ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis [Economie en Bedrijfseconomie]

The effects of Track Recommendations on the secondary school career

Name student: Karzan Schippers

Student ID number: 548390

Supervisor: Bastian Ravesteijn

Second assessor: Coen van de Kraats

Date final version:

1. Abstract

Students in The Netherlands are sorted relatively early into separate educational tracks before the beginning of secondary school. Track recommendations are given by the primary school and a final primary school test called the CET can be used to increase this track recommendation. I exploit the thresholds on this test to perform a 'fuzzy' RDD to estimate the causal effect of receiving the havo (senior general) and vwo (pre-university) track recommendations on track attendance and the chance to repeat a grade during secondary school. I find strong, statistically significant effects across the board with an exception for receiving a vwo track recommendation on the chance to repeat a grade. Policy implications are discussed.

2. Introduction:

The practice of grouping students by ability (tracking) is controversial to say the least. Doubly so for countries where students are put in tracks where they no longer interact with those outside their designated educational pathways. This is often the case in early tracking systems where students are 'sorted' as early as the end of elementary school (ages 10-12). More subtle tracking systems can be found across the globe. In the UK or US, it is common for students to essentially self-select into more academically challenging tracks or classes. In contrast, certain areas in Germany make their early track recommendations legally binding and even when students are not legally bound, it is rare that they deviate far from their recommended tracks, at least while entering secondary school (Dustmann et al., 2017).

Due to the nature of early tracking misallocations are more likely to occur than if tracking was delayed. Such misallocations may be very inefficient in an educational system that does not allow its students to shift across tracks easily. Grade retention costs the state money per student and there are the costs of missed wages to be considered when delaying school completion. Furthermore, if students at the margin between two track recommendations (who by assumption are essentially the same) experience vastly different chances to attend certain tracks this would signal there to be likely misallocation issues that are not easily fixed.

Dustmann et al., (2017) investigates the effect of attending a more advanced track in Germany by exploiting a quasi-random shift between tracks due to date of birth and finds no evidence for favourable long term outcomes. Borghans et al., (2019) performs a similar study in The Netherlands (but using test scores as the running variable) and finds positive effects on

earnings. Both of the previous studies have some issues. Firstly, both studies only examine the effects of attending a higher track on later in life outcomes. Missing out on shorter-term outcomes of interest such as the potential for grade retention in students at the margin. Furthermore, results are less informing for policy decisions as the educational systems in both countries will have already gone through multiple iterations of reforms by the time the studies publish their findings. Thirdly, Borgans et al.,(2019) suffers from potentially unreliable data as they find many missing datapoints and end up imputing test scores based on earlier tests.

This paper focuses on estimating the effect of specific teacher track recommendations on secondary school track attendance and the chance of grade retention for recent cohorts. Chapter 3 reviews the current literature regarding tracking and elaborates on the ways the Dutch educational system works. Chapter 4 and 5 outline the research methodology and data used. Chapter 6 provides a robustness test and examines the characteristics of parental income in the sample. Chapter 7 presents the results while chapter 8 discuss policy implications and limitations. Finally, chapter 9 is my conclusion.

3. Background

3.1 Current literature

Research on tracking has traditionally focused on estimating the overall effect of having different tracking systems in place and how having those systems influenced school performance, educational completion rates and inequality. Gamoran and Mare (1989) found evidence by controlling for observable characteristics that supports the notion that tracking reduces inequality and that students are sorted to the track that provide the greatest reward to them. More robust results have been found exploiting differences across countries with different educational systems that suggest that early tracking increase educational inequality by improving the achievements of better performing students which comes at a cost to those already performing more poorly than their peers (Brunello & Checchi, 2007; Hanushek & Wößmann, 2006). Surprisingly, more recent studies have found that an increase in the amount of comprehension in a school system increases the amount of people eligible for higher tertiary education such as university but does not change post-secondary school outcomes such as earnings or enrolment in university (Hall, 2012; Malamud & Pop-Eleches, 2011). The studies listed above examine the overall effects of different levels of tracking both within and between countries. They do not examine the effects of attending a specific track at an allocation margin

such as test scores. It becomes especially important to know more about the potential costs of misallocation at the margin of such allocation. As the risk of such errors are likely the highest then. A wide array of studies have found evidence that teachers are biased in their grading and track recommendations regarding gender, Socioeconomic status and ethnicity (Dijks et al., 2020; Geven et al., 2018; Glock et al., 2013; Nürnberger et al., 2016; Sprietsma, 2013; Timmermans et al., 2015; Van Leest et al., 2021). If there are substantial effects of attending certain tracks on long-term outcomes misallocations as a result of previously mentioned bias may partially explain general inequality of opportunities present in society. There are a few studies that focus on the long-term effects of attending specific tracks during secondary school. As discussed in the introduction, Dustmann et al., (2017) finds no evidence for attending certain tracks to have a positive effect on long term educational outcomes and earnings in Germany. They attribute this to students having a lot of opportunity to swap tracks during and after their secondary school career. While Borghans et al., (2019) does find positive effects on post-education earnings in The Netherlands. These studies estimate such effects in an attempt to estimate the value of attending certain tracks for students at the allocation margin. But equally important is examining the effect of teacher's track recommendations to understand if they have a significant effect on secondary school outcomes. Depending on the result it may become more or less pressing that teachers are biased in their track recommendations for example.

3.2 *The Dutch Education System*

The Netherlands employs an educational system with multiple tracks where tracking begins once students finish primary school after six years (and two in kindergarten) and lasts until the end of tertiary education.

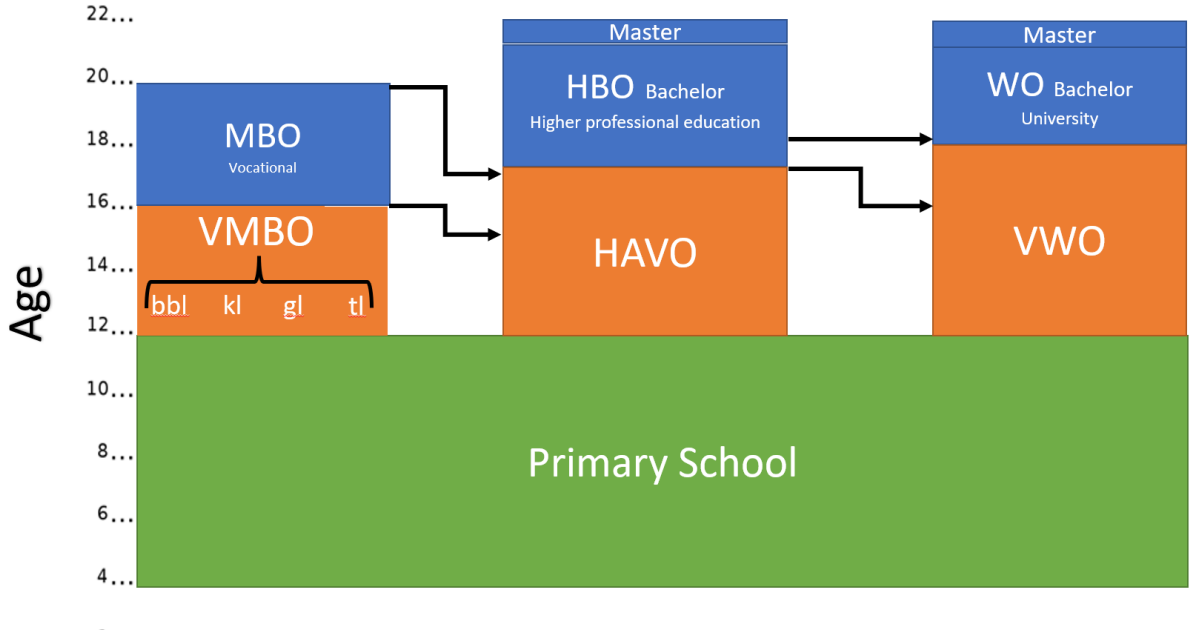

Figure 1. The Dutch education system

Notes. The figure gives an overview of the Dutch education system and the expected ages of students at various points throughout. Once students graduate primary school, they enter secondary school and are (eventually) sorted into separate tracks. These tracks are: practical education, vmbo, havo and vwo. Practical education is excluded from this paper due to its unique structure. Those who attend the vmbo sub tracks can be further subdivided into vmbo-bbl, vmbo-kl, vmbo-gl and vmbo-tl. Vmbo tracks usually take four years to complete. The havo track is designed to last five years and vwo six. It is possible to be held back or downgrade tracks during secondary school. To upgrade tracks towards the last few years of secondary school it is common for students to first finish their current track and then redo their graduation year, this time at the upward adjacent track and then complete that track's graduation year before they can graduate once more. It is also possible to swap tertiary tracks as shown in the figure but this is not relevant to this paper.

Figure 1 shows a general overview of the Dutch education system and some of the paths that students may take to move to a more advanced track. Secondary schools may offer multiple or only singular tracks. Those who finish the vmbo track (which consists of four different tracks) after four years move on to vocational schools. Havo graduates are prepared for higher professional education over the span of five years. The six-year vwo track prepares for university. Track selection usually happens before secondary school starts, but may be postponed until secondary school through the process of students attending mixed classes where multiple, adjacent tracks are grouped together (CBS, 2022). These mixed classes usually last one or two years but may take three in some cases too. Afterwards students are sorted into final, singular tracks. Students may swap tracks in either direction or repeat a grade during their time in secondary school. Repeating grades is more common in the havo track, as 38 % of havo

students experience grade retainment during secondary school, compared to vmbo's 27% and vwo's 23%. (Vuuren & Wiel, 2015) Grade retention is most common during the year preceding graduation (year 3 for vmbo, 4 for havo and 5 for vwo). Shifting tracks upwards outside the early years of secondary school is usually only possible after completing the track, repeating the same year at the upper track and then graduating the year afterwards. (A year 4 vmbo-tl student may graduate, move on to year 4 havo and graduate once more during havo's fifth year)

Educational paths start to diverge towards the end of primary school once a student's teacher recommends a track appropriate for them. This could be a crucial point in these students' lives. A better recommendation may land someone in a different track or school, which could result in a different tertiary institute and through that a different job or adult life. These recommendations are formed in a three-step process. First, the primary school makes an initial recommendation based on their experience with the student and previous test results. This needs to be done before the 1st of March. Second, students take a compulsory final test between the 15th of March and the 15th of April, which tests the students in arithmetic and language skills and calculates a final standardized score. This score corresponds to a track recommendation as well. Third, the primary school determines a final recommendation. This final recommendation may not be a lower track than the initial recommendation. If the track recommendation resulting from the final test is equal or lower to the one from the initial recommendation then the final recommendation equals the initial one. They must reconsider their initial recommendation if the student's test recommendation is higher than the school's. This include 'half levels' like havo being adjusted to havo/vwo (Rijksoverheid, 2022). This reconsideration does not guarantee an upwards adjustment and the new recommendation as a result of the reconsideration does not have to be as advanced as the track suggested by the final primary school test score.

Table 1. Scores and corresponding track recommendations

Track recommendation	Centrale Eindtoets score
Vmbo-bbl	501-518
Vmbo-bbl and vmbo-kl	519-525
Vmbo-kl	526-528
Vmbo-gl/vmbo-tl	529-532
Vmbo-gl/vmbo-tl and havo	533-536
Havo	537-539
Havo/vwo	540-544
Vwo	545-550

Notes. The table shows which scores correspond to which track recommendations for the 2015/2016 schoolyear for the Centrale Eindtoets (CET). Which is the test taken at the end of primary school to help determine what track students should be sorted into for secondary school. The values in this table are sourced from the Centrale Eindtoets 2016 yearly report.

The compulsory final test and reconsiderations came to be in April 2015 due to an administrative change in The Netherlands. Before this it was not uncommon for students to take a final test but it was not compulsory. Note that the 2014/2015 schoolyear was a transitionary period and the new policies did not truly become mandatory until the 2015/2016 school year. Also, as a part of the administrative change came the introduction of a new final primary school test called the Centrale Eindtoets' (CET), which replaced the old 'Cito toets'. In the years following the introduction of the new mandatory tests it has become more common for primary schools to not use the CET, which is the test most approved by the government. While in 2015 85% of primary schools used the CET, by 2018 this had dropped to 56% (CPB, 2019) Instead they often pick other approved final tests like DIA, Route 8, IEP eindtoets and AMN eindtoets. Since the 2018/2019 schoolyear all final tests use the same recommendation categories, but this was not the case in the years that this paper examines. (CPB, 2019) Table 1 depicts which scores correspond with separate track recommendations for the CET test in 2016. The threshold values between track recommendations may vary slightly between years. But this is rare in the sample used for this paper.

4. Empirical Strategy

The thresholds and their associated track recommendations as seen in Table 1 generate an excellent opportunity to study the effects of track recommendations on multiple outcome variables. In the remainder of this section, I will explain how I will estimate the causal effect of receiving a higher secondary school recommendation. Second, I will discuss the relevant assumptions.

4.1 The Empirical framework

The goal of this thesis is to estimate the effect of receiving a higher secondary school recommendation on track attendance and grade retention during secondary school. The outcome equation describing this relationship is estimated separately for each track and once for grade retention:

$$Y_i = \beta_0 + \beta_1 R_i + \epsilon_i \quad (1)$$

Where the dependent variable Y_i either reflects the educational track student *i* attended during secondary school or if the student repeated a grade during secondary school. R_i is a sample specific dummy that equals 1 for student *i* if their final track recommendation was at least the track we estimate the equation for and ϵ_i is the error term. Estimating a simple regression like this is clearly tainted with selection bias. It is unlikely that students with specific track recommendations do not differ in many observable and unobservable ways with students with other track recommendations. Simply running this regression therefore yields a biased estimator. To solve this issue we use a fuzzy regression discontinuity design (Imbens & Lemieux, 2008; Lee & Lemieux, 2010) To this end I estimate the first stage, which estimates the effect of crossing a recommendation threshold conditional on CET score on track recommendation received.

$$R_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 C_i + \mu_i \tag{2}$$

Where Z_i is a dummy that equals 1 if student i's CET score is at or above the threshold value needed for their score to correspond with the track recommendation we are estimating the effect for in (1), C_i functions as the running variable for student i's final test score and is relative to the relevant cut-off. μ_i is the error term. I now add the reduced form equation which estimates the effect of crossing a recommendation threshold on our outcome variables.

$$Y_i = \delta_0 + \delta_1 Z_i + \delta_2 C_i + \omega_i \tag{3}$$

Where ω_i is the error term. It is then possible to estimate the causal effect of R_i on Y_i by simply dividing the reduced form by the first stage through the use of Two-stage least squares (2SLS):

$$\widehat{\beta_1} = \frac{\widehat{\delta_1}}{\widehat{\gamma_1}} \tag{4}$$

4.2 Assumptions

The key identifying assumption is that the potential outcomes in educational track taken by the students runs smooth through our threshold values. Conditional on the running variable students should be as good as randomly assigned around the threshold value. This is also known as the independence assumption and is slightly weaker in our study than if we had used non rounded CET standard scores. The assumption could be violated if students or parents would be able to influence which side of the threshold they fall on. While both groups would have considerable incentives to attempt such an action (and hopefully increase the student's

track), they would somehow have to either cheat during the CET examination or manage to influence the student's official score after the test. Both of these seem quite infeasible, especially with the presence of mixed track classes in early secondary school which provides additional opportunities for students to promote track-wise. We do test this assumption by examining the parental income of children around both sides of the thresholds in chapter 4.

