In this case half or a third of the students are randomly assigned to the eligible HAVO pool.

The cohort we analyse enters secondary education in the school year 2011/2012. We use the third-year (grade 9; age 15) secondary education cohort (2013/2014) as a criterion of *selection.* In this grade 9, there are hardly any comprehensive classes, so each individual can be linked to a certain track in the school. Of those students, we observe the primary school teacher's recommendation and test score of the school year 2010/2011, retention and acceleration in secondary education of those students entering secondary education in school year 2011/2012, switching of track and the final exam results. Most of the exam results are nominally obtained in school year 2015/2016 for HAVO and in school year 2016/2017 for VWO (see also the operationalisation section). Table 1 gives an overview of the cohort structure of the data.

Table 1. Cohort design of the data.

2010/2011	2011/2012	2012/2013	**2013/2014**	2014/2015	2015/2016	2016/2017	2017/2018
Grade 6	Grade 7	Grade 8	**Grade 9**	Grade 10	Grade 11	Grade 12	
Primary schoolteacher recommendation and test score	First year	Second year	**Third year entry cohort**	Exam year VMBO	Exam year HAVO	Exam year VWO	

Starting from our individual level dataset, we made the following selections. First, we only included students that were either in a VWO or HAVO track in grade 9. Second, we excluded students that had a primary school teacher's recommendation that consisted of three or more adjacent tracks, because for these cases it was not possible to calculate the opportunity score (for VWO this amounts to 0.8% of all recommendations in 2010; for HAVO to 1.1%). This results in 43,278 students in the third year of the VWO-track. For HAVO this number is 42,020. Third, after aggregating the data to the school level, small schools with tracks that have less than 20 students in each observed cohort year are excluded as well. Lastly, we excluded schools that did not have any information on any of the variables used in the

multivariate analysis. This results in 413 schools that offer a HAVO-track and 429 schools that offer a VWO-track.

In Table 2 below, we show descriptive statistics for the analyses for both HAVO and VWO. The dependent variables are standardised, hence the mean of 0. The potential student composition in the school's supply radius varies strongly, if we consider the relative high standard deviation. Furthermore, the percentage of students that received a recommendation that is higher than the test score implies also varies substantially between schools. The track size, school denomination, urbanisation, school community, single-track school and comprehensive classes are dummy variables, therefore the mean indicates the percentage of schools in the sample. Finally, the difference scores on selection variables have a theoretical range from -100 to +100 and show quite some variation given the size of the standard deviation.

The regression models are built-up as follows: first, we enter factors that describe the context that schools have to deal with when making policy decisions. Schools can hardly influence those contextual factors and therefore these factors should be controlled for: 1) potential student composition within the supply radius of a school; 2) the urbanisation level and 3) the extent to which the supplying primary schools are systematically giving a recommendation that is higher than their national test scores suggest. Second, we enter characteristics of schools that can be changed in the medium to long-term: track size, denomination and the structure of tracks within the school. Third, we enter specific selection of schools or the self-selection of students (or their parents) in choosing a specific school into the model.

Table 2. Descriptive Statistics for both HAVO and VWO

	HAVO			VWO		
	N	Mean/ Proportion	Std. Dev.	N	Mean/ Proportion	Std. Dev
Dependent variables						
Opportunity	413	24,659	14,454	429	39,632	17,889
Performance	413	65,329	12,379	429	67,945	11,191
Quality: performance + opportunity	413	0,000	1,319	429	0,000	1,159
Relative preference: performance - opportunity	413	0,000	1,503	429	0,000	1,630
Composition of 5km supply radius						
% girls	413	0,503	0,041	429	0,475	0,050
% students with non-western migration background	413	0,144	0,118	429	0,103	0,076
% one-parent households	413	0,162	0,059	429	0,131	0,052
% high SES	413	0,459	0,086	429	0,568	0,084
Track size						
0-60 students	413	0,295		429	0,336	
60-120 students	413	0,598		429	0,529	
More than 120 students	413	0,107		429	0,135	
Urbanisation						
Not urban	413	0,041		429	0,030	
Somewhat urban	413	0,169		429	0,168	
Moderately urban	413	0,245		429	0,231	
Strongly urban	413	0,344		429	0,354	
Very strongly urban	413	0,201		429	0,217	
Denomination						
Public schools	413	0,305		429	0,310	
Non-religious	413	0,097		429	0,093	
Religious	413	0,598		429	0,597	
Single-track school				429	0,093	
School community	413	0,874				
One year comprehensive classes	413	0,431		429	0,401	
Two years comprehensive classes	413	0,487		429	0,422	
Social selectivity						
Difference score girls	413	0,841	8,649	429	5,975	8,948
Difference score non-western	413	-1,236	11,033	429	-0,897	8,106
Difference score high SES	413	2,499	10,089	429	3,562	9,929
Difference score one-parent households	413	1,763	6,246	429	2,328	5,315

4. Results

We first discuss the results regarding the trade-off between opportunity and performance. Thereafter, we discuss the results of school characteristics on quality and relative preferences. We discuss the results of school characteristics on opportunity and performance separately in Appendix 2.

The trade-off

The correlation between opportunity and performance is -0.328 in the case of VWO and -0.129 in the case of HAVO, as can be seen in Table 3. This indicates a low to modest trade-off, but by no means a strong one. A visual representation of the trade-off is given in Figure 3, where clusters of schools are depicted in the two-dimensional graph. The clusters of schools are scattered over the graph and do not indicate a strong negative relationship.

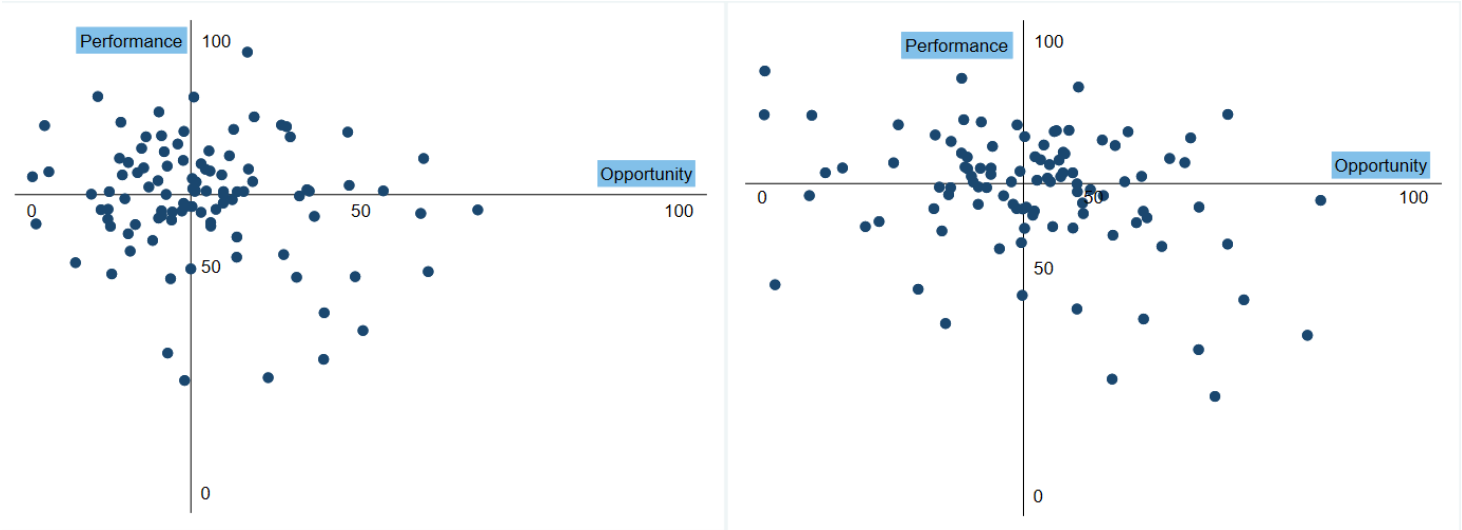

Figure 3: Opportunity and Performance graph for HAVO (left) and VWO (right).
Note: the dots are midpoints of clusters of schools

Another clue towards having leeway in avoiding trade-offs lies in the distribution of the schools over the quadrants when opportunity is set against performance in a graph. In the case of the VWO-track, the numbers in the quadrants that trade performance for opportunity (27%) and opportunity for performance (29%) together consist of 56% of all schools; 24%

are high quality schools and 20% low quality schools. For HAVO, the quadrants that trades off performance for opportunity (33%) or opportunity for performance (22%) together consist of 55% of all schools as well, whereas 24% of the schools could be labelled as schools with high quality and 21% as schools with low quality.