The second assumption of note is the exclusion restriction. Here it is important that a student's CET score being at or above the track's score threshold only affects the outcome variables through the track recommendation they received at the end of primary school. This assumption seems relatively safe to make. One way for the assumption to be violated would be for secondary school teachers to be influenced by their student's CET score. It is possible for secondary school teachers to make decisions about a student's track or if the student should repeat a grade and also be able to see their CET score (and through that, know about if a student scored above a track threshold). Thus, these teachers might partially base their decisions regarding the student's secondary school educational experience on if the student scored in a specific track threshold. However, there are some mitigating circumstances. First, secondary school teachers likely base their decisions much more on their own experiences with their students instead of a CET score multiple years old. Second, since when I estimate the effect of getting a track recommendation, I use bandwidths that do not encroach on CET scores corresponding with other track intervals. This eliminates any compliers that score at other track intervals, vastly reducing the odds of secondary school teachers being truly influenced by their student's CET score.

5 Data

5.1 Sources

For the entire empirical analysis, I use 5 separate non-public microdata datasets from Statistics Netherlands (CBS). The primary dataset used is called inschrwpotab and concerns primary school track recommendation data and CET scores and I use four separate cohorts for analysis. Those that have taken the CET In 2015, 2016, 2017 and 2018. I then merge this dataset with one called onderwijsinschrtab, concerning characteristics of educational institutes to obtain data regarding which school year students are in on October 1st between 0 and 5 years after they took the CET. Then the data is merged once more with a dataset called hoogsteopltab which contains information regarding the highest level of education people have taken on October 1st between 0 and 5 years after they took the CET. Not all of these datapoint are

available for each cohort since not enough time has passed for all of them. Finally, the children are matched with their parents through the usage of the kindoudertab dataset and use this to find the parents yearly gross income during the year their child took the CET in a dataset called inpatab. After removing all observations where any of the variables of interest were missing the final dataset contains 486,406 observations of which between 90,000 and 150,000 belong to each cohort. The reason for this disparity between cohort sizes is that the later cohorts have more students taking other final primary school tests besides the CET.

5.2 Descriptive Statistics

Here I depict some information regarding the data and set up the further analysis in chapter 4 and 5. First, Figure 2 depicts the relative distribution of CET scores for the entire dataset. Note that these CET scores are rounded to whole numbers. The distribution of these scores is generally the same across cohorts but some spikes at specific CET scores do occur in the 2016 and 2017 cohorts, this is likely due to the rounding to whole numbers. Individual histograms for each cohort can be found in the appendix. Second, see Table 2 for statistics regarding the data. Notable is that around 6.2% of the sample has attained an increase in track recommendation after taking the CET. This number is the lowest in 2015 as expected since the administrative change discussed earlier was not fully rolled out yet during the first year in our sample. Furthermore, as seen in Table 3, almost one-fifth of our sample consists of students who received an initial recommendation of vwo, the highest track that has no way to upgrade its recommendation through the CET. Table 3 depicts a relative distribution table to see how such increases in track recommendation are distributed. It shows that mixed track recommendations are rarer than their singular-track counterparts both in initial and final recommendations, that track increases are mostly limited to a one-track adjustment and that per track between fifteen and five percent experience an upward adjustment in their track recommendation (with the exception of vwo). The appendix includes the same table sporting the absolute frequencies instead of percentages.

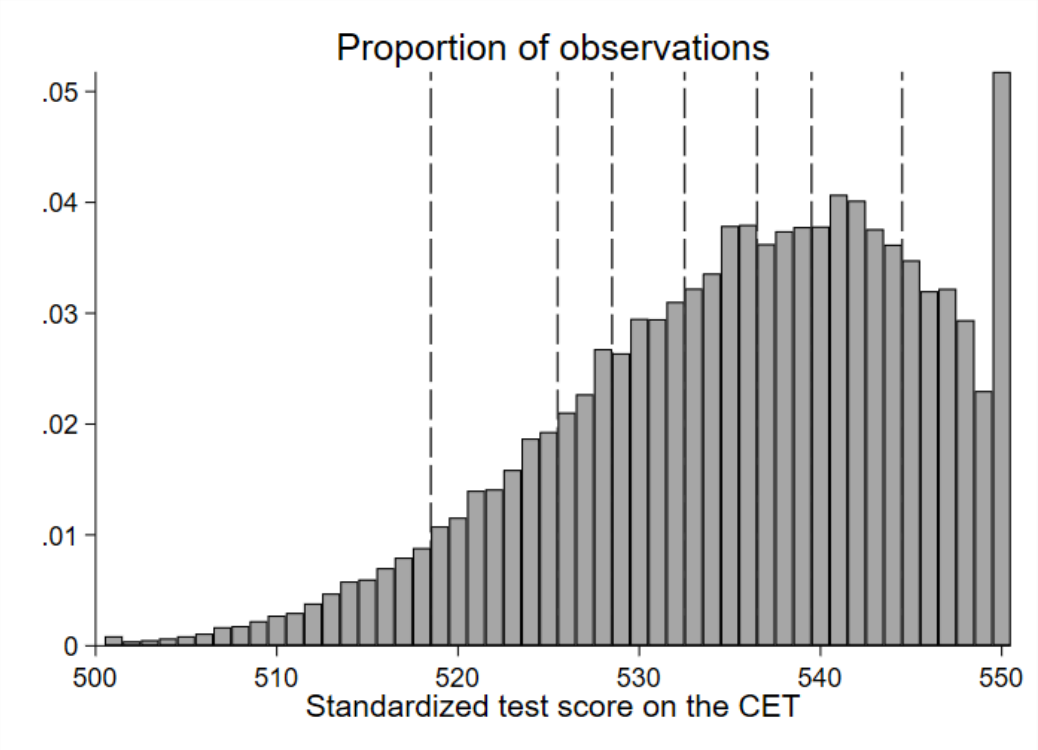

Figure 2. Distribution of CET scores

Notes. This graph depicts the distribution of CET scores in the sample. The sample uses data from the 2014/2015 school year until the 2017/2018 school year. CET scores range from 501-550 and scores correspond to specific secondary school track recommendations. The thresholds between track recommendations are marked by dashed vertical lines before scores: 519, 526, 529, 533, 537, 540 and 545. During the 2014/2015 school year the threshold found at 519 was 521 and the 533 threshold was 534.

Table 2: Descriptive statistics

	Mean	**N**
CET score	535.502 (9.826)	486,406
Parental Income	98189.25 (79265.92)	486,406
Track recommendation increase	0.062 (0.242)	486,406

Notes. This table depicts information on various variables for those that took the CET test from the 2014/2015 school year until the 2017/2018 school year. Means are shown and standard deviations can be found in brackets below the means. CET scores can range from 501 to 550. Parental income is gross annual income of the legal father and mother in the year the student took the CET. Track recommendation increase is a dummy that equals one if the student's final track recommendation exceeded their initial one.

Next, I examine the track recommendation increases further. Figure 3 is a binned scatter plot showing the relationship between the proportion of students receiving such a track recommendation increase per CET score. The figure also shows the thresholds that can be found in Table 1. Since 2015 had different threshold values for the first and fourth thresholds Figure 3 only shows cohorts 2016, 2017 and 2018 pooled. The same graph for individual cohorts can be found in the appendix. Notably, we see a relatively strong 'pseudo first-stage' at every threshold, which is a good early sign for my later analysis. The figure also displays an increase per threshold in how steep the drop off is for the likelihood of students to experience a track recommendation increase past the threshold. A possible explanation for this phenomenon is that those that score higher on the CET are also more likely to already have an initial track recommendation at or above the track corresponding to the threshold found below their CET score. This then also makes them less likely to qualify for a track recommendation increase. Additionally, the thresholds found at higher CET scores have less room for upward outliers since they are closer to the maximum score allowed.

Table 3. The distribution of final track recommendations

Initial Track recommendation %	*Bbl*	*Bbl-kl*	*Kl*	*Kl-gl*	*Kl-tl*	*Gl*	*Gl-tl*	*Gl-havo*	*Tl*	*Tl-havo*	*Havo*	*Havo/vwo*	*Vwo*	*TOTAL*
Bbl	93.98	2.4	3.25	0.05	0.15	<0.03	0.06	<0.03	0.21	<0.03	<0.03	<0.03	0.00	100.00
Bbl-kl	0.00	92.38	3.64	0.28	1.00	0.23	0.75	<0.07	1.45	0.16	<0.07	<0.07	<0.07	100.00
Kl	0.00	0.00	90.18	0.25	1.54	0.58	0.82	0.04	6.18	0.33	0.07	<0.02	<0.02	100.00
Kl-gl	0.00	0.00	0.00	89.31	<0.50	1.86	3.06	<0.50	3.81	0.70	<0.50	<0.50	<0.50	100.00
Kl-tl	0.00	0.00	0.00	0.00	85.65	0.17	0.75	0.20	11.08	1.64	0.35	0.16	<0.09	100.00
Gl	0.00	0.00	0.00	0.00	0.00	93.64	1.07	0.90	2.87	0.71	0.77	<0.22	<0.22	100.00
Gl-tl	0.00	0.00	0.00	0.00	0.00	0.00	93.47	0.86	1.23	1.71	2.60	<0.12	<0.12	100.00
Gl-havo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	89.48	<0.57	<0.57	7.56	2.39	<0.57	100.00
Tl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	91.88	4.22	3.71	0.17	0.02	100.00
Tl-havo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	87.50	10.79	1.66	0.06	100.00
Havo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	94.27	3.85	1.88	100.00
Havo/vwo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	93.62	6.38	100.00
Vwo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
Total	6.14	2.83	10.15	0.40	2.24	0.98	1.78	0.36	17.50	6.64	20.52	8.46	22.00	100.00

Notes. This table depicts relative percentages of initial track recommendations and how they are distributed

along final track recommendations for students that took the Centrale Eindtoets (CET) from the 2014/2015 school year until the 2017/2018 school year. Cells with fewer than ten observations have been suppressed and some total values are rounded slightly for privacy purposes.

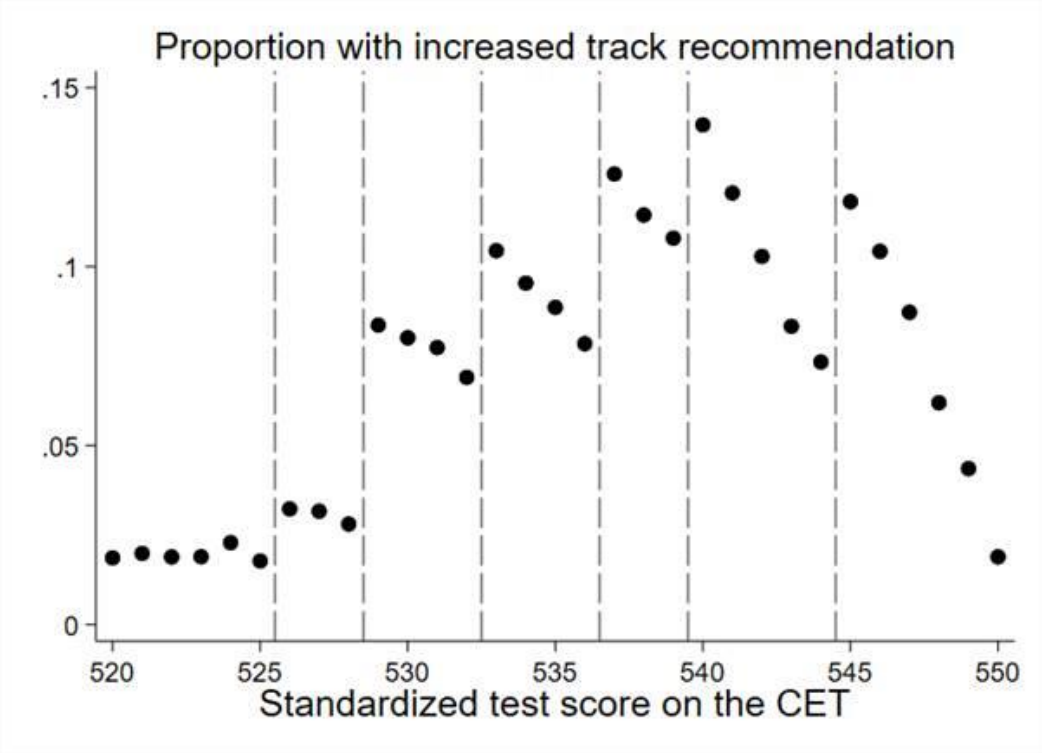

Figure 3. Increases in track recommendation

Notes. This graph depicts the proportion of students per CET score that experienced a track recommendation increase. This is the case if a student's final track recommendation after the CET is higher than their initial track recommendation. The graph consists of data from the 2015/2016 school year until the 2017/2018 school year. Thresholds can be found at certain scores which are the minimum value for a specific track recommendation category on the CET. All thresholds were identical throughout these years and dashed vertical lines at CET scores 519, 526, 529, 533, 537, 540 and 545 indicate where these thresholds are placed.

5.2.1 Removal of outliers

Besides removing all observations from the sample that had a missing value for any of the variables it was also necessary to remove some unrealistic values from the data. Due to merging two separate datasets, one containing data regarding which secondary school grade a student is in and one regarding a student's highest level of education followed up to that point I am able to cross reference the data to itself and remove all observations with conflicting datapoints. To this end I removed two observations who were reported to reside in the fifth grade of secondary school 2 years after they took the CET but also had a highest level of education followed of vmbo at the time. Since vmbo can only last four years this is not possible.

Furthermore, I removed 311 students who were listed to go straight to tertiary education and 575 students whose highest level of education followed was claimed to be elementary school mere months after the CET. 2,675 observations were removed due to conflicting information within the dataset regarding school grade or highest level of education followed. These observations mostly consisted of students being listed as attending tertiary education while they were simultaneously listed as presiding in a secondary school grade and students whose highest level of education followed was downgraded at some point. Which goes against the purpose of this variable. Note that all descriptive statistics seen above are based on the data post removal of these outliers.

6. Parental income

Here I examine the relationship between parental income and CET scores. Parental income is defined as the combined annual gross income of the legal father and mother on January 1st during the year their child took the CET. Parental income is then sorted into 100 percentiles by comparing it to the other parental incomes during the same CET examination year. As a first step I perform a visual analysis of percentile parental income rank around the various thresholds in our sample. I then move on to further exploration of parental income in my sample.

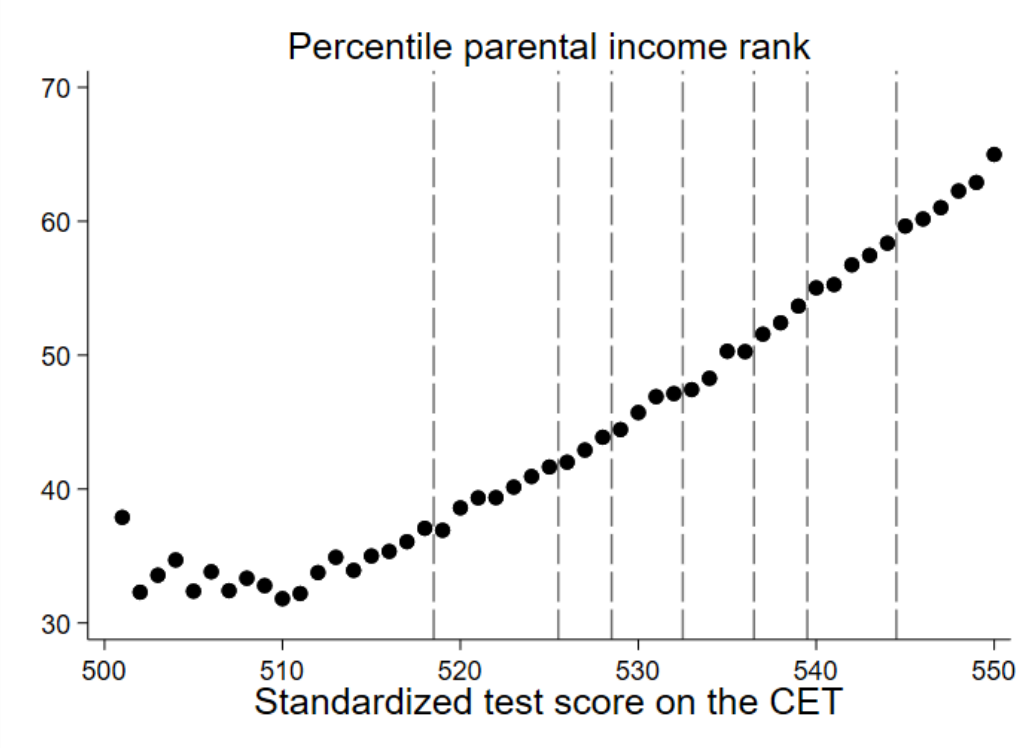

Figure 4. Relative parental income for test takers.

Notes. This graph depicts the average percentile parental income of students per CET score. Percentile parental income is gross income of the father and mother which is then split into 100 percentiles relative to all other parental incomes during the year the child took the CET. The graph consists of data from students who took the CET from the 2015/2016 until 2017/2018 school years. Thresholds can be found at CET scores 519, 526, 529, 533, 537, 540 and 545 which are the minimum values for specific track recommendation categories on the CET, they are represented by dashed lines.

Figure 4 examines the average percentile parental income per CET score for the 2016, 2017 and 2018 cohorts. The 2015 cohort is excluded from the graph due to having two thresholds that differ slightly. The same graph for the 2015 cohort can be found in the appendix. The percentiles are relative to the other parents whose children took the CET during that year. The figure also tests the independence assumption of parents and students not being able to affect which side of the thresholds they end up on. The figure shows a relatively smooth line where the average percentile of parental income is higher for those that score higher on the CET. Such correlations are to be expected. Note that the lower ends of the CET score distribution contain relatively few observations which may be a cause for the non-smooth look at the earlier scores. The figure also seems to show support for the independence assumption as the percentile of parental income is not particularly higher directly above the thresholds compared to directly below them. It seems unlikely that richer parents are able to influence which side of the threshold their child's CET score ends up on. In our dataset there is a negative association between a student's percentile parental income and their odds to receive a track recommendation increase after the CET. The graph depicting this relationship can be found in the appendix. The coefficient of this decrease is -0.0004. On average a child whose parents have a percentile parental income one percentile higher will be associated with having a 0.04% smaller chance to receive a track recommendation increase. This association may seem counter-intuitive at first. But it makes perfect sense once one considers that the children of the richest parents likely did not need such an increase in track recommendation in the first place as they are more likely to already have received the highest possible track recommendation before the CET. Further graphs depicting the average test score per percentile parental income and distribution of income in my sample can be found in the appendix. The 10^{th} percentile of the income distribution has a median gross yearly income of around €28,679 while the 90^{th} percentile has one of around €164,616.

7. Results

In this section I visually and statistically examine the first stage, reduced form and final IV estimates for multiple outcome variables, thresholds and sample compositions.

7.1 The First Stage

Table 4 depicts the first stages for each of the seven thresholds shown in Table 1. It consists of seven different regressions, each regressing a dummy that equals 1 if a student's final track recommendation is at least the track specified in the column on the threshold dummy that equals 1 if a student's CET score was high enough to qualify on the CET for the track specified in the column. All regressions use OLS and heteroskedasticity-robust standard errors. Two control variables are added, one controls for the student's CET score relative to the cut-off. This also compensates for the threshold values being slightly different for column (1) and (4) in 2015 compared to the rest of the sample. The second is an interaction variable between the threshold dummy and the relative CET score. This allows for the slopes of our regression to be different on both sides of the threshold. Bandwidths used are specified and are chosen in an attempt to only ever include one cut-off. Statistical analysis finds five out of the seven first stages to be statistically significant. Of these I will use two for further analysis. These two both have sufficiently strong first stages and both are composed of singular track recommendations. This is useful since we are mostly looking for effects regarding outcomes for specific tracks that occur later during secondary school compared to mixed tracks that seize to exist once students are sorted further.

Table 4: The first stage

	(1) Bb/Kl	(2) Kader	(3) Gl/Tl	(4) Gl/Tl/Havo	(5) Havo	(6) Havo/Vwo	(7) Vwo
Threshold	0.055***	-0.002	0.024**	0.008	0.061***	0.041***	0.071***
	(0.011)	(0.010)	(0.011)	(0.006)	(0.009)	(0.007)	(0.005)
Test Score	0.024***	0.049***	0.061***	0.053***	0.065***	0.048***	0.079***
	(0.005)	(0.006)	0.006)	(0.002)	(0.005)	(0.004)	(0.001)
Threshold*Test Score	0.013**	-0.022***	-0.003	-0.015***	0.014**	0.033***	0.001
	(0.005)	(0.006)	(0.007)	(0.003)	(0.006)	(0.005)	(0.002)
Constant	0.645***	0.860***	0.592***	0.808***	0.494***	0.276***	0.497***
	(0.010)	(0.009)	(0.010)	(0.005)	(0.008)	(0.007)	(0.004)
Number of observations	40,435	52,783	80,680	94,227	91,106	94,308	148,851
Prop. before threshold	0.633	0.811	0.531	0.756	0.429	0.228	0.099
Bandwidth	517-523	524-528	527-532	530-535	535-539	538-542	541-549

Notes. This table depicts the first stage regressions for seven different thresholds. The sample consists of students who took the Centrale Eindtoets (CET) between the 2014/2015 and 2017/2018 schoolyears. The outcome variables are dummies that equal 1 if a student receives a final track recommendation at least at the level specified in the column. Threshold is a dummy that equals 1 if a student's test score was at least the threshold value needed to qualify for the track recommendation category on the CET. Test Score reflects the CET score which ranges from 501-550. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The first can be found in column (5) and uses the cut-off between the mixed recommendation vmbo-gl/tl and havo and the singular havo recommendation. The second resides in column (7) and exploits the cut-off between the havo/vwo recommendation and the vwo recommendation. In the later parts of the analysis, I use these same thresholds but for a sample composed of the 2015, 2016 and 2017 cohorts. Those additional first stages can be found in the appendix but the coefficient for the havo threshold shifts from 0.061 to 0.054 and for the vwo threshold it shifts from 0.071 to 0.077. The first stage becoming slightly weaker when excluding the 2018 cohort is expected as the amount of track recommendation increases after students took the CET was also the strongest in the 2018 cohort. Below in Figure 5 and 6 the first stages that represent column (5) and (7) respectively are shown visually. The appendix includes additional graphs with an increased range of CET scores shown and for the sample composition that excludes 2015.

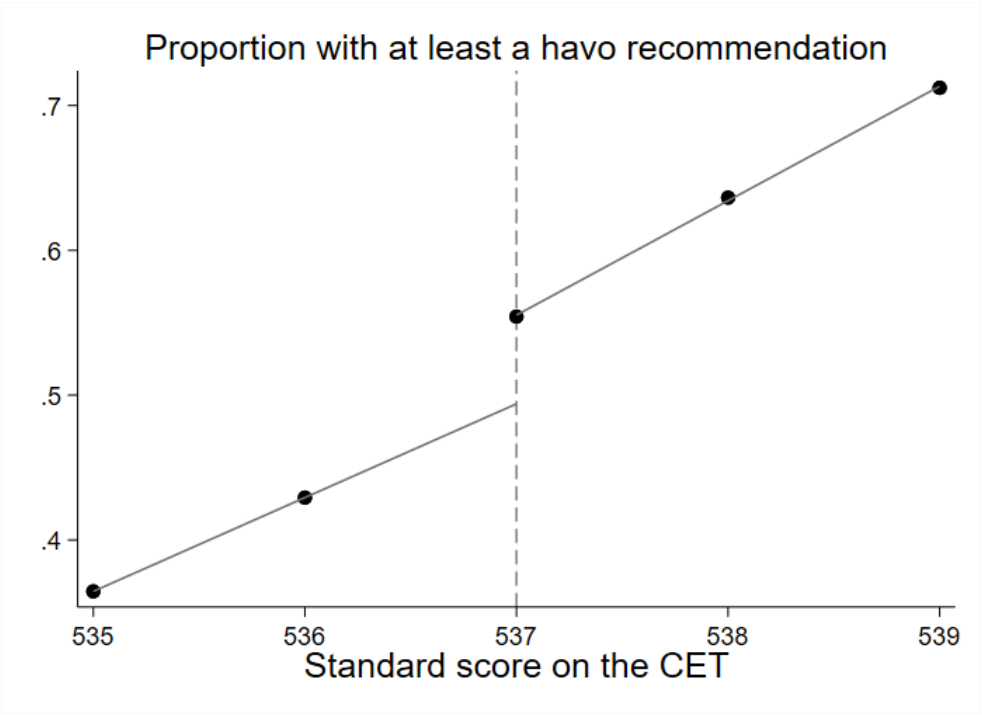

Figure 5. The First stage for the havo threshold

Notes. This figure depicts the proportion of students who received at least a havo recommendation as their final track recommendation for the 2014/2015 until 2017/2018 school years per Centrale Eindtoets (CET) score. Students may receive this final track recommendation before the CET or after. The score 537 is marked with a dashed line due to this being the lowest value that is associated with the havo track for the CET.

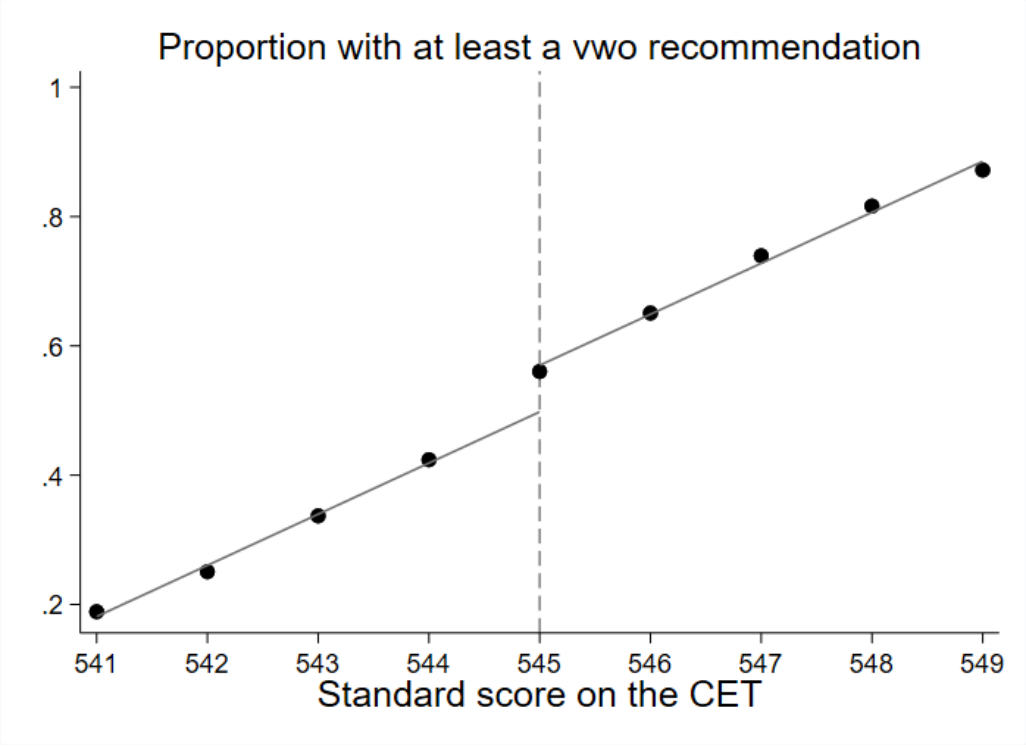

Figure 6. The first stage for the vwo threshold

Notes. This figure depicts the proportion of students who received at least a vwo recommendation as their final track recommendation for the 2014/2015 until 2017/2018 school years per Centrale Eindtoets (CET) score. Students may receive this final track recommendation before the CET or after. The score 545 is marked with a dashed line due to this being the lowest value that is associated with the vwo track for the CET.