Table 3. Correlations between the dependent variables

HAVO	Quality	Relative preferences	Opportunity	Performance
Relative preferences	0,000			
Opportunity	**0,660**	**-0,751**		
Performance	**0,660**	**0,751**	**-0,129**	

VWO	Quality	Relative preferences	Opportunity	Performance
Relative preferences	0,000			
Opportunity	**0,580**	**-0,815**		
Performance	**0,580**	**0,815**	**-0,328**	

Relative preferences and quality

The results for the highest academic track VWO suggest the following. Schools with more high SES students in their supply radius have a higher score on the quality indicator. Schools that have more students with a recommendation that is higher than the test score are negatively correlated with the quality score. Schools that have a comprehensive class in the first or first two years score higher on the quality dimension, as well as schools that have selected more girls and high SES students than their respective supply radius averages would suggest. The effect sizes of especially one- and two-year comprehensive classes is large: 0.469 and 0.448, respectively.

With regard to the relative preferences of the VWO-schools, schools in high SES radii have a preference for performance over opportunity, and the same holds for single-track schools. Schools that have a religious denomination have a preference for opportunity over performance, as well as schools with one- or two-year year comprehensive classes. Moreover, if schools select more girls than would have been expected given the supply radius

average, they prefer opportunity over performance. The opposite holds for schools that select more high SES students than the supply radius average: they are preferring performance over opportunity. Here, single-track schools, comprehensive classes and selection on high SES have the largest effect sizes, ranging in absolute value between 0.191 and 0.297.

These results suggest that the comprehensive ability classes hypothesis is confirmed; the single-track hypothesis is only confirmed when it comes to preferences. Moreover, it seems that the denominational preference selection hypothesis is confirmed as well as the quality selection hypothesis, at least when it comes to selecting girls and high SES students.

For HAVO, the results show that the composition of the supply radius correlates positively with quality when it comes to girls and high SES students and negatively when it comes to one-parent households. Also, if a school has more students whose recommendation is higher than the national test score, the quality indicator is lower. When a track is part of a school community it is positively related to quality; the effect being quite substantial 0.179. Lastly, all selection indicators are significantly different from zero, meaning that if a school selects more girls, and high SES students than the average of the supply radius indicates, this is associated with a higher score on the quality dimension. Their effect sizes are also substantial, in absolute value ranging up to 0.257. Interestingly, this also holds for a positive selection on non-western students: this has a positive effect of 0.118, after controlling for selection on high SES[2]. For the selection on students from one-parent households the correlation is negative.

When it comes to the relative preference of the HAVO-schools, we see that the more students with a high SES live in the supply radius of a school, the more the school prefers

[2] The effect size of selection on non-western migrants without controlling for high SES parents is not significant: 0.001.

performance over opportunity. A school that is located in a more urbanised municipality correlates positively with a preference for opportunity over performance than schools in more rural municipalities; the effect being quite substantial ranging up to -0.365 for the most urbanised municipality. Medium (60-120) and large-sized (120 and more) school tracks tend to prefer performance over opportunity, while the reverse holds for tracks that are part of a school community. When schools select more girls than expected on the basis of the supply radius, it correlates with a preference of opportunity over performance. Instead, schools that are selecting more high SES students have a preference of performance over opportunity. The effect size of the last two indicators are also quite relevant, -0.180 and 0.278 respectively. These results suggest that the school community hypothesis is confirmed. Moreover, it seems that the quality selection hypothesis is confirmed, as well as the high SES preference selection hypothesis.

These regressions explain a large percentage of the variance in the outcome measure, with an explained variance for the quality dimension of 0.224 for HAVO and 0.264 for VWO, whereas for the relative preference dimension these numbers are 0.225 and 0.374, respectively.

Table 4. Regressions on Quality and Relative Preferences for VWO

	Quality										Relative Preferences									
VWO	M1				M2			M3			M1			M2			M3			
	b		se	beta	b		se	beta	b		se	beta	b		se	beta	b		se	beta
Constant	0,983		0,760		0,038		0,738		-1,527		0,830		-0,577		1,094		0,169		0,966		-1,397		1,076	
Composition of 5km supply radius																								
% girls	-1,037		1,113	-0,045	-0,968		1,038	-0,042	0,638		1,073	0,027	-0,946		1,602	-0,029	-1,151		1,358	-0,035	-1,311		1,391	-0,040
% students with non-western migration background	1,145		1,101	0,075	0,480		1,053	0,032	0,621		1,049	0,041	-0,865		1,586	-0,040	-0,216		1,378	-0,010	-0,319		1,359	-0,015
% one-parent households	-1,671		1,459	-0,075	-2,078		1,379	-0,093	-2,228		1,411	-0,099	0,600		2,101	0,019	0,696		1,804	0,022	2,081		1,829	0,066
% high SES	1,716	*	0,704	0,125	1,089		0,674	0,079	2,293	**	0,788	0,167	1,730		1,013	0,090	2,111	*	0,882	0,109	4,615	***	1,021	0,239
Urbanisation (ref. Not urban)																								
Somewhat urban	-0,154		0,348	-0,050	0,044		0,327	0,014	0,001		0,319	0,000	0,673		0,500	0,155	0,428		0,428	0,098	0,386		0,413	0,089
Moderately urban	-0,346		0,349	-0,126	-0,089		0,331	-0,032	-0,133		0,323	-0,048	0,702		0,503	0,182	0,342		0,433	0,088	0,204		0,418	0,053
Strongly urban	-0,197		0,357	-0,081	0,174		0,338	0,072	0,116		0,331	0,048	0,953		0,514	0,280	0,308		0,443	0,091	0,135		0,429	0,040
Very strongly urban	-0,301		0,412	-0,107	0,112		0,391	0,040	0,084		0,382	0,030	0,860		0,593	0,218	0,081		0,512	0,020	0,007		0,495	0,002
% recommendation higher than test score	-0,047	***	0,010	-0,246	-0,040	***	0,009	-0,211	-0,038	***	0,009	-0,198	-0,031	*	0,014	-0,116	-0,032	**	0,012	-0,119	-0,018		0,012	-0,068
Track size (ref. (0-60 students)																								
60-120 students					-0,174		0,119	-0,075	-0,197		0,119	-0,085					0,332	*	0,155	0,102	0,107		0,155	0,033
More than 120 students					0,018		0,175	0,005	-0,037		0,174	-0,011					0,541	*	0,229	0,114	0,273		0,226	0,057
Denomination (ref. Public schools)																								
Non-religious					0,300		0,200	0,075	0,247		0,195	0,062					-0,129		0,261	-0,023	-0,250		0,253	-0,045
Religious					0,149		0,117	0,063	0,111		0,115	0,047					-0,361	*	0,153	-0,109	-0,392	**	0,149	-0,118
Single-track schools					0,070		0,250	0,017	0,150		0,248	0,038					1,947	***	0,327	0,348	1,661	***	0,322	0,297
One year comprehensive class					1,115	***	0,196	0,472	1,108	***	0,191	0,469					-0,740	**	0,256	-0,223	-0,778	**	0,247	-0,234
Two year comprehensive class					1,040	***	0,195	0,444	1,051	***	0,191	0,448					-0,724	**	0,256	-0,220	-0,629	*	0,248	-0,191
Social selectivity																								
Difference score girls									0,023	***	0,006	0,181									-0,016	*	0,008	-0,089
Difference score non-western									0,004		0,007	0,030									-0,007		0,009	-0,034
Difference score one-parent households									0,015	*	0,007	0,131									0,040	***	0,010	0,244
Difference score high SES									-0,020		0,010	-0,092									-0,014		0,014	-0,046
R2	0,080				0,215				0,264				0,035				0,320				0,374			
N	429				429				429				429				429				429			