7.2 Reduced Form

The statistical analysis for the reduced forms is done similarly to the first stages. Regressions are done using OLS and heteroskedasticity-robust standard errors. Relative test scores, threshold dummies and an interaction variable between the two are used as independent variables for the same reasons as before. Table 5 consists of four columns each with a separate outcome dummy that equals 1 if the highest level of education a student has taken so far by their third or fourth year in secondary school is at least the track specified in the column. Table 6 consists of three separate reduced forms that regress if the student has ever repeated a grade by a certain year on the same independent variables as in Table 4 and 5. Students are considered to have repeated a grade by a certain year in secondary school if they have progressed slower through secondary school than one grade a year. Students who finished a vmbo or havo track and then moved on to then attend the havo or vwo track respectively are not considered to repeat a grade even if they technically spend two years at the same grade. Promoting tracks is a positive thing and should not be classified as repeating a grade, which is generally seen as a negative. There are three years of grade data available for the 2018 cohort, four for the 2017 cohort, five for 2016 and six for the 2015 cohort. Variables such as 'repeated a grade by year 4' uses four years of data for each student in the sample. Students with missing data are assumed to not repeat a grade for those missing years if they have not yet done so in the data. This means that all coefficients for these grade retainment variables are minimum values since some students who I assume do not experience grade retainment do so in reality. Table 5 shows statistically significant estimates for the effect of crossing the cut-offs on having attended at least the track specified in the column by a certain year. Note that 'at least havo attended by year 4' and 'at least vwo attended by year 4' means that the student had to attend the senior (bovenbouw) version of the track which students attend once they follow any track during grade 4 of secondary school. Since students can generally only reach grade 4 of secondary school after at least three years these outcome dummies essentially only equal one if the student attends at least the track specified without repeating a grade. Column (1) uses an outcome dummy that equals one if the highest track attended by a student was either havo, havo/vwo or vwo by their third year of secondary school. However, since students may enter secondary

school at these tracks and the variable only shows the highest track attended the coefficient in column (1) tells us more about the initial differences in track attendance and how much the group at the lower side of the threshold managed to catchup in track attendance after two years of secondary school. Column (4) and (5) are the same but column (5) uses a smaller bandwidth to see if the estimates change when reducing the bandwidth.

Table 5: The reduced form for Track attendance

	(1) Havo/vwo by year 3	(2) Havo by year 4	(3) Vwo by year 4	(4) Vwo by year 4
Threshold	0.035***	0.032***	0.028***	0.026**
	(0.009)	(0.010)	(0.006)	(0.010)
Test Score	0.056***	0.046***	0.068***	0.069***
	(0.005)	(0.006)	(0.002)	(0.006)
Threshold*Test Score	0.002	0.017***	-0.002	-0.005
	(0.006)	(0.006)	(0.002)	(0.007)
Constant	0.530***	0.425***	0.520***	0.523***
	(0.008)	(0.009)	(0.005)	(0.009)
Number of observations	91,106	71,542	117,856	66,111
Bandwidth	535-539	535-539	541-549	543-547

Notes. This table depicts the reduced form regressions for two different thresholds. The sample consists of students who took the Centrale Eindtoets (CET) between the 2014/2015 and 2017/2018 schoolyears. The outcome variables are dummies that equal 1 if a student has attended least at the track specified in the column by the time, they start the year specified in the column. Threshold is a dummy that equals 1 if a student's test score was at least the threshold value needed to qualify for the track recommendation category on the CET. Test Score reflects the CET score which ranges from 501-550. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

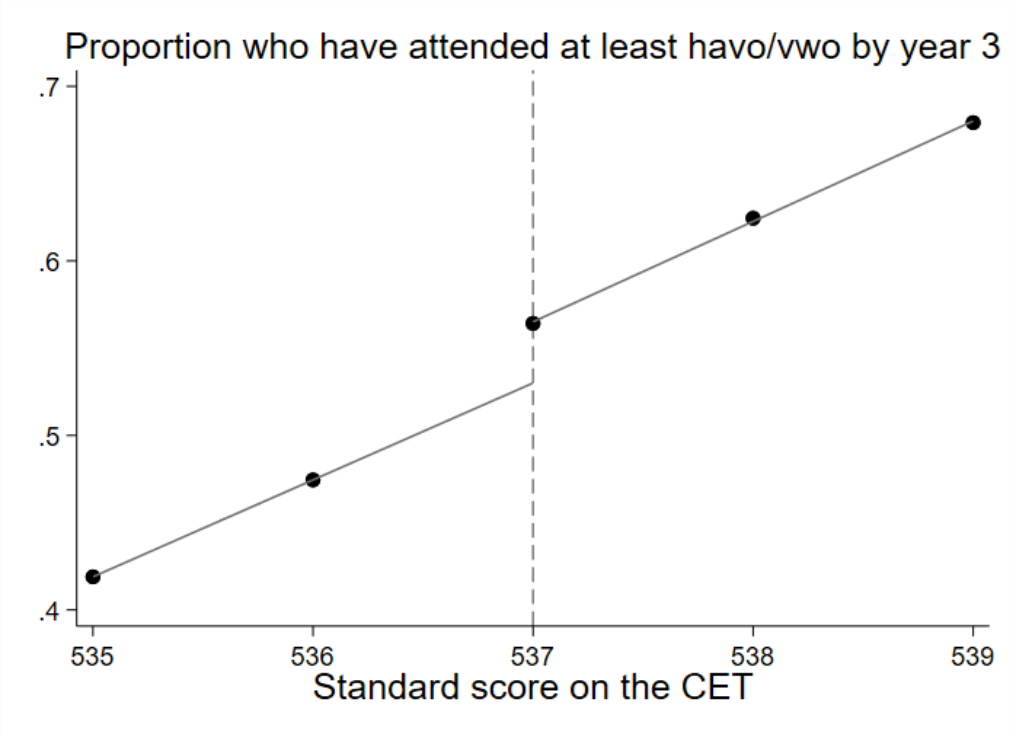

Figure 7. The reduced form for havo/vwo attendance

Notes. This figure depicts the proportion of students who attended at least havo/vwo by year 3 of secondary school for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2017/2018 school years per CET score. The outcome variable is measured on October 1st, during the students' third school year. The bandwidth used is 535-539 with the threshold residing at 537

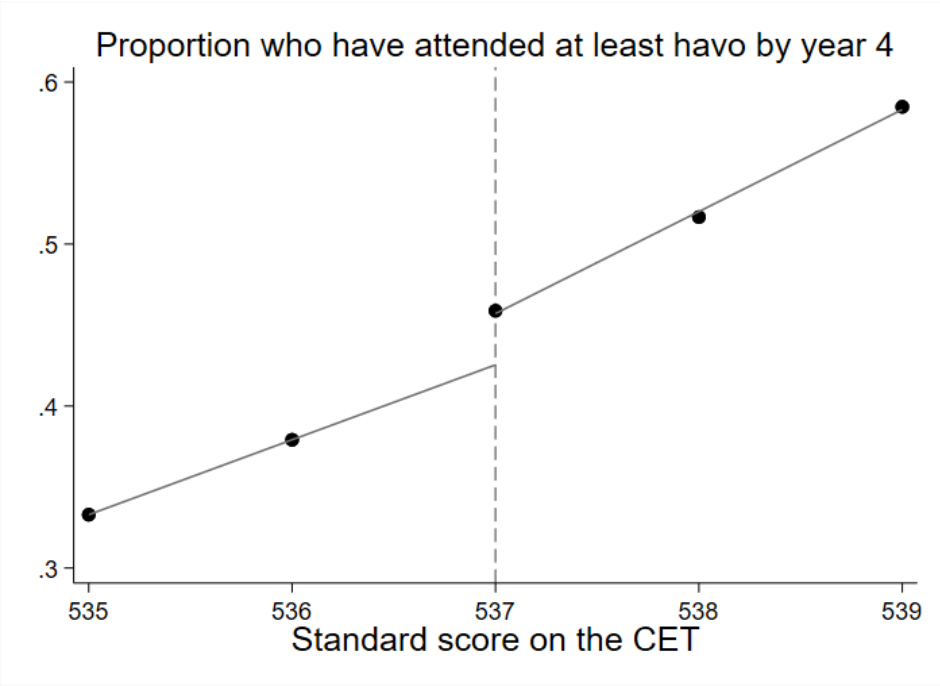

Figure 8. The reduced form for havo attendance in year 4

Notes. This figure depicts the proportion of students who attended at least the havo senior track by year 4 of secondary school for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2016/2017 school years per CET score. The outcome variable is measured on October 1st, during the students' fourth school year. The bandwidth used is 535-539 with the threshold residing at 537

Even though column (5) uses a sample size almost half the size of column (4) the estimates are not very different, indicating the significance and sign of the results are robust to changes to the bandwidth around the vwo threshold. Figures 7 through 10 give compelling visual evidence for jumps in the outcome variables at the thresholds.

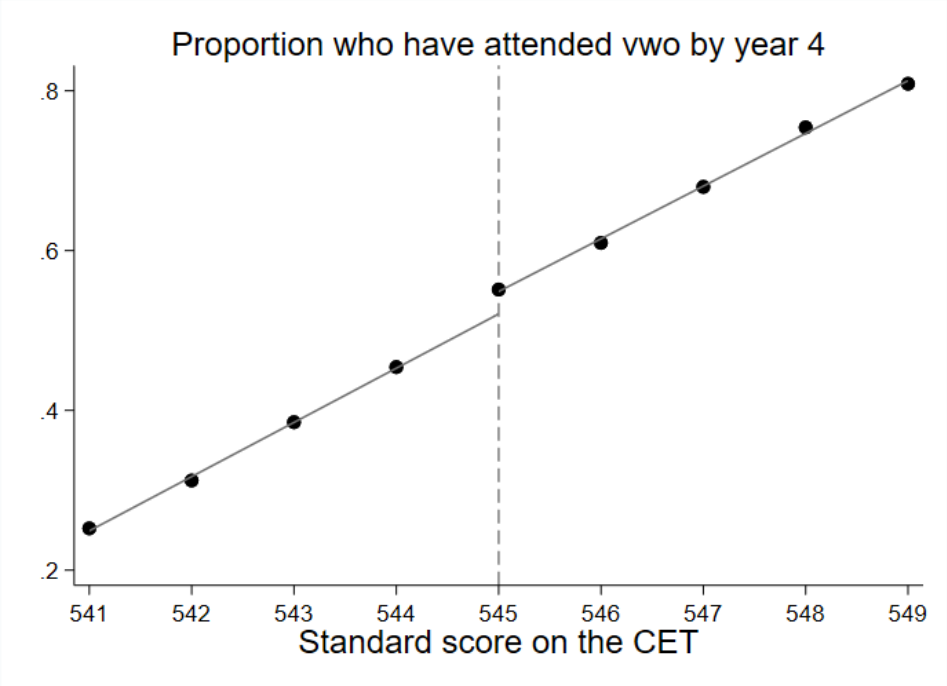

Figure 9. The reduced form for vwo attendance in year 4

Notes. This figure depicts the proportion of students who attended the vwo senior track by year 4 of secondary school for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2016/2017 school years per CET score. The outcome variable is measured on October 1st, during the students' fourth school year. The bandwidth used is 541-549 with the threshold residing at 545

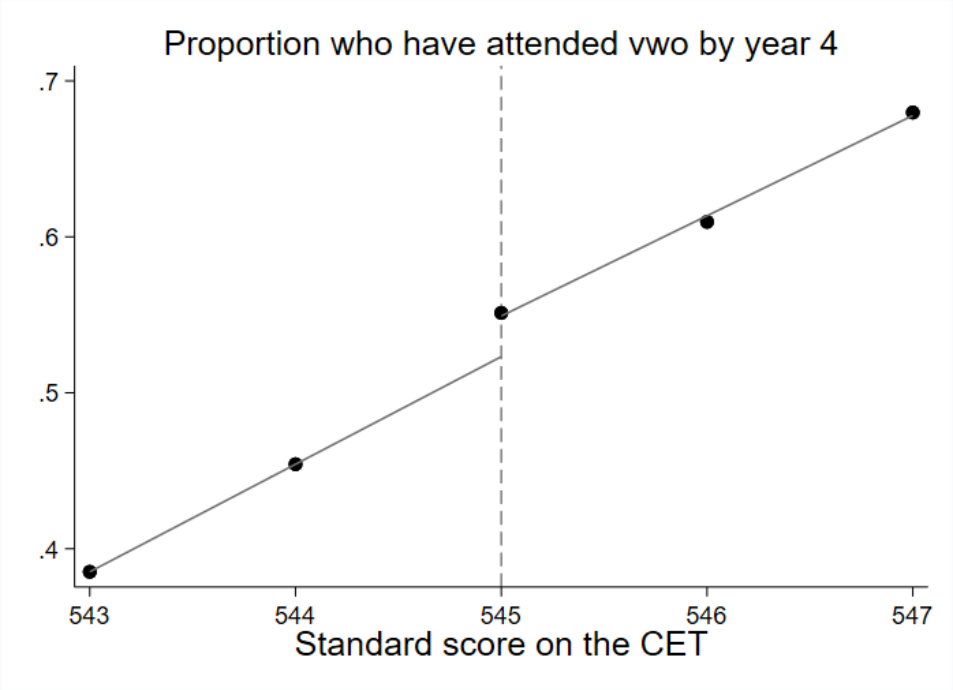

Figure 10. The reduced form for vwo attendance in year 4 at a smaller bandwidth

Notes. This figure depicts the proportion of students who attended the vwo senior track by year 4 of secondary school for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2016/2017 school years per CET score. The outcome variable is measured on October 1st, during the students' fourth school year. The bandwidth used is 541-549 with the threshold residing at 545

Table 6: The reduced form for grade retainment

	(1) **Repeated a grade by year 5**	**(2)** **Repeated a grade by year 6**	**(3)** **Repeated a grade by year 6**
Threshold	0.029***	0.030***	-0.005
	(0.007)	(0.007)	(0.004)
Test Score	-0.008**	-0.009**	-0.005***
	(0.004)	(0.004)	(0.001)
Threshold*Test Score	0.005	0.005	-0.007***
	(0.004)	(0.004)	(0.002)
Constant	0.147***	0.164***	0.160***
	(0.006)	(0.006)	(0.003)
Number of observations	91,106	91,106	148,851
Bandwidth	535-539	535-539	541-549

Notes. This table depicts the reduced form regressions for two different thresholds. The sample consists of students who took the Centrale Eindtoets (CET) between the 2014/2015 and 2017/2018 schoolyears. The outcome variables are dummies that equal 1 if a student has repeated a grade by the time they start the year specified in the column. Track promotions that delay secondary school completion do not count as repeating a grade. Threshold is a dummy that equals 1 if a student's test score was at least the threshold value needed to qualify for the track

recommendation category on the CET. Test Score reflects the CET score which ranges from 501-550. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6 shows significant estimates for the effect of crossing the vmbo-gl/tl/havo to havo threshold on the chance to repeat a grade by year 4 and 5. Remarkably, the estimate for crossing the havo/vwo to vwo threshold is not significant. Figures 11 through 14 show the reduced form visually. When looking at Figure 12, which is the same as Figure 11 but for a wider range of CET scores a parabolic pattern emerges in the likelihood to repeat a grade that peaks at the CET score that corresponds to the threshold for the havo track. The appendix includes reduced form figures for wider ranges across the board.