Table 5. Regressions on Quality and Relative Preferences for HAVO

	Quality									Relative Preferences								
HAVO	M1			M2			M3			M1			M2			M3		
	b	se	beta	b	se	beta	b	se	beta	b	se	beta	b	se	beta	b	se	beta
Constant	-0,661	1,052		-1,857	1,064		-3,301 **	1,060		0,206	1,210		1,225	1,215		0,875	1,207	
Composition of 5km supply radius																		
% girls	1,712	1,559	0,053	2,123	1,534	0,066	4,585 **	1,534	0,143	0,752	1,792	0,021	0,308	1,751	0,008	-1,303	1,746	-0,036
% students with non-western migration background	-0,021	0,996	-0,002	0,499	0,988	0,045	1,092	0,952	0,098	1,726	1,145	0,135	1,509	1,128	0,118	1,808	1,084	0,142
% one-parent households	-4,366 **	1,666	-0,194	-4,205 *	1,667	-0,187	-4,453 **	1,656	-0,198	0,162	1,915	0,006	-0,130	1,903	-0,005	-0,168	1,885	-0,007
% high SES	1,784 *	0,865	0,116	2,403 **	0,866	0,156	3,430 ***	0,906	0,223	1,956 *	0,994	0,111	1,246	0,989	0,071	3,079 **	1,032	0,175
Urbanisation (ref. Not urban)																		
Somewhat urban	-0,054	0,351	-0,015	-0,100	0,344	-0,028	-0,063	0,329	-0,018	-0,445	0,404	-0,111	-0,501	0,392	-0,125	-0,494	0,375	-0,124
Moderately urban	0,214	0,348	0,070	0,193	0,342	0,063	0,200	0,329	0,065	-0,819 *	0,401	-0,235	-1,035 **	0,391	-0,296	-0,965 *	0,374	-0,276
Strongly urban	0,348	0,350	0,125	0,344	0,344	0,124	0,392	0,329	0,141	-0,946 *	0,402	-0,299	-1,111 **	0,392	-0,352	-1,092 **	0,375	-0,345
Very strongly urban	0,393	0,427	0,119	0,431	0,420	0,131	0,455	0,404	0,138	-1,369 **	0,491	-0,365	-1,511 **	0,479	-0,403	-1,366 **	0,460	-0,365
% recommendation higher than test score	-0,022	0,011	-0,099	-0,030 **	0,011	-0,135	-0,032 **	0,011	-0,147	-0,035 **	0,013	-0,140	-0,023	0,013	-0,093	-0,016	0,012	-0,063
Track size (ref. (0-60 students)																		
60-120 students				0,091	0,144	0,034	-0,008	0,141	-0,003				0,648 ***	0,164	0,212	0,475 **	0,161	0,155
More than 120 students				0,127	0,231	0,030	-0,024	0,224	-0,006				0,801 **	0,263	0,165	0,612 *	0,255	0,126
Denomination (ref. Public schools)																		
Non-religious				-0,005	0,240	-0,001	0,000	0,233	0,000				0,060	0,274	0,012	-0,103	0,266	-0,020
Religious				-0,231	0,145	-0,086	-0,253	0,140	-0,094				-0,034	0,166	-0,011	-0,076	0,159	-0,025
School community				0,960 ***	0,194	0,242	0,710 ***	0,193	0,179				-0,897 ***	0,222	-0,198	-0,643 **	0,219	-0,142
One year comprehensive class				0,111	0,237	0,042	0,119	0,227	0,045				-0,190	0,270	-0,063	-0,336	0,258	-0,111
Two year comprehensive class				-0,013	0,235	-0,005	0,012	0,225	0,005				-0,272	0,268	-0,091	-0,337	0,256	-0,112
Social selectivity																		
Difference score girls							0,039 ***	0,007	0,257							-0,031 ***	0,008	-0,180
Difference score non-western							0,014 *	0,007	0,118							0,003	0,008	0,018
Difference score one-parent households							0,019 *	0,008	0,148							0,041 ***	0,010	0,278
Difference score high SES							-0,028 **	0,011	-0,132							0,000	0,012	0,001
R2	0,077			0,142			0,224			0,060			0,138			0,225		
N	413			413			413			413			413			413		

Robustness checks

We scrutinize our results by executing a few robustness checks. All results are presented in Appendix 3. First, we replace our performance indicator with an average deviation from the nationwide yearly central exam average for the subjects math and Dutch, to get another measure of performance in secondary education. We observe a similar correlation between opportunity and exam results for both HAVO and VWO and the analyses of relative preference and quality yield somewhat lower explained variances. Regression results yield similar conclusions about our hypotheses. For both HAVO and VWO schools, the exception is that the quality selection hypothesis does not hold anymore when it comes to high SES parents. Moreover, there is a preference for opportunity when selecting on non-western migrants at HAVO.

We also checked whether the results would be different if we analysed the exam results for Dutch and mathematics separately. For both HAVO and VWO we observe a negative correlation with quality for schools that select more non-western students than the supply radius suggests in the case of subject Dutch. For the over-selection of girls, it holds that for both HAVO and VWO it is positively correlated with quality and with a preference for opportunity.

Secondly, we checked for the VWO schools whether offering Latin and Ancient Greek, and thus being considered a gymnasium, mattered for the quality and relative preference. It does not correlate significantly with the quality variable and for relative preference it holds that it correlates with a preference for performance.

5. Conclusions and discussion

In this paper we tried to address the question to what extent there is a trade-off between performance and opportunity in the academic tracks in Dutch secondary education. Our

results suggest that performance and opportunity are only modestly correlated and not more than approximately 55% of the schools experience performance-opportunity trade-offs. This indicates that a trade-off between opportunity and performance in education is not always the case. While this in itself is an important finding, we think we contribute to the literature in two important ways.

The first innovation of our study is that we explicitly model two dimensions on which schools differ: quality and relative preference. Quality can be defined as a situation in which schools simultaneously increase opportunity and performance, while relative preference is defined as increasing one at the cost of the other. We constructed both dimensions in an orthogonal way, so that we can separately look at which school characteristics are related to quality and which are related to relative preferences.

The explained variance of the regressions shows that differences between schools are quite well related with our explanatory variables: the explained variance ranges between 22% and 37%. Given that our limited set of indicators focuses more on structural school characteristics and student composition, and not on the quality of the education process itself, such as learning materials, classroom and teacher characteristics, we suggest that the input of schools together with the structural choices of schools are important when explaining quality, and even more important when explaining the relative preference of a school in pursuing either opportunity or performance. The fact that the relative preference of schools is strongly explained by our analysis, suggests that this is a meaningful dimension alongside the quality dimension. At the same time, we suggest that future research should incorporate indicators on the education process itself.

In the study we used register data to answer our questions. The register data allowed us to get information on all publicly funded schools. Having register data allows for sufficient statistical power and lowers the risk of response bias that often hampers research based on

surveys. The use of register data also allowed for a second innovation in our analyses: how we deal with the effects of student composition. Since there is no catchment area policy in the Netherlands, and governmental funding of schools is equally distributed over schools, the actual school composition of students could be an artefact of the potential student composition in the school's supply radius. However, which students from the supply radius are selected is partly dependent on the policy of the schools. By explicitly distinguishing between the potential student composition in the school's supply radius and the actual student composition as a result of (self-)selection into the school, we get a better understanding of the effects of student composition on school's quality or school's relative preference. The results show that (self-)selection on girls and high SES parents are associated with a higher quality for both HAVO and VWO. A relative preference for opportunity over performance is associated with a stronger (self-)selection on girls, while the opposite is true for a stronger (self-)selection on high SES parents.

The most important conclusion is related to how schools organise track placement in the first few years. The analyses show that in VWO-tracks comprehensive ability classes are associated with a higher quality as well as a relative preference for opportunity. Moreover, in VWO, single-track schools are associated with a relative preference for performance over opportunity. In HAVO-tracks we see something similar. Here schools that are part of school communities (these are schools that offer three tracks or more) are associated with a higher quality as well as a relative preference for opportunity. This can be explained by the fact that such school communities can more easily accommodate between-track mobility than for example single-track schools.

The key message is that our analyses show that schools can evade a possible unwanted trade-off between opportunity and performance by installing comprehensive

classes or being part of a school community. More in general, to organise the track placement such that options to change tracks are still open after the first decision moment.

Research shows that a consequence of early selection into tracks is that socioeconomic inequalities increase (e.g. Chmielewski, 2019; Van de Werfhorst, 2018). The consequential inequalities of selection are said to be tolerable if they stem from meritocratic principles as effort, ability and preference, and not from factors that are beyond the student's influence, such as ethnicity, gender and socioeconomic status of the parents. However, scholars pointed towards the possibility that even then these tolerable inequalities might be problematic, since the selection instruments are prone to error along the lines of the intolerable inequalities (see also Cameron, Daga, & Outhred, 2018; Montt, 2010).