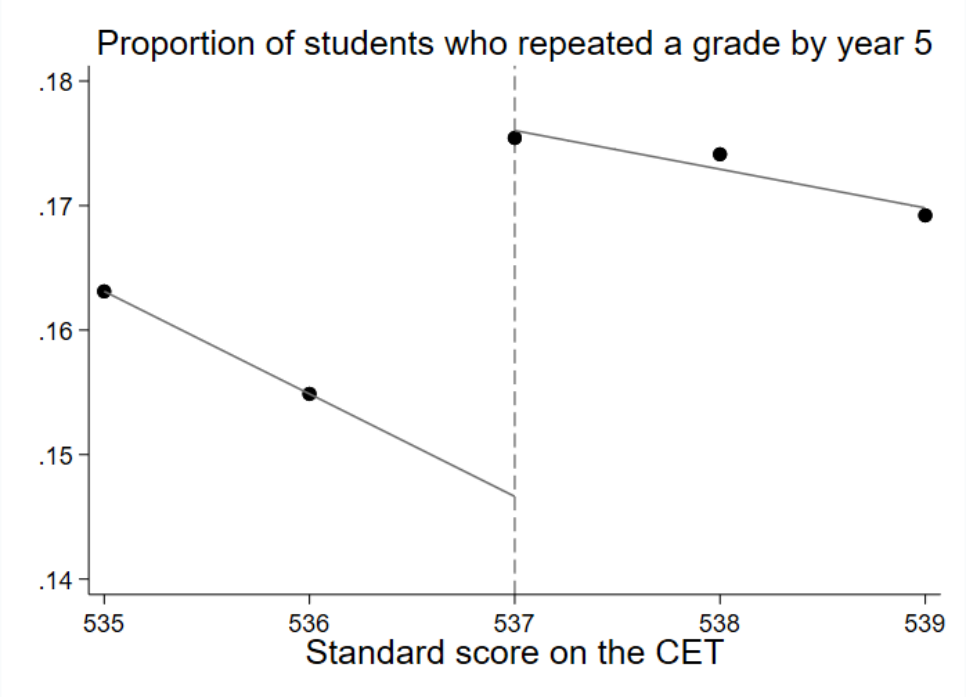

Figure 11. The reduced form for repeating grades by year 5

Notes. This figure depicts the proportion of students who repeated a grade by year 5 for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2017/2018 school years per CET score. The outcome variable is measured on October 1st, during the students' fifth school year. If cohorts did not have all the years of data available the available years were used and it was assumed they did not repeat a grade during the missing years. Track promotions that delayed secondary school completions do not count as repeating a grade. The bandwidth used is 535-539 with the threshold residing at 537. The threshold represents the minimum value where a CET score corresponds to the havo track.

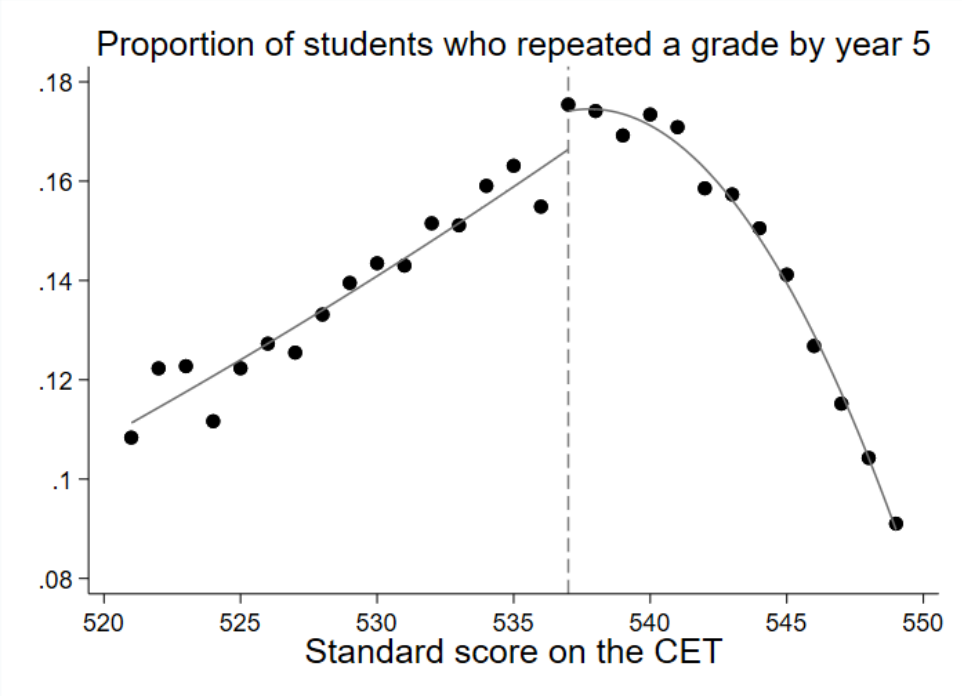

Figure 12. The reduced form for repeating grades by year 5 across a wider range

Notes. This figure depicts the proportion of students who repeated a grade by year 5 for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2017/2018 school years per CET score. The outcome variable is measured on October 1st, during the students' fifth school year. If cohorts did not have all the years of data available the available years were used and it was assumed they did not repeat a grade during the missing years. Track promotions that delayed secondary school completions do not count as repeating a grade. The bandwidth used is 521-549 with the threshold residing at 537. The threshold represents the minimum value where a CET score corresponds to the havo track.

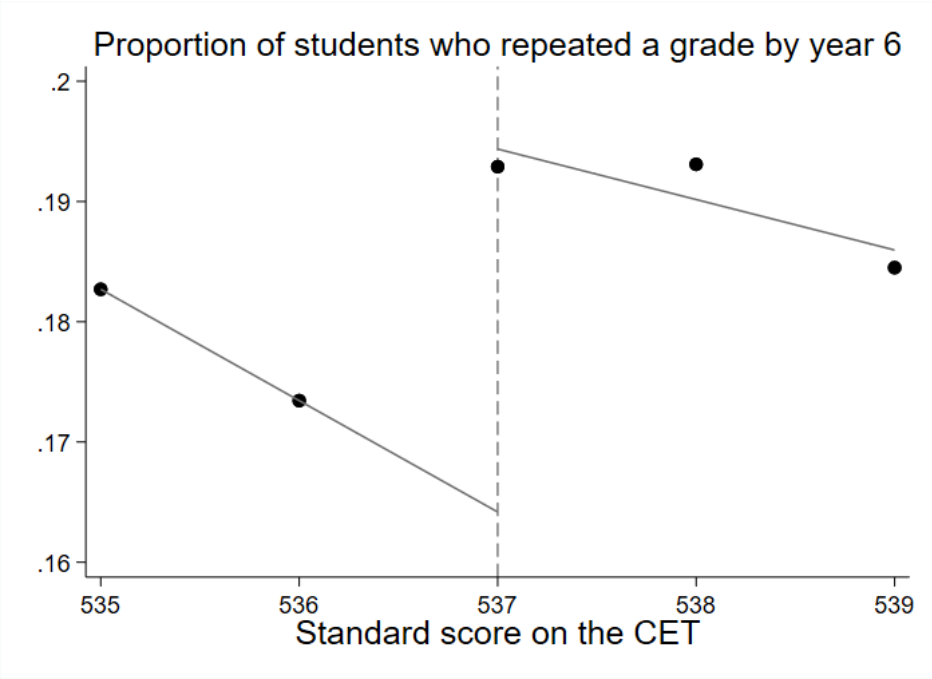

Figure 13. The reduced form for repeating grades by year 6 for the havo threshold

Notes. This figure corresponds to the second column in Table 6 and depicts the proportion of students who repeated a grade by year 6 for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2017/2018 school years per CET score. The outcome variable is measured on October 1st, during the students' sixth school year. If cohorts did not have all the years of data available the available years were used and it was assumed they did not repeat a grade during the missing years. Track promotions that delayed secondary school completions do not count as repeating a grade. The bandwidth used is 535-539 with the threshold residing at 537. The threshold represents the minimum value where a CET score corresponds to the havo track.

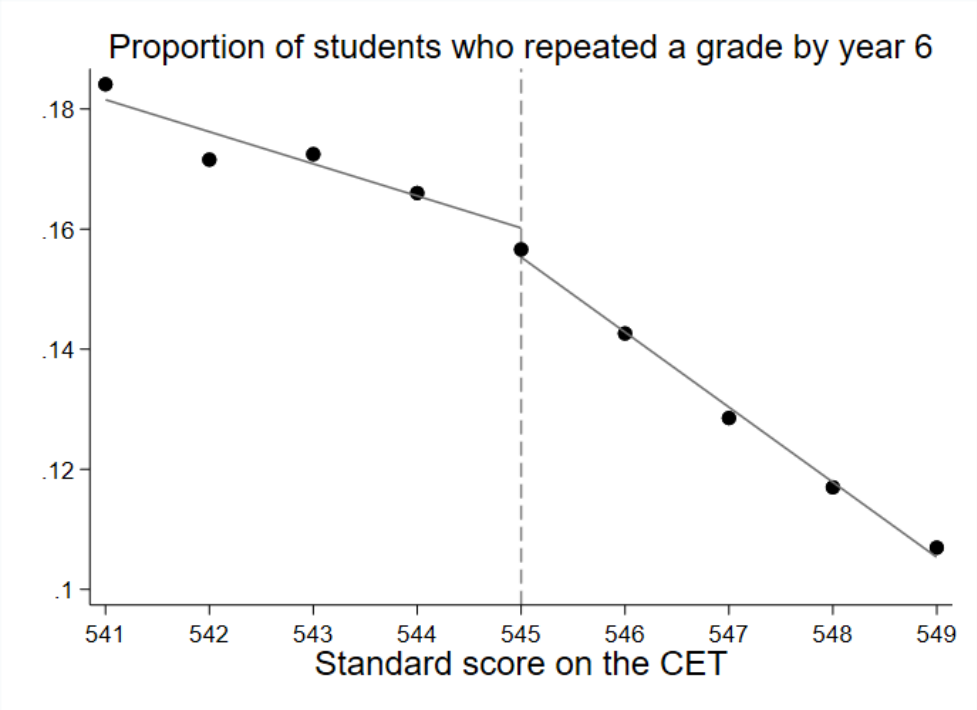

Figure 14. The reduced form for repeating grades by year 6 for the vwo threshold

Notes. This figure corresponds to the third column in Table 6 and depicts the proportion of students who repeated a grade by year 6 for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2017/2018 school years per CET score. The outcome variable is measured on October 1st, during the students' sixth school year. If cohorts did not have all the years of data available the available years were used and it was assumed they did not repeat a grade during the missing years. Track promotions that delayed secondary school completions do not count as repeating a grade. The bandwidth used is 541-549 with the threshold residing at 545. The threshold represents the minimum value where a CET score corresponds to the vwo track.

7.1 The causal effect of Track recommendations on Track attendance and the chance to repeat a grade.

Table 7 and 8 show my IV estimates for the causal effect of specific track recommendations on Track attendance by a specific year of secondary school and the chance to repeat a grade by a specific year of secondary school respectively. 2SLS and heteroskedasticity-robust standard errors are used. Table 7 excludes cohort 2018 in the sample for 3 out of the 4 columns due to

these cohorts being too young at this point in time. Overall, the estimated effects in Table 7 seem large and statistically significant across the board. Estimated effects are of course Local Average Treatment Effects (LATE) and should only be interpreted for the bandwidth chosen. The effects being LATE indicates that the estimated effects only apply to compliers, those that received their track recommendation because of their CET score and would have entered secondary school with a lower track recommendation if they had scored below the threshold.

Table 7: The effect of Track Recommendations on Track attendance

	(1) Havo/vwo by year 3	(2) Havo by year 4	(3) Vwo by year 4	(4) Vwo by year 4
Track recommendation	0.573***	0.581***	0.395***	0.510***
	(0.123)	(0.162)	(0.078)	(0.184)
Test Score	0.018	0.008	0.037***	0.026
	(0.012)	(0.015)	(0.007)	(0.020)
Track recommendation*Test Score	-0.006	0.010**	-0.004**	-0.010*
	(0.004)	(0.005)	(0.002)	(0.005)
Constant	0.247***	0.138	0.329***	0.271***
	(0.067)	(0.088)	(0.041)	(0.099)
Number of observations	91,106	71,542	117,856	66,111
Bandwidth	535-539	535-539	541-549	543-547

Notes. This table depicts IV estimates of the effect of receiving a havo recommendation for the first three columns and a vwo recommendation for the last two columns on the chance to attend the track specified in the column. The sample consists of students who took the Centrale Eindtoets (CET) between the 2014/2015 and 2017/2018 schoolyears. Column (2), (4) and (5) exclude the 2018 cohort. Column (3) excludes the 2017 and 2018 cohorts. The outcome variables are dummies that equal 1 if a student has attended least at the track specified in the column by the time they start the year specified in the column. Threshold is a dummy that equals 1 if a student's test score was at least the threshold value needed to qualify for the track recommendation category on the CET. Test Score reflects the CET score which ranges from 501-550. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The interpretation for the LATE coefficient in column (1) is as follows. On average, receiving a havo recommendation raises one's chance to have attended at least the havo/vwo track by the third year of secondary school by 57 percentage points in the sample. As discussed earlier column (1) gives less room for shifts across tracks for people as those that enter secondary school at havo/vwo may drop down in tracks but this won't be reflected in the variable in question. Interpretations for columns (2) and (3) are similar and seem to confirm the idea that track recommendations are a leading cause of track attendance, even multiple years after the recommendation has taken place. Flexibility for shifting what tracks one attends without

delaying secondary school by one or more years seems limited. Column (4) is the same as (3) but uses a smaller bandwidth and estimates a stronger effect at the cost of raising the standard errors substantially. Overall, there seems to be a strong, positive and significant effect of track recommendation on track attendance towards the later stages of secondary school.