Our results suggest that selection of students into schools indeed induces differences between schools, which violates the meritocratic principle outlined above. Moreover, the finding that flexibility mechanisms within the institutional structure of a school, such as school communities (HAVO) and comprehensive classes (VWO), are fruitful for both giving opportunity without harming performance is an important finding. Moreover, it is at odds with a trend in Dutch secondary education to increase the number of homogeneous first year classes and single-track schools (Inspectorate of Education, 2016). It is also at odds with a common understood assumption that giving opportunities would have a detrimental effect on the quality scores of schools. Our research shows that rather the opposite is true. In other words, our research suggests that reversing the trend towards more homogenous tracked classes could affect both opportunity and performance positively.

References

Allen, J., Bijlsma, I., Borghans, L., & Poulissen, D. (2016). *Schoolkeuzemotieven van ouders en leerlingen in het voortgezet onderwijs. [School choice motives of parents and students in secondary education.]* Maastricht: Research Centre for Education and the Labour Market.

Armor, D. J., Marks, G. N., & Malatinszky, A. (2018). The Impact of School SES on Student Achievement: Evidence From U.S. Statewide Achievement Data. *Educational Evaluation and Policy Analysis, 40*(4), pp. 613-630. doi:10.3102/0162373718787917

Borghans, L., Diris, R., Smits, W., & de Vries, J. (2020). Should we sort it out later? The effect of tracking age on long-run outcomes. *Economics of Education Review, 75*doi:10.1016/j.econedurev.2020.101973

Breen, R., & Goldthorpe, J. H. (1997). Explaining Educational Differentials. *Rationality and Society, 9*(3), pp. 275-305. doi:10.1177/104346397009003002

Bronneman-Helmers, R., Herweijer, L., & Vogels, R. (2002). *Voortgezet onderwijs in de jaren negentig [Secondary education in the nineties]* Den Haag: Sociaal en Cultureel Planbureau.

Brunello, G., Giannini, M., & Ariga, K. (2007). The optimal timing school tracking: A general model with calibration for germany. In L. Woesmann & P. E. Peterson (Eds.), *Schools and the equal opportunity problem*: MIT Press.

Büchner, C., & Van der Velden, R. K. W. (2013). How social background affects educational attainment over time in the Netherlands. In M. Jackson (Ed.), *Determined to succeed?* (pp. 89-115). Stanford, California: Stanford University Press.

Cabus, S. J., & Cornelisz, I. (2017). Competition, student sorting and performance gains in local education markets: The Dutch secondary sector. *European Journal of Education, 52*(3), pp. 365-386. doi:10.1111/ejed.12221

Cameron, S., Daga, R., & Outhred, R. (2018). Setting out a conceptual framework for measuring equity in learning *Handbook on Measuring Equity in Education*. Montreal: UNESCO Institute for Statistics.

Chmielewski, A. K. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. *American Sociological Review, 84*(3), pp. 517-544. doi:10.1177/0003122419847165

De Wolf, I. F., & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education, 33*(3), pp. 379-396. doi:10.1080/03054980701366207

Dustmann, C., Puhani, P. A., & Schönberg, U. (2017). The Long-Term Effects of Early Track Choice. *The Economic Journal, 127*(August), pp. 1348–1380. doi:10.1111/ecoj.12419

Ehren, M. C. M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools—describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability, 25*(1), pp. 3-43. doi:10.1007/s11092-012-9156-4

Everson, K. C. (2017). Value-Added Modeling and Educational Accountability. *Review of Educational Research, 87*(1), pp. 35-70. doi:10.3102/0034654316637199

Forster, A. G., & van de Werfhorst, H. G. (2019). Navigating Institutions: Parents' Knowledge of the Educational System and Students' Success in Education. *European Sociological Review*doi:10.1093/esr/jcz049

Glazerman, S., & Dotter, D. (2017). Market Signals: Evidence on the Determinants and Consequences of School Choice From a Citywide Lottery. *Educational Evaluation and Policy Analysis, 39*(4), pp. 593-619. doi:10.3102/0162373717702964

Haelermans, C., Huijgen, T., Jacobs, M., Levels, M., van der Velden, R., van Vugt, L., & van Wetten, S. (2020). Using Data to Advance Educational Research, Policy, and Practice: Design, Content, and Research Potential of the Netherlands Cohort Study on Education. *European Sociological Review*doi:10.1093/esr/jcaa027

Hanushek, E. A., & Wößmann, L. (2006). Does Educational Tracking Affect Performance And Inequality? Differences-In-Differences Evidence Across Countries. *The Economic Journal, 116*, pp. C63-C76.

Havermans, N., Vanassche, S., & Matthijs, K. (2017). Children's Post-Divorce Living Arrangements and School Engagement: Financial Resources, Parent–Child Relationship, Selectivity and

Stress. *Journal of Child and Family Studies, 26*(12), pp. 3425-3438. doi:10.1007/s10826-017-0843-0
Inspectorate of Education. (2016). *Staat van het Onderwijs. Onderwijsverslag 2014-2015 [State of Education. Education Report 2014-2015]*. Utrecht: Inspectorate of Education.
Inspectorate of Education. (2017). *Onderwijsresultaten Voortgezet Onderwijs 2017 Technische Toelichting [Education Results Secondary Education 2017 Technical Report]*. Utrecht: Inspectorate of Education,.
Inspectorate of Education. (2019). *Staat van het Onderwijs. Onderwijsverslag 2017/2018 [State of Education. Education Report 2017-2018]* Utrecht: Inspectorate of Education.
Jennings, J. L. (2010). School Choice or Schools' Choice? *Sociology of Education, 83*(3), pp. 227-247. doi:10.1177/0038040710375688
Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *PNAS, 103*(27), pp. 10155–10162. doi:10.1073/pnas.0600888103
Korpershoek, H., Naayer, H. M., & Bosker, R. J. (2017). *De inrichting van de onderbouw [The set-up of lower secondary education]* Gronininger: GION Onderwijs/Onderzoek.
Korthals, R. A., & Dronkers, J. (2016). Selection on performance and tracking. *Applied Economics, 48*(30), pp. 2836-2851. doi:10.1080/00036846.2015.1130789
Leist, S. A., & Perry, L. B. (2019). Quantifying segregation on a small scale: how and where locality determines student compositions and outcomes taking Hamburg, Germany, as an example. *School Effectiveness and School Improvement*, pp. 1-25. doi:10.1080/09243453.2019.1688845
LeTendre, G. K., Hofer, B. K., & Shimizu, H. (2003). What Is Tracking? Cultural Expectations in the United States,Germany, and Japan. *American Educational Research Journal, 40*(1), pp. 43-89.
Luyten, H., Schildkamp, K., & Folmer, E. (2009). Cognitive development in Dutch primary education, the impact of individual background and classroom composition. *Educational Research and Evaluation, 15*(3), pp. 265-283. doi:10.1080/13803610902955333
Montt, G. (2010). Cross-national Differences in Educational Achievement Inequality. *Sociology of Education, 84*(1), pp. 49-68. doi:10.1177/0038040710392717
Mühlenweg, A. M. (2007). *Educational Effects of Early or Later Secondary School Tracking in Germany*. ZEW Discussion Paper Series, (79, 7). Mannheim.
Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), pp. 231-256. doi:10.1080/09243453.2014.885451
Nieuwenhuis, J., & Hooimeijer, P. (2016). The association between neighbourhoods and educational achievement, a systematic review and meta-analysis. *J Hous Built Environ, 31*(2), pp. 321-347. doi:10.1007/s10901-015-9460-7 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29355196
OECD. (2016). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. Paris: O. Publishing.
Oomens, M., Scholten, F., & Luyten, H. (2017). *Evaluatie Wet Eindtoetsing PO. Tussenrapportage [Evaluation Primary School End Test Act. Interim Report]* Utrecht: Oberon.
Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), pp. 197-230. doi:10.1080/09243453.2014.885450
Rosenbaum, J. E. (1979). Tournament Mobility: Career Patterns in a Corporation. *Administrative Science Quarterly, 24*(2), pp. 220-241.
Ruijs, N., & Oosterbeek, H. (2019). School Choice in Amsterdam: Which Schools are Chosen When School Choice is Free? *Education Finance and Policy, 14*(1), pp. 1-30. doi:10.1162/edfp_a_00237
Sheskin, D. (2004). *Parametric and Nonparametric statistical procedures* (Third edition ed.) Boca Raton: Chapman & Hall/crc.

Smolkowski, K., & Cummings, K. D. (2015). Evaluation of Diagnostic Systems. *Assessment for Effective Intervention, 41*(1), pp. 41-54. doi:10.1177/1534508415590386

Timmermans, A. C., de Boer, H., Amsing, H. T. A., & van der Werf, M. P. C. (2018). Track recommendation bias: Gender, migration background and SES bias over a 20-year period in the Dutch context. *British Educational Research Journal, 44*(5), pp. 847-874. doi:10.1002/berj.3470

Turner, R. H. (1960). Sponsored and Contest Mobility and the School System. *American Sociological Review, 25*(6), pp. 855-867.

Van de Werfhorst, H. G. (2018). Early tracking and socioeconomic inequality in academic achievement: Studying reforms in nine countries. *Research in Social Stratification and Mobility, 58*, pp. 22-32. doi:10.1016/j.rssm.2018.09.002

Van der Velden, R. K. W. (2011). De effecten van betrouwbaarheid van onderwijsdiploma's op arbeidsproductiviteit: toepassing van een simulatiemodel. [The effects of reliablitity of diploma's on labour productivity: application of a simulation model]. In J. Dronkers (Ed.), *Goede bedoelingen in het onderwijs: Kansen en missers* (pp. 27-50). Amsterdam: Amsterdam University Press.

Van der Velden, R. K. W., & Weßling, K. (forthcoming). Flexibility in Educational Systems - concept, indicators and measures *Oxford Handbook of Social Stratification*. Oxford.

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), pp. 1009-1021. doi:10.1016/j.econedurev.2011.04.014

Van Hek, M., Kraaykamp, G., & Pelzer, B. (2018). Do schools affect girls' and boys' reading performance differently? A multilevel study on the gendered effects of school resources and school practices. *School Effectiveness and School Improvement, 29*(1), pp. 1-21. doi:10.1080/09243453.2017.1382540

Wenger, M., Gärtner, H., & Brunner, M. (2020). To what extent are characteristics of a school's student body, instructional quality, school quality, and school achievement interrelated? *School Effectiveness and School Improvement*, pp. 1-28. doi:10.1080/09243453.2020.1754243

Weßling, K., & Bechler, N. (2019). Where do regional influences matter? The impact of socio-spatial indicators on transitions from secondary school to university. *Review of Regional Research, 39*(2), pp. 163-188. doi:10.1007/s10037-019-00131-5

Appendix 1

Operationalisation dependent variable[3]

Operationalisation exam results

First, we use the deviation from the nationwide within-track average of both Dutch and mathematics on the central exit exams. This is because these courses are mandatory for everyone, and thus useful in comparable regard. The theoretical scale on which the exam results are expressed ranges from 1 to 10. As a first step, we calculate the individual student's deviation from the nationwide average for each school year for Dutch and mathematics separately. Next, we aggregated that to the school level. Lastly, these differences are, equally weighted, averaged together to get a final school difference score. A negative score means that the school scores on Dutch and mathematics are below the nationwide average, a positive score means that these scores are above the nationwide average.

Operationalisation independent variables

If the independent variables were constructed at the individual, student level, they are subsequently aggregated to the school level. Since the timeframe of student observations in the register ranges from 2010/2011 until 2017/2018, the independent variables will be observed in the year they obtain their first secondary education degree. Ethnicity, SES, type of household and gender are variables that are used to count students in school's supply radius. As explained in the main text, we selected only potential students in the school supply radius by looking at students that have a primary school recommendation which would allow access to the track.

[3] The operationalisation of dependent variables performance, opportunity, quality and relative preference are described in the main text.

Ethnicity – The ethnicity of the student was inferred by looking at the country of birth of the student and his/her parents. If the student or at least one of his/her parents was born in one of the countries of the continent of Africa, Latin America and Asia (excluding Indonesia and Japan) or Turkey, the student is considered to be a non-western migrant. The variable was obtained from the register.

Socio-economic status – The paternal household income was measured at the year of obtaining the secondary education diploma in percentiles of the nationwide distribution of incomes and obtained from the register. For the school's supply radius, the percentage of students with a paternal income in the 80^{th} or higher percentile was counted. Since most mothers and fathers live in the same household, the high correlation between the paternal and maternal household income forced us to only use paternal income.

Type of household – This indicator consists of a category of (1) one-parent household, (2) two-parent household and a category (3) other. The categories are derived from the register. We took at the type of household in the year of graduation. The one-parent household percentage in the school's supply radius was subsequently calculated.

Gender – Derived from the register: (0) is male and (1) female. The percentage females in the school's supply radius was subsequently calculated.

Comprehensive ability classes – Although using the third-year entry cohort as a selection criterion, we incorporate the history of the students of a school in that entry cohort, by assigning the school to the type of class they have in the first two years of secondary education. This variable has been obtained by counting the more precise element codes in the registration of the school in the first two years. For each year, we considered a school offering a comprehensive class if more than 10 students in the register were coded as being in a class with at least two adjacent tracks. This threshold of 10 is arbitrary but has been applied because it is very unlikely to have a class with less than 10 students. In the case that less than

10 students are observed, most likely these students were wrongly registered as being in a comprehensive class. The categories of this indicator are (0) no comprehensive class, (1) only first year comprehensive class and (2) first- and second-year comprehensive class[4].

Track size – The track size within the school was inferred by aggregating the number of students that was linked to the school identifier in the register of the same track level in year three. Subsequently, we created a set of dummies: 0-60, 61 – 120, 121 or more, which is more or less equivalent to 1-2 classes, 3-4 classes, 5 or more classes.

Urbanisation level of school municipality – from the register, we linked the urbanisation level of the municipality to the secondary school. The variable contains five categories (1) not urban, (2) somewhat urban, (3) moderately urban, (4) strongly urban and (5) very strongly urban.

Denomination of school – The register contains the specific religious denomination of the school. We grouped them into (0) public, (1) non-religious denomination (2) religious denomination. The public category also included a couple of mixed denomination schools. Analyses with and without those mixed denomination schools yielded no differences in results.

School community – We sorted the schools based on their structure into schools that offer a school community, that is (1) three or more tracks available or (0) less than three tracks available. This indicator is only used at HAVO.

Single-track schools - We sorted the schools based on their structure into schools that offer (1) only one track or (0) more than one track. This indicator is only used at VWO.

[4] It is known that this division is most likely a lower bound effect, since it might be that school register their separate HAVO and VWO classes as being HAVO/VWO classes. Therefore, we only know for sure that the homogeneous classes are homogeneous, but not that heterogeneous classes are heterogeneous per se.

Primary school recommendation versus test score – Some primary schools give systematically higher recommendations than other primary schools. If a secondary education school recruits more students from such lenient primary schools, they are in a disadvantaged situation compared to a secondary education school whose student inflow consists more of students that come from more strict primary schools. To correct for this, we look at the difference between the primary school teachers' recommendation and the recommendation based on test score. The test taken at the end of primary school was the CITO-test. The test was not compulsory but the vast majority of schools took that test (Timmermans, et al., 2018). The scores correspond to a certain level of track that the pupil could enter in secondary education. We counted the percentage of students that had a higher primary school teachers' recommendation than the test recommendation.

Appendix 2

Performance and opportunity

In the VWO track, schools score higher on the performance indicator if their supply radius contains more high SES students, if they are single-track schools and if they select more high SES students than expected in their supply radius. The effect sizes (beta's) of those correlations are also large, between 0.264 and 0.291. The performance of schools is lower when they are confronted with students that come from primary schools that systematically give recommendations that are too high compared to the test results. For the opportunity indicator it holds that schools score lower if in their supply radius there are more students from a one-parent household, if the school is a single-track school, and if they select on high SES students. They score higher if their denomination is religious, when they offer one- or two-year comprehensive tracks and when they have more girls than their supply ratio would suggest. Single-track schools and one- or two-year comprehensive classes have a quite large effect size of -0.220, 0.463 and 0.415 respectively.