Table 8: The effect of Track recommendations on the chance to repeat a grade

	(1) Repeated a grade by year 5	(2) Repeated a grade by year 6	(3) Repeated a grade by year 6
Threshold	0.480***	0.493***	-0.064
	(0.120)	(0.127)	(0.057)
Test Score	-0.039***	-0.041***	-0.000
	(0.012)	(0.012)	(0.006)
Threshold*Test Score	-0.002	-0.002	-0.007***
	(0.004)	(0.004)	(0.002)
Constant	-0.090	-0.079	0.192***
	(0.066)	(0.069)	(0.031)
Number of observations	91,106	91,106	148,851
Bandwidth	535-539	535-539	541-549

Notes. This table depicts IV estimates two different thresholds. Column (1) and (2) estimate the effect of receiving a havo track recommendation and use the threshold belonging to the havo track. Column (3) estimates the effect of receiving a vwo track recommendation and uses the threshold belonging to the vwo track. The sample consists of students who took the Centrale Eindtoets (CET) between the 2014/2015 and 2017/2018 schoolyears. The outcome variables are dummies that equal 1 if a student has repeated a grade by the time they start the year specified in the column. Track promotions that delay secondary school completion do not count as repeating a grade. Threshold is a dummy that equals 1 if a student's test score was at least the threshold value needed to qualify for the track recommendation category on the CET. Test Score reflects the CET score which ranges from 501-550. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8 estimates the causal effect of receiving specific track recommendations for compliers around the threshold on the chance to repeat a grade during secondary school. Column (1) and (2) use the havo track recommendation threshold but column 2 estimates the effect by a student's sixth year in secondary school compared to column (1)'s fifth year. This means that column (2) uses an additional year of data for the 2015 cohort but assumes that those who have not yet repeated a grade by then will not do so in this additional year for the other cohorts. The estimates between the two columns do not differ much, possibly due to this exact reason. The two estimated effects that exploit the havo track recommendation threshold both indicate a

significant positive effect of receiving a havo track recommendation on the likelihood to experience repeating a grade. The estimated effect that exploits the vwo track recommendation thresholds finds no significant effects of receiving a vwo recommendation on the chance to repeat a grade. Intuitively, barely making it into a track would also likely increase one's chance to repeat a grade as shown in column (1) and (2). But this isn't the case for this later estimated effect. This might be because the track recommendation below the vwo recommendation is havo/vwo, a mixed track recommendation. Students at the peak of havo/vwo are likely to be able to handle vwo essentially the same as those at the bottom of vwo. That I find insignificant effects here but find significant effects in Table 7, columns (4) and (5) might indicate that havo and vwo students get treated relatively similarly regarding grade retention policies. Compared to the disparity found in the comparison of the havo track recommendation with its lower counterpart the idea comes to mind that students who are not very different experience a significantly different chance to repeat a grade. This might further indicate that there is limited possibility for students at the peak of the track below havo to move up tracks without the use of the promotion vehicle, which usually delays their educational career by a year.

8 Discussion

Here I discuss the implications of the results presented above, the external and internal validity of the results and discuss limitations. First, the significant positive effects found of receiving a track recommendation on the odds to attend at least that track during secondary school imply that for the marginal student, who has the greatest chance of being misallocated there is not enough flexibility during secondary school to shift tracks to the 'right' one without repeating a grade. If the only problem for students to sort themselves efficiently is this flexibility between tracks, then it seems quite relevant for policymakers to re-examine how promoting tracks works in The Netherlands. The current system which causes delays in secondary school completion may be quite inefficient and causes major additional costs in the Dutch educational system. However, if a major cause of the track attendance differences lies in early secondary school experiences where those with a higher track recommendation get treated differently through instruments such as receiving better teaching, doing harder coursework and experiencing more positive peer effects this may indicate that in The Netherlands not enough children receive a proper opportunity to attend the higher tracks. Mixed track classes do exist in the earlier years of secondary school but the significant positive effects found above indicate they are not enough to compensate for the rift caused by a difference in track recommendations. At the same time the results in Table 8 indicate that perhaps not enough students are shifted down

from the havo track when they need to repeat a grade to complete the track. This seems especially true when considering that this study finds effects for compliers, those that needed a specific CET performance to increase their track recommendation. It is this group of people that is most likely to be overestimated, which is reflected in the results on the odds to repeat a grade. Policy decisions based on the results should factor in the cost of repeating a grade and if it is worth it for students to 'barely' complete the havo track if it comes with the price of grade retention. Likewise, implementing additional ways to decrease primary teacher track recommendation bias are likely worth it due to the current low flexibility and increased costs due to this in secondary education. Overall, the results seem to imply that track recommendations have a strong effect on the secondary school experience.

The internal validity of this paper is quite high due to the regression discontinuity design used as discussed in chapter 4. The external validity is less strong. The effects found only apply to those who needed a sufficient CET score to increase their track recommendation and scored close to the thresholds. This already excludes a relatively large portion of the student population. Furthermore, it is unlikely that these results are easily generalizable to other countries with other tracking systems as the Dutch systems is relatively unique. Germany has quite a comparable tracking system but track attendance there also likely is a major determinant of track choice by the student and parents. In The Netherlands track attendance is more based on track recommendations and acceptance by secondary schools. Furthermore, not every primary school student in The Netherlands takes the CET, it is known that schools that perform worse on the CET are more likely to start employing other final primary school tests. Which adds another limitation to the external validity to my results.

A limitation of this study is that I do not have data of the entire secondary school experience and how the sample changes directly post-secondary school as the population studied is simply too young. Dustmann et al., (2017) found similar results in Germany compared to mine but found that differences in outcomes in secondary school often vanished afterwards. I cannot check for such a possibility. There is also a possibility that Covid-19 affected some tracks more than others, which could potentially skew the results. Further research may exploit the valid first stages found here to estimate other, potentially long-run effects or may re-examine my results in a few years once the cohorts have completed secondary school.

9 Conclusion

In conclusion, I find significant positive effects of receiving a havo or vwo track recommendations on the chance to attend that track towards the later stages of secondary school and the odds to repeat a grade. These effects are LATE and apply around the thresholds. The exception is the threshold dividing the havo/vwo and vwo recommendations where no significant effects on the odds to repeat a grade is found. The results suggest that the marginal students are impacted significantly by their track recommendations and cast doubt on the efficiency of the Dutch education system during secondary school. Misallocations of students to the 'wrong' lower track can easily cause delays in track completions due to the way promoting tracks works, which often causes delays of a year. Simultaneously, the students who were overestimated seem to not be shifted tracks downwards enough when the goal is for these students to complete their track without repeating a grade. These implications are the strongest for the threshold between the highest vmbo track and the havo track recommendations.

10 References

Centraal Bureau Statistiek. (2022). *Vo; leerlingen, onderwijssoort in detail, leerjaar.* Retrieved from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80040ned/table?dl=6400F

Centraal Plan Bureau , (2019). *De waarde van eindtoetsen in het primair onderwijs.* Retrieved from:https://www.cpb.nl/sites/default/files/omnidownload/CPB-policy-brief-2019-03-de-waarde-van-eindtoetsen_0.pdf

College voor Toetsen en Examens. (2015) Terugblik 2015; Resultaten Centrale Eindtoets 2015. Retrieved from: https://www.centraleeindtoetspo.nl/publicaties/publicaties/2016/09/09/terugblik-centrale-eindtoets-2015

College voor Toetsen en Examens. (2016) Terugblik 2016; Resultaten Centrale Eindtoets 2016. Retrieved from: https://www.centraleeindtoetspo.nl/actueel/nieuws/2017/04/04/terugblik-centrale-eindtoets-2016

College voor Toetsen en Examens. (2017) Terugblik 2017; Resultaten Centrale Eindtoets 2017. Retrieved from: https://www.centraleeindtoetspo.nl/actueel/nieuws/2018/02/12/terugblik-centrale-eindtoets-2017

College voor Toetsen en Examens. (2018) Terugblik 2018; Resultaten Centrale Eindtoets 2018. Retrieved from: https://www.centraleeindtoetspo.nl/publicaties/publicaties/2018/12/12/terugblik-centrale-eindtoets-2018

Borghans, L., Diris, R., Smits, W., & de Vries, J. (2019). The long-run effects of secondary school track assignment. *PloS One*, *14*(10), e0215493.

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, *22*(52), 782–861.

Dijks, M. A., Warrens, M. J., Fleur, E., Korpershoek, H., Wichgers, I. J. M., & Bosker, R. J. (2020). The predictive power of track recommendations in Dutch secondary education. *Pedagogische Studien*, *97*(4), 263–280.

Geven, S., Batruch, A., & van de Werfhorst, H. (2018). Inequality in teacher judgements, expectations and track recommendations: A review study. *Amsterdam: Universiteit van Amsterdam*.

Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013). Beyond judgment bias: How students' ethnicity and academic profile consistency influence teachers' tracking judgments. *Social Psychology of Education*, *16*(4), 555–573.

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

Hanushek, E. A., & Wößmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences- in-Differences Evidence Across Countries. *The Economic Journal*, *116*(510), C63–C76. https://doi.org/10.1111/j.1468-0297.2006.01076.x

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355.

Malamud, O., & Pop-Eleches, C. (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics*, *95*(11), 1538–1549. https://doi.org/10.1016/j.jpubeco.2011.03.006

Nürnberger, M., Nerb, J., Schmitz, F., Keller, J., & Sütterlin, S. (2016). Implicit gender stereotypes and essentialist beliefs predict preservice teachers' tracking recommendations. *The Journal of Experimental Education*, *84*(1), 152–174

Rijksoverheid. (2022). *Schooladvies toelating voortgezet onderwijs*. Retrieved from: https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-eindtoets-basisschool/toelating-voortgezet-onderwijs-gebaseerd-op-definitief-schooladvies

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, *45*(1), 523–538.

Timmermans, A. C., Kuyper, H., & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do D utch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, *85*(4), 459–478.

Van Leest, A., Hornstra, L., van Tartwijk, J., & van de Pol, J. (2021). Test-or judgement-based school track recommendations: Equal opportunities for students with different socio-economic backgrounds? *British Journal of Educational Psychology*, *91*(1), 193–216.

Vuuren, D., Wiel, K. (2015). Zittenblijven in het primair en voortgezet ondewijs. Een inventarisatie van de voor- en nadelen. *CPB Policy brief, Retrieved from: https://www.cpb.nl/sites/default/files/publicaties/download/cpb-policy-brief-2015-01-zittenblijven-het-primair-en-voortgezet-onderwijs.pdf*.

8 Appendix

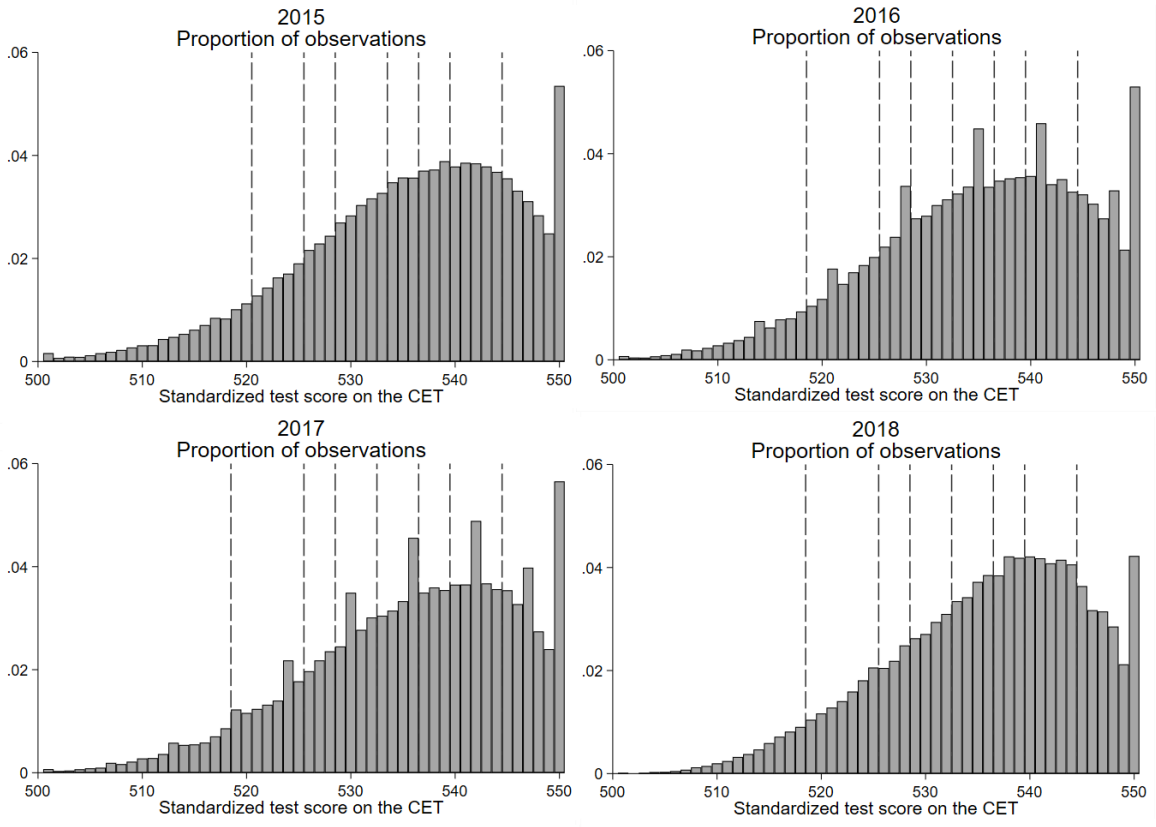

Figure A1. Distribution of CET scores for individual cohorts

Notes. This graph depicts the distribution of CET scores for each year in the sample. 2015 refers to the 2014/2015 school year. CET scores range from 501-550 and scores correspond to specific secondary school track recommendations. The thresholds between track recommendations are marked by dashed vertical lines before scores: 519, 526, 529, 533, 537, 540 and 545. During the 2014/2015 school year the threshold found at 519 was 521 and the 533 threshold was 534.