For the HAVO track, the results suggest that the performance score is positively correlated with the share of non-western and high SES students in the supply radius[5], with a track size between 60 and 120 students and with schools that select more high SES students than the supply radius composition would suggest. It is negatively correlated with an inflow of students whose recommendation is higher than the end test score would indicate. The effect size of selection on high SES is particularly large, the coefficient is .306. Schools score higher on the opportunity indicator if there are more girls in the supply radius of the school, if

[5] Although on face value this might seem counterintuitive, the negative coefficient of non-western migrants turns positive once SES enters the regression equation. This suggests that a supply radius with more poorer students explain performance and not school supply radius with a high share of non-western students.

the school is located in a more urbanised municipality, when the school is part of school community and when the school selects more girls than the supply radius would suggest. A medium sized school track between 60 and 120 students is negatively correlated with opportunity. Especially urbanisation and more selection of girls have large effect sizes of around .3.

For opportunity and performance at both academic tracks it holds that our explanatory variables in the regression explain substantive amount of variance (for HAVO 0.245 and 0.204 and for VWO 0.439 and 0.236, respectively). Results are to be found in Tables A1 and Table A2.

Table A1 VWO Regressions on Performance & Opportunity

	Performance										Opportunity									
	M1				M2			M3			M1			M2			M3			
	b		se	beta	b		se	beta	b		se	beta	b		se	beta	b		se	beta
Constant	0,203		0,654		0,104		0,656		-1,462	*	0,730		0,780	0,677		-0,065		0,556		-0,065		0,625	
Composition of 5km supply radius																							
% girls	-0,991		0,958	-0,050	-1,060		0,922	-0,053	-0,336		0,943	-0,017	-0,046	0,992	-0,002	0,092		0,782	0,005	0,975		0,808	0,049
% students with non-western migration background	0,140		0,948	0,011	0,132		0,935	0,010	0,151		0,922	0,011	1,005	0,982	0,076	0,348		0,793	0,026	0,470		0,790	0,036
% one-parent households	-0,535		1,256	-0,028	-0,691		1,224	-0,036	-0,074		1,241	-0,004	-1,135	1,301	-0,059	-1,387		1,039	-0,072	-2,155	*	1,063	-0,111
% high SES	1,723	**	0,606	0,145	1,600	**	0,599	0,135	3,454	***	0,692	0,291	-0,007	0,627	-0,001	-0,511		0,508	-0,043	-1,161		0,593	-0,098
Urbanisation (ref. Not urban)																							
Somewhat urban	0,260		0,299	0,097	0,236		0,290	0,088	0,194		0,280	0,072	-0,413	0,310	-0,155	-0,192		0,246	-0,072	-0,192		0,240	-0,072
Moderately urban	0,178		0,301	0,075	0,127		0,294	0,053	0,035		0,284	0,015	-0,524	0,311	-0,221	-0,215		0,249	-0,091	-0,169		0,243	-0,071
Strongly urban	0,378		0,307	0,181	0,241		0,301	0,116	0,125		0,291	0,060	-0,575	0,318	-0,275	-0,067		0,255	-0,032	-0,010		0,249	-0,005
Very strongly urban	0,280		0,355	0,115	0,096		0,347	0,040	0,045		0,336	0,019	-0,581	0,367	-0,240	0,015		0,295	0,006	0,038		0,287	0,016
% recommendation higher than test score	-0,039	***	0,008	-0,237	-0,036	***	0,008	-0,220	-0,028	***	0,008	-0,170	-0,008	0,009	-0,048	-0,004		0,007	-0,025	-0,010		0,007	-0,059
Track size (ref. (0-60 students)																							
60-120 students					0,079		0,105	0,040	-0,045		0,105	-0,022				-0,253	**	0,089	-0,126	-0,152		0,090	-0,076
More than 120 students					0,279		0,156	0,096	0,118		0,153	0,040				-0,262	*	0,132	-0,090	-0,155		0,131	-0,053
Denomination (ref. Public schools)																							
Non-religious					0,085		0,177	0,025	-0,002		0,172	0,000				0,214		0,150	0,062	0,249		0,147	0,072
Religious					-0,106		0,104	-0,052	-0,141		0,101	-0,069				0,255	**	0,088	0,125	0,252	**	0,087	0,124
Single-track schools					1,008	***	0,222	0,293	0,905	***	0,218	0,264				-0,939	***	0,188	-0,273	-0,756	***	0,187	-0,220
One year comprehensive class					0,187		0,174	0,092	0,165		0,168	0,081				0,927	***	0,147	0,455	0,943	***	0,144	0,463
Two year comprehensive class					0,158		0,174	0,078	0,211		0,168	0,104				0,882	***	0,147	0,436	0,840	***	0,144	0,415
Social selectivity																							
Difference score girls									0,004		0,005	0,033								0,020	***	0,005	0,177
Difference score non-western									-0,001		0,006	-0,010								0,006		0,005	0,045
Difference score one-parent households									0,028	***	0,007	0,275								-0,012	*	0,006	-0,123
Difference score high SES									-0,017		0,009	-0,091								-0,003		0,008	-0,016
R2	0,083				0,169				0,236				0,017			0,401				0,439			
N	429				429				429				429			429				429			

Table A2 HAVO Regressions on Performance & Opportunity

	Performance										Opportunity									
	M1				M2			M3			M1			M2			M3			
	b		se	beta	b		se	beta	b		se	beta	b		se	beta	b		se	beta
Constant	-0,227		0,789		-0,316		0,811		-1,213		0,814		-0,433		0,814		-1,541		0,804		-2,088	**	0,793	
Composition of 5km supply radius																								
% girls	1,232		1,168	0,051	1,216		1,169	0,050	1,641		1,177	0,067	0,480		1,206	0,020	0,908		1,158	0,037	2,944	*	1,147	0,121
% students with non-western migration background	0,853		0,747	0,101	1,004		0,753	0,118	1,450	*	0,731	0,171	-0,873		0,771	-0,103	-0,505		0,746	-0,060	-0,358		0,712	-0,042
% one-parent households	-2,102		1,249	-0,123	-2,168		1,271	-0,127	-2,311		1,271	-0,135	-2,264		1,289	-0,133	-2,038		1,259	-0,119	-2,143		1,238	-0,125
% high SES	1,870	**	0,648	0,160	1,824	**	0,660	0,156	3,254	***	0,696	0,279	-0,086		0,669	-0,007	0,579		0,654	0,050	0,175		0,678	0,015
Urbanisation (ref. Not urban)																								
Somewhat urban	-0,249		0,263	-0,094	-0,301		0,262	-0,113	-0,279		0,253	-0,105	0,196		0,272	0,073	0,201		0,260	0,075	0,216		0,246	0,081
Moderately urban	-0,303		0,261	-0,130	-0,421		0,261	-0,181	-0,382		0,252	-0,165	0,517		0,270	0,222	0,614	*	0,259	0,264	0,582	*	0,246	0,251
Strongly urban	-0,299		0,262	-0,142	-0,384		0,262	-0,183	-0,350		0,253	-0,166	0,647	*	0,271	0,307	0,727	**	0,260	0,346	0,742	**	0,246	0,353
Very strongly urban	-0,488		0,320	-0,196	-0,540		0,320	-0,217	-0,456		0,310	-0,183	0,881	**	0,330	0,353	0,971	**	0,317	0,390	0,910	**	0,302	0,365
% recommendation higher than test score	-0,029	***	0,008	-0,171	-0,027	**	0,009	-0,159	-0,024	**	0,008	-0,144	0,007		0,009	0,040	-0,003		0,009	-0,020	-0,008		0,008	-0,050
Track size (ref. (0-60 students)																								
60-120 students					0,370	***	0,110	0,181	0,233	*	0,108	0,115					-0,279	*	0,109	-0,137	-0,242	*	0,106	-0,119
More than 120 students					0,464	**	0,176	0,143	0,294		0,172	0,091					-0,337		0,174	-0,104	-0,318		0,168	-0,098
Denomination (ref. Public schools)																								
Non-religious					0,028		0,183	0,008	-0,052		0,179	-0,015					-0,033		0,182	-0,010	0,051		0,174	0,015
Religious					-0,133		0,111	-0,065	-0,165		0,107	-0,081					-0,098		0,110	-0,048	-0,089		0,104	-0,043
School community					0,032		0,148	0,010	0,033		0,148	0,011					0,929	***	0,147	0,309	0,677	***	0,144	0,225
One year comprehensive class					-0,039		0,180	-0,019	-0,109		0,174	-0,054					0,150		0,179	0,075	0,228		0,170	0,113
Two year comprehensive class					-0,143		0,179	-0,071	-0,162		0,172	-0,081					0,130		0,177	0,065	0,174		0,168	0,087
Social selectivity																								
Difference score girls									0,004		0,006	0,034									0,035	***	0,006	0,305
Difference score non-western									0,008		0,005	0,092									0,006		0,005	0,064
Difference score one-parent households									0,030	***	0,006	0,306									-0,011		0,006	-0,112
Difference score high SES									-0,014		0,008	-0,086									-0,014		0,008	-0,088
R2	0,097				0,132				0,204				0,038				0,148				0,245			
N	413				413				413				413				413				413			

Appendix 3 Results robustness checks

Table A3. VWO Regressions Exam, Exam Quality and Exam Relative preferences

	Exam				Exam Quality				Exam Relative Preferences			
	Model 3				Model 3				Model 3			
	b		se	beta	b		se	beta	b		se	beta
(Constant)	-0,756		0,776		-0,821		0,907		-0,691		1,079	
Composition of 5km supply radius												
% girls	0,103		1,003	0,005	1,077		1,172	0,045	-0,872		1,395	-0,027
% students with non-western migration background	-0,400		0,980	-0,030	0,070		1,145	0,005	-0,869		1,363	-0,041
% one-parent households	-0,758		1,319	-0,039	-2,913		1,541	-0,127	1,396		1,834	0,045
% high SES	2,085	**	0,736	0,176	0,924		0,860	0,066	3,246	**	1,024	0,170
Urbanisation (ref. Not urban)												
Somewhat urban	-0,114		0,298	-0,043	-0,307		0,348	-0,097	0,078		0,414	0,018
Moderately urban	-0,119		0,302	-0,050	-0,288		0,353	-0,103	0,049		0,420	0,013
Strongly urban	-0,025		0,309	-0,012	-0,035		0,361	-0,014	-0,015		0,430	-0,005
Very strongly urban	0,097		0,357	0,040	0,135		0,417	0,047	0,058		0,496	0,015
% recommendation higher than test score	-0,011		0,009	-0,066	-0,021	*	0,010	-0,106	-0,001		0,012	-0,004
Track size (ref. (0-60 students)												
60-120 students	0,230	*	0,112	0,115	0,078		0,131	0,033	0,382	*	0,155	0,118
More than 120 students	0,094		0,163	0,032	-0,061		0,190	-0,018	0,249		0,227	0,053
Denomination (ref. Public schools)												
Non-religious	-0,126		0,183	-0,037	0,122		0,213	0,030	-0,375		0,254	-0,068
Religious	0,011		0,108	0,005	0,262	*	0,126	0,109	-0,241		0,150	-0,073
Single-track schools	0,466	*	0,232	0,136	-0,289		0,271	-0,071	1,222	***	0,323	0,221
One year comprehensive class	-0,205		0,178	-0,101	0,738	***	0,208	0,306	-1,148	***	0,248	-0,349
Two year comprehensive class	-0,284		0,179	-0,140	0,556	**	0,209	0,232	-1,124	***	0,248	-0,345
Social selectivity												
Difference score girls	-0,004		0,006	-0,032	0,016	*	0,007	0,123	-0,023	**	0,008	-0,130
Difference score non-western	-0,009		0,007	-0,076	-0,004		0,008	-0,026	-0,015		0,009	-0,075
Difference score one-parent households	0,005		0,007	0,049	-0,007		0,008	-0,062	0,017		0,010	0,107
Difference score high SES	-0,007		0,010	-0,037	-0,010		0,011	-0,045	-0,004		0,014	-0,013
N	413				413				413			
R2	0,136				0,158				0,357			

Table A4. Regressions on Quality and Relative Preferences dimensions for Language and Math subjects separately at VWO track.

	Language Quality				Language Relative Preferences				Math Quality				Math Relative Preferences			
	Model 3				Model 3				Model 3				Model 3			
	b		se	beta	b		se	beta	b		se	beta	b		se	beta
(Constant)	0,322		0,929		0,452		1,033		-0,980		0,932		-0,849		1,054	
Composition of 5km supply radius																
% girls	1,812		1,200	0,071	-0,138		1,335	-0,004	0,591		1,205	0,024	-1,358		1,363	-0,043
% students with non-western migration background	-2,773	*	1,173	-0,166	-3,712	**	1,305	-0,183	1,189		1,177	0,073	0,250		1,332	0,012
% one-parent households	-2,080		1,579	-0,084	2,229		1,756	0,075	-3,705	*	1,584	-0,155	0,604		1,793	0,020
% high SES	-1,140		0,881	-0,076	1,182		0,980	0,065	1,752	*	0,884	0,120	4,075	***	1,000	0,218
Urbanisation (ref. Not urban)																
Somewhat urban	-0,325		0,356	-0,095	0,060		0,396	0,015	-0,236		0,358	-0,072	0,149		0,405	0,035
Moderately urban	-0,319		0,361	-0,106	0,018		0,402	0,005	-0,303		0,362	-0,104	0,034		0,410	0,009
Strongly urban	-0,001		0,370	0,000	0,019		0,411	0,006	-0,085		0,371	-0,033	-0,066		0,420	-0,020
Very strongly urban	0,000		0,427	0,000	-0,076		0,475	-0,020	0,132		0,429	0,044	0,056		0,485	0,015
% recommendation higher than test score	-0,024	*	0,010	-0,114	-0,004		0,011	-0,017	-0,021	*	0,010	-0,104	-0,002		0,012	-0,006
Track size (ref. (0-60 students)																
60-120 students	-0,123		0,134	-0,048	0,182		0,149	0,059	0,069		0,134	0,028	0,373	*	0,152	0,118
More than 120 students	-0,370		0,195	-0,099	-0,060		0,217	-0,013	-0,004		0,196	-0,001	0,306		0,221	0,067
Denomination (ref. Public schools)																
Non-religious	0,104		0,219	0,024	-0,393		0,243	-0,074	0,117		0,219	0,028	-0,380		0,248	-0,070
Religious	0,263	*	0,129	0,101	-0,240		0,143	-0,076	0,240		0,129	0,095	-0,263		0,146	-0,082
Single-track schools	-0,265		0,278	-0,061	1,247	***	0,309	0,235	-0,528		0,279	-0,125	0,984	**	0,316	0,182
One year comprehensive class	0,731	***	0,213	0,282	-1,155	***	0,237	-0,367	0,744	***	0,214	0,296	-1,142	***	0,242	-0,356
Two year comprehensive class	0,832	***	0,214	0,323	-0,848	***	0,238	-0,272	0,455	*	0,214	0,182	-1,226	***	0,243	-0,385
Social selectivity																
Difference score girls	0,020	**	0,007	0,141	-0,020	*	0,008	-0,114	0,015	*	0,007	0,112	-0,024	**	0,008	-0,138
Difference score non-western	-0,028	***	0,008	-0,181	-0,040	***	0,009	-0,208	0,012		0,008	0,081	0,001		0,009	0,006
Difference score one-parent households	-0,015		0,008	-0,118	0,010		0,009	0,062	0,000		0,008	-0,001	0,025	**	0,010	0,156
Difference score high SES	-0,015		0,012	-0,065	-0,010		0,013	-0,033	-0,012		0,012	-0,050	-0,006		0,013	-0,019
N	413				413				413				413			
R2	0,236				0,357				0,180				0,356			

Table A5. VWO Regressions on Quality and Relative Preferences with Gymnasium indicator