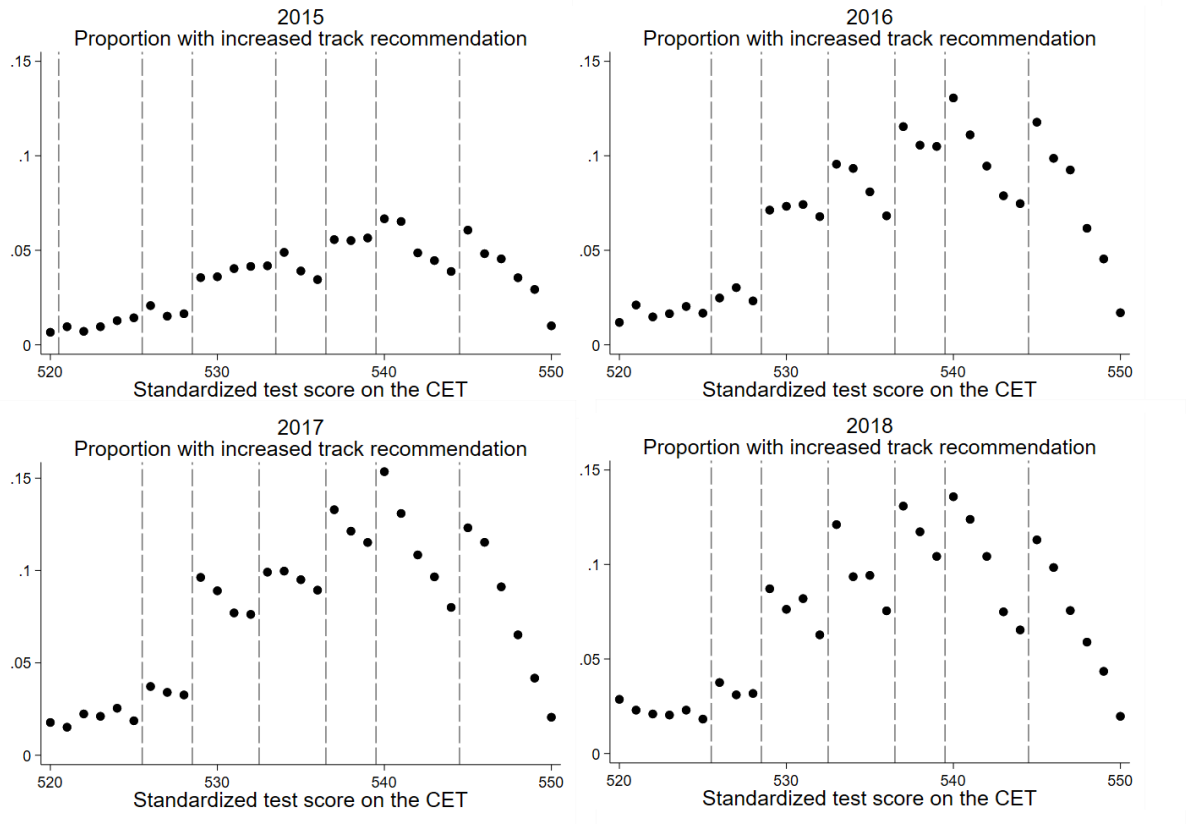

Figure A2. Increases in track recommendation for individual cohorts

Notes. This graph depicts the proportion of students per CET score that experienced a track recommendation increase for the individual cohorts in the sample. 2015 refers to the 2014/2015 school year. A student increases their track recommendation if their final track recommendation after the CET is higher than their initial track recommendation. The thresholds between track recommendations are marked by dashed vertical lines before scores: 519, 526, 529, 533, 537, 540 and 545. During the 2014/2015 school year the threshold found at 519 was 521 and the 533 threshold was 534.

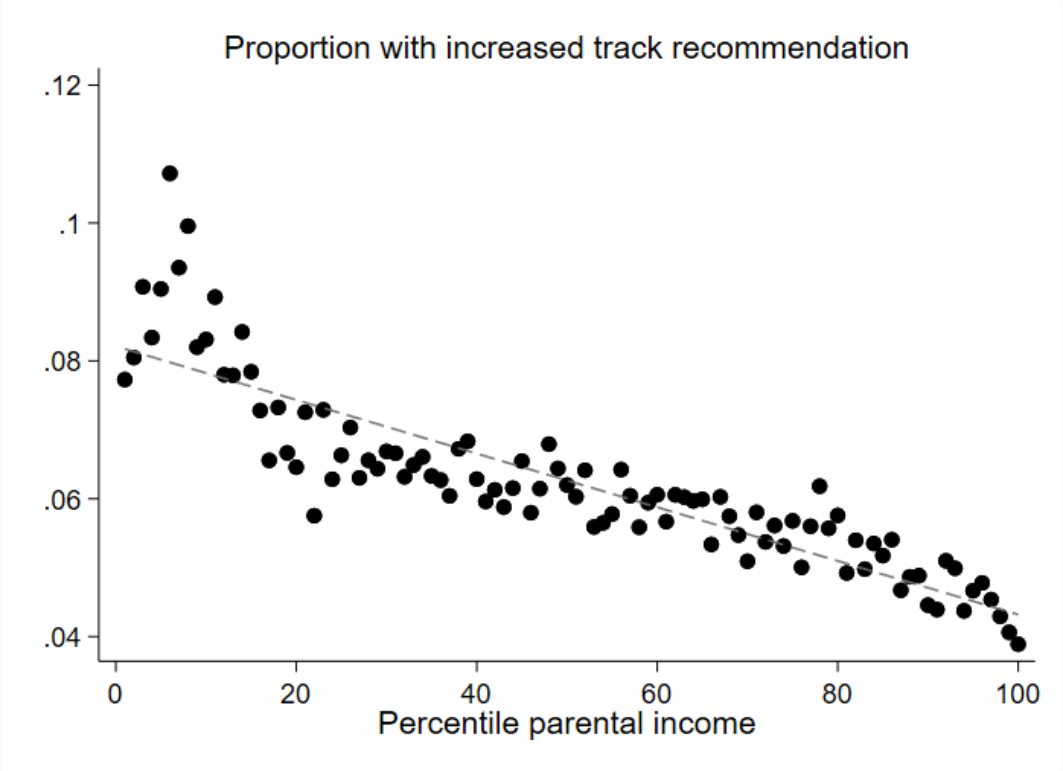

Figure A3. Proportion.

Notes. This figure depicts the proportion of each percentile in parental income that increased their track recommendation after taking the Centrale Eindtoets (CET). The sample consists of students who took the CET between the 2014/2015 and 2017/2018 school years. Percentile parental income is the combined annual gross income of the legal father and mother measured on January 1st on the year that the child takes the CET.

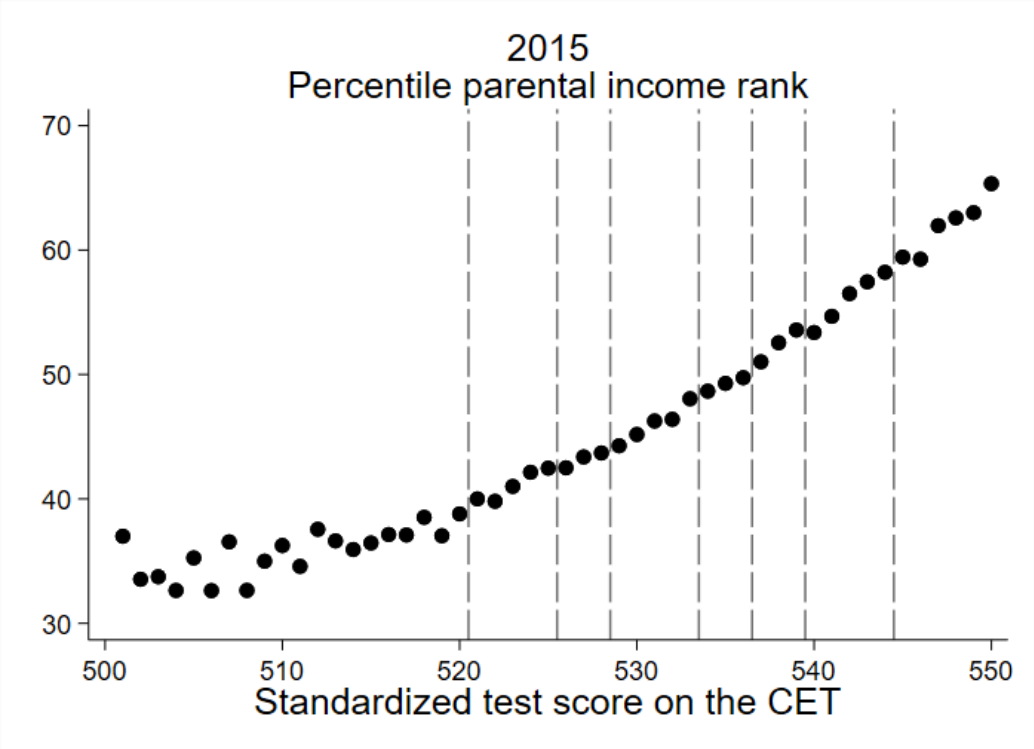

Figure A4. Relative parental income for test takers in 2015.

Notes. This graph depicts the average percentile parental income of students per CET score. Percentile parental income is gross income of the father and mother which is then split into 100 percentiles relative to all other parental incomes during the year the child took the CET. The graph consists of data from students who took the CET during the 2014/2015 school year. The thresholds between track recommendations are marked by dashed vertical lines before CET scores: 521, 526, 529, 534, 537, 540 and 545.

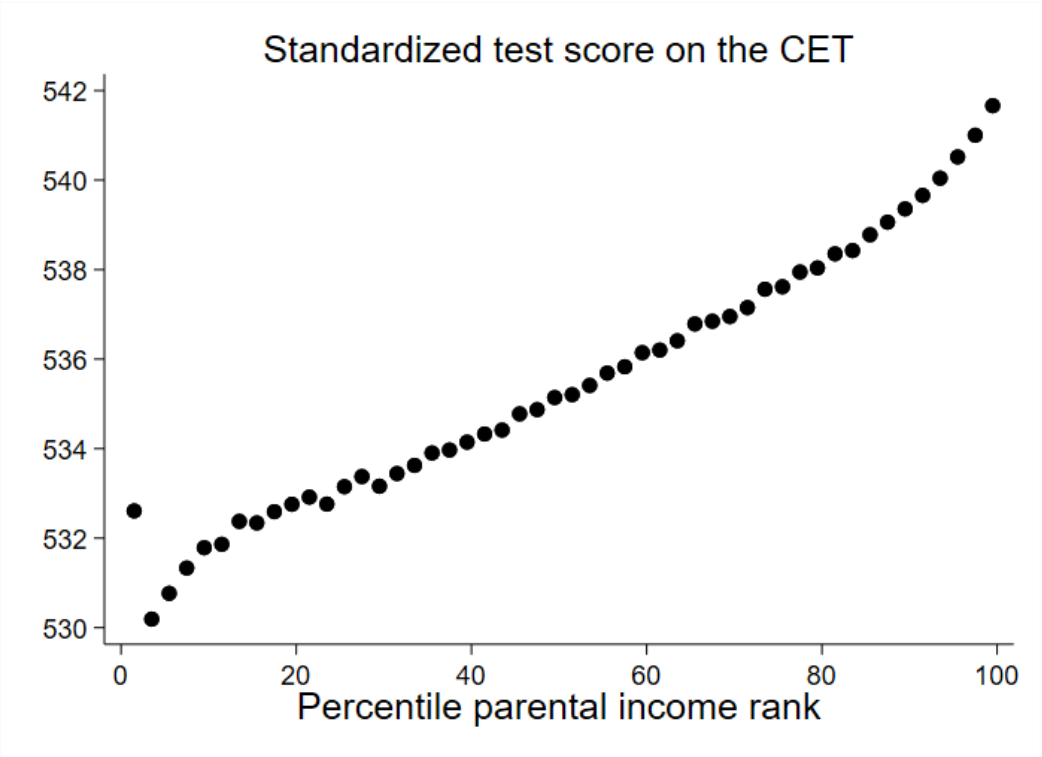

Figure A5. Average CET score per percentile parental income

Notes. This graph depicts the average Centrale Eindtoets (CET) score per percentile parental income of the students that took the CET between the 2014/2015 school year and the 2017/2018 school year. Percentile parental income is gross income of the father and mother which is then split into 100 percentiles relative to all other parental incomes during the year the child took the CET. CET scores range from 501-550.