	Gymnasium Quality				Gymnasium Relative Preferences			
	Model 3				Model 3			
	b		se	beta	b		se	beta
(Constant)	-1,463		0,831		-1,647		1,060	
Composition of 5km supply radius								
% girls	0,594		1,073	0,026	-1,137		1,368	-0,035
% students with non-western migration background	0,664		1,049	0,044	-0,489		1,337	-0,023
% one-parent households	-2,168		1,411	-0,097	1,844		1,800	0,059
% high SES	2,287	**	0,787	0,166	4,637	***	1,004	0,240
Urbanisation (ref. Not urban)								
Somewhat urban	-0,014		0,319	-0,005	0,446		0,406	0,102
Moderately urban	-0,153		0,323	-0,056	0,280		0,412	0,073
Strongly urban	0,094		0,331	0,039	0,219		0,422	0,064
Very strongly urban	0,062		0,382	0,022	0,091		0,487	0,023
% recommendation higher than test score	-0,037	***	0,009	-0,193	-0,022		0,012	-0,081
Track size (ref. (0-60 students)								
60-120 students	-0,173		0,121	-0,075	0,014		0,154	0,004
More than 120 students	-0,014		0,175	-0,004	0,181		0,223	0,038
Denomination (ref. Public schools)								
Non-religious	0,260		0,196	0,065	-0,302		0,249	-0,054
Religious	0,115		0,115	0,049	-0,408	**	0,147	-0,123
Single-track schools	0,375		0,306	0,094	0,783	*	0,390	0,140
% of track gymnasium	-0,003		0,003	-0,085	0,013	***	0,003	0,235
One year comprehensive class	1,100	***	0,191	0,465	-0,747	**	0,243	-0,225
Two year comprehensive class	1,023	***	0,192	0,436	-0,519	*	0,245	-0,158
Social selectivity								
Difference score girls	0,023	***	0,006	0,176	-0,014		0,008	-0,074
Difference score non-western	0,004		0,007	0,031	-0,007		0,009	-0,036
Difference score one-parent households	0,015	*	0,007	0,130	0,041	***	0,010	0,248
Difference score high SES	-0,019		0,011	-0,086	-0,019		0,013	-0,061
N	413				413			
R2	0,267				0,397			

Table A6. HAVO Regression with Exam averages as dependent variable and incorporated in the Quality and Relative Preferences dimensions.

	Exam				Exam Quality				Exam Relative Preferences			
	Model 3				Model 3				Model 3			
	b		se	beta	b		se	beta	b		se	beta
(Constant)	-0,935		0,853		-3,023	**	1,058		1,153		1,262	
Composition of 5km supply radius												
% girls	0,236		1,234	0,010	3,180	*	1,530	0,100	-2,708		1,825	-0,073
% students with non-western migration background	-0,318		0,766	-0,038	-0,676		0,950	-0,061	0,039		1,133	0,003
% one-parent households	-1,016		1,332	-0,059	-3,159		1,652	-0,142	1,127		1,971	0,043
% high SES	1,527	*	0,729	0,131	1,703		0,904	0,112	1,352		1,079	0,076
Urbanisation (ref. Not urban)												
Somewhat urban	0,152		0,265	0,057	0,368		0,329	0,106	-0,064		0,392	-0,016
Moderately urban	0,053		0,264	0,023	0,635		0,328	0,210	-0,529		0,391	-0,150
Strongly urban	0,042		0,265	0,020	0,784	*	0,329	0,286	-0,700		0,392	-0,219
Very strongly urban	0,106		0,325	0,043	1,017	*	0,403	0,313	-0,804		0,481	-0,212
% recommendation higher than test score	-0,009		0,009	-0,054	-0,017		0,011	-0,080	-0,001		0,013	-0,002
Track size (ref. (0-60 students)												
60-120 students	0,117		0,114	0,057	-0,124		0,141	-0,047	0,359	*	0,168	0,116
More than 120 students	-0,102		0,180	-0,031	-0,420		0,224	-0,100	0,216		0,267	0,044
Denomination (ref. Public schools)												
Non-religious	-0,192		0,188	-0,057	-0,141		0,233	-0,032	-0,243		0,278	-0,047
Religious	0,185		0,112	0,091	0,096		0,139	0,036	0,274		0,166	0,088
School community	0,221		0,155	0,073	0,898	***	0,192	0,229	-0,456	*	0,229	-0,100
One year comprehensive class	0,104		0,183	0,052	0,332		0,227	0,126	-0,124		0,270	-0,040
Two year comprehensive class	0,195		0,181	0,098	0,369		0,224	0,142	0,021		0,267	0,007
Social selectivity												
Difference score girls	-0,002		0,006	-0,018	0,033	***	0,007	0,221	-0,037	***	0,009	-0,213
Difference score non-western	-0,017	**	0,006	-0,184	-0,011		0,007	-0,092	-0,022	**	0,008	-0,163
Difference score one-parent households	-0,001		0,007	-0,014	-0,012		0,008	-0,097	0,010		0,010	0,064
Difference score high SES	-0,013		0,008	-0,079	-0,027	*	0,010	-0,129	0,001		0,012	0,006
N	413				413				413			
R2	0,126				0,207				0,169			

Table A7. Regressions on Quality and Relative Preferences dimensions for Language and Math subjects separately at HAVO track.

	Language Quality				Language Relative Preferences				Math Quality				Math Relative Preferences			
	Model 3				Model 3				Model 3				Model 3			
	b		se	beta	b		se	beta	b		se	beta	b		se	beta
(Constant)	-2,782	**	1,074		1,394		1,237		-2,273	*	1,090		1,903		1,219	
Composition of 5km supply radius																
% girls	4,026	**	1,553	0,123	-1,862		1,790	-0,052	2,125		1,577	0,064	-3,763	*	1,763	-0,105
% students with non-western migration background	-2,371	*	0,964	-0,208	-1,655		1,111	-0,132	0,641		0,979	0,056	1,357		1,094	0,109
% one-parent households	-0,199		1,677	-0,009	4,087	*	1,933	0,161	-6,307	***	1,703	-0,272	-2,021		1,904	-0,081
% high SES	-0,047		0,918	-0,003	-0,398		1,058	-0,023	2,037	*	0,932	0,129	1,686		1,042	0,098
Urbanisation (ref. Not urban)																
Somewhat urban	0,245		0,333	0,069	-0,186		0,384	-0,047	0,380		0,339	0,105	-0,052		0,378	-0,013
Moderately urban	0,640		0,333	0,205	-0,525		0,384	-0,152	0,638		0,338	0,202	-0,527		0,378	-0,154
Strongly urban	0,776	*	0,334	0,275	-0,708		0,384	-0,227	0,758	*	0,339	0,266	-0,726		0,379	-0,235
Very strongly urban	1,022	*	0,409	0,305	-0,799		0,471	-0,216	1,049	*	0,415	0,310	-0,772		0,464	-0,211
% recommendation higher than test score	-0,012		0,011	-0,052	0,005		0,013	0,020	-0,024	*	0,011	-0,107	-0,008		0,013	-0,031
Track size (ref. (0-60 students)																
60-120 students	-0,273		0,143	-0,100	0,210		0,165	0,069	-0,086		0,145	-0,031	0,397	*	0,162	0,133
More than 120 students	-0,471	*	0,227	-0,108	0,165		0,262	0,034	-0,417		0,230	-0,095	0,219		0,258	0,046
Denomination (ref. Public schools)																
Non-religious	-0,094		0,236	-0,021	-0,196		0,272	-0,039	-0,239		0,240	-0,052	-0,342		0,268	-0,069
Religious	0,090		0,142	0,033	0,267		0,163	0,088	-0,001		0,144	0,000	0,176		0,161	0,059
School community	0,877	***	0,195	0,217	-0,477	*	0,225	-0,107	1,072	***	0,198	0,262	-0,282		0,221	-0,064
One year comprehensive class	0,212		0,230	0,078	-0,244		0,265	-0,082	0,389		0,233	0,142	-0,067		0,261	-0,023
Two year comprehensive class	0,197		0,228	0,073	-0,152		0,262	-0,051	0,394		0,231	0,145	0,046		0,258	0,016
Social selectivity																
Difference score girls	0,050	***	0,007	0,320	-0,021	*	0,009	-0,121	0,024	**	0,008	0,155	-0,046	***	0,008	-0,272
Difference score non-western	-0,020	**	0,007	-0,165	-0,032	***	0,008	-0,236	0,002		0,007	0,013	-0,010		0,008	-0,075
Difference score one-parent households	-0,015		0,008	-0,109	0,008		0,010	0,052	-0,011		0,009	-0,078	0,012		0,010	0,079
Difference score high SES	-0,022	*	0,011	-0,105	0,006		0,012	0,024	-0,028	**	0,011	-0,130	0,000		0,012	0,000
N	413				413				413				413			
R2	0,232				0,162				0,224				0,173			