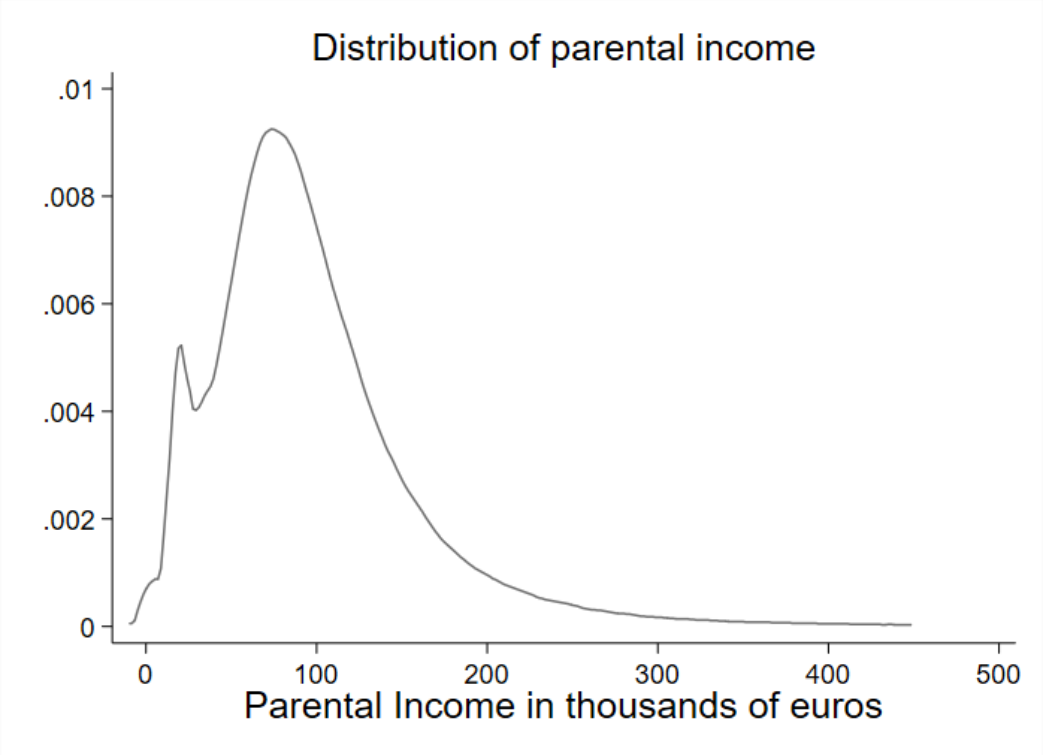

Figure A6. Distribution of parental income

Notes. This graph depicts the distribution of parental income. Parental income is the combined gross income of the legal father and mother during the year their child took the Centrale Eindtoets (CET). The sample consists of students who took the CET between the 2014/2015 and 2017/2018 school years.

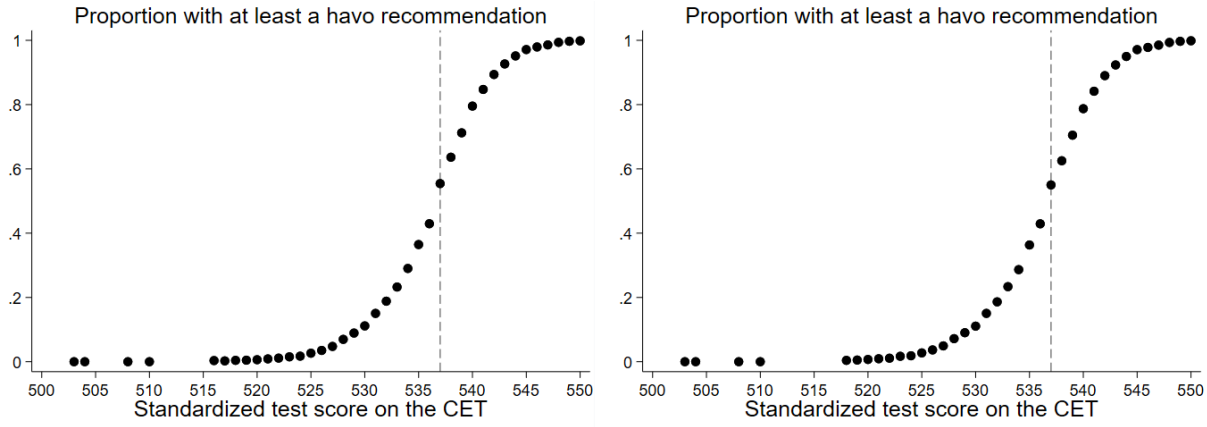

Figure A7. Expanded first stages for the havo threshold

Notes. This figure depicts the proportion of students who received at least a havo recommendation as their final track recommendation per Centrale Eindtoets (CET) score. The left figure uses the 2014/2015 – 2017/2018 school years while the right figure uses the 2014/2015 – 2016/2017 school years. Students may receive this track recommendation before the CET or after. The score 537 is marked with a dashed line due to this being the lowest value that is associated with the havo track for the CET. Missing values are excluded due to privacy reasons relating to low sample sizes.

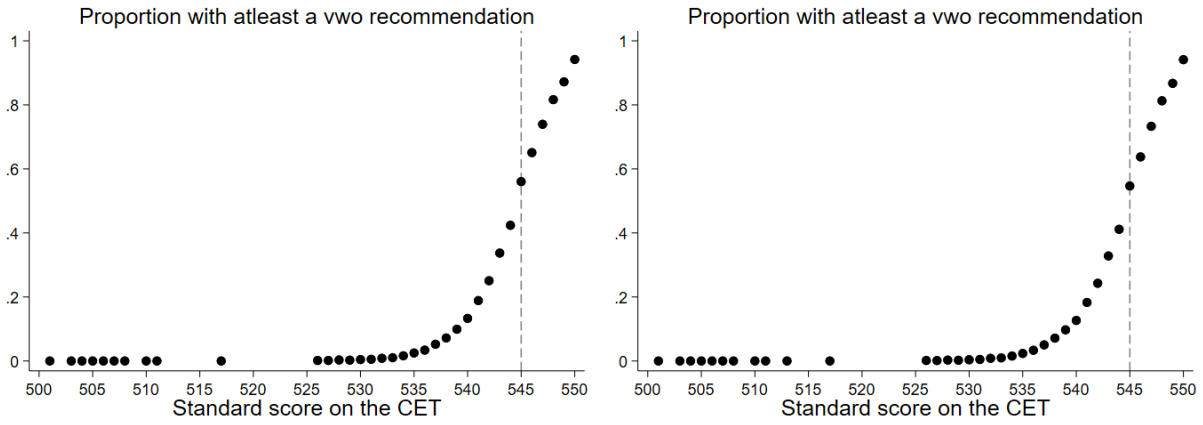

Figure A8. Expanded first stages for the vwo threshold

Notes. This figure depicts the proportion of students who received a vwo recommendation as their final track recommendation per Centrale Eindtoets (CET) score. The left figure uses the 2014/2015 – 2017/2018 school years while the right figure uses the 2014/2015 – 2016/2017 school years. Students may receive this track recommendation before the CET or after. The score 545 is marked with a dashed line due to this being the lowest value that is associated with the vwo track for the CET. Missing values are excluded due to privacy reasons relating to low sample sizes.

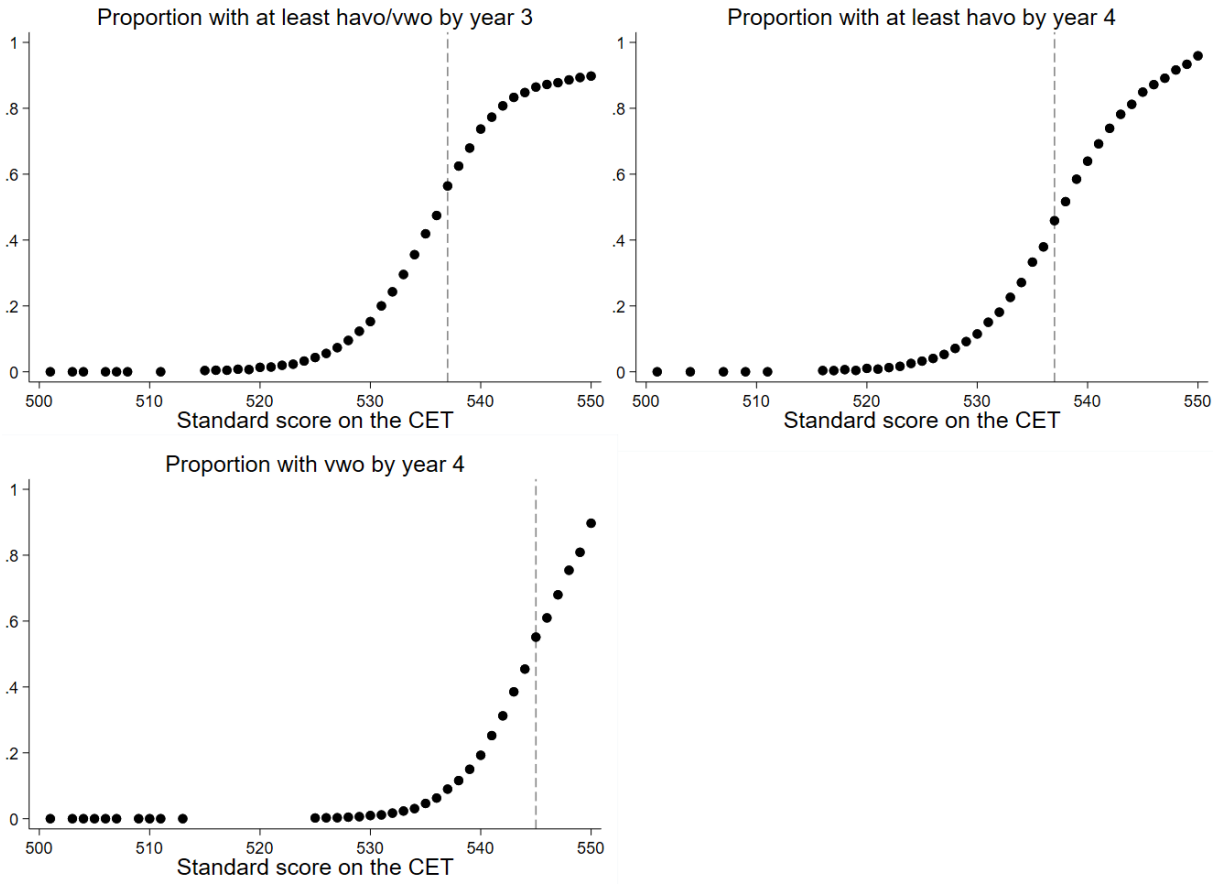

Figure A9. Expanded reduced forms for track attendance

Notes. This figure depicts the proportion of students who attended havo/vwo by year 3 (top left), senior havo (top right) by year 4 and senior vwo (bottom left) by year 4 of secondary school for those that took the Centrale

Eindtoets (CET). The top left graph uses the students who took the CET between the 2014/2015 – 2017/2018 school years The top right and bottom left graphs use students who took the CET between the 2014/2015 – 2016/2017 school years. The outcome variable is measured on October 1st, during the school year mentioned in the title. The threshold used for the top graphs is 537 and corresponds with the havo track and 545 for the bottom graph, which corresponds with the vwo track. Missing values are excluded due to privacy reasons relating to low sample sizes.

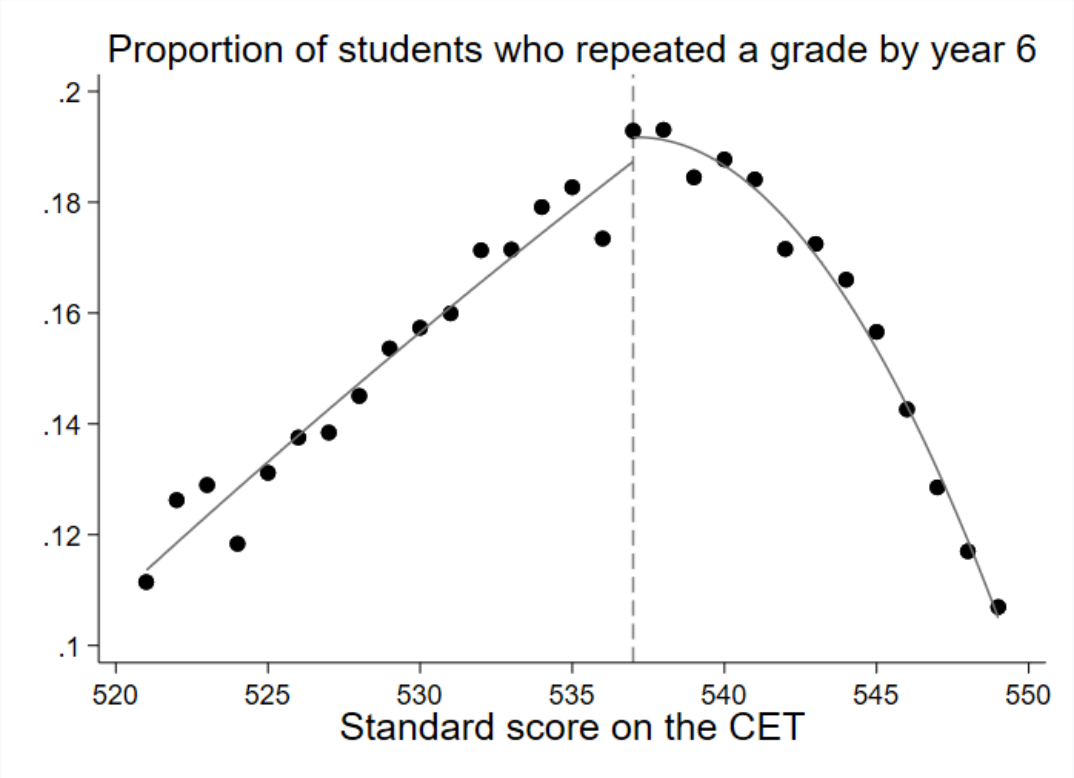

Figure A10. Expanded reduced form for grade retention.

Notes. This figure depicts the proportion of students who repeated a grade by year 6 for those that took the Centrale Eindtoets (CET) between the 2014/2015 until the 2017/2018 school years per CET score. The outcome variable is measured on October 1st, during the students' sixth school year. If cohorts did not have all the years of data available the available years were used and it was assumed they did not repeat a grade during the missing years. Track promotions that delayed secondary school completions do not count as repeating a grade. The bandwidth used is 521-549 with the threshold residing at 537. The threshold represents the minimum value where a CET score corresponds to the havo track.

Table A1. The first stage for the smaller samples

	(1) Havo	(2) Vwo
Threshold	0.054***	0.077***
	(0.010)	(0.002)
Test Score	0.066***	0.071***
	(0.006)	(0.006)
Threshold*Test Score	0.011*	0.006***
	(0.006)	(0.002)
Constant	0.495***	0.483***
	(0.009)	(0.005)
Number of observations	71.542	117.856
Bandwidth	535-539	541-549

Notes. This table depicts the first stage regressions for two different thresholds. These regressions correspond to Table 4, column 5 and 7 but use a smaller sample. The sample consists of students who took the Centrale Eindtoets (CET) between the 2014/2015 and 2016/2017 schoolyears. The outcome variables are dummies that equal 1 if a student receives a final track recommendation at least at the level specified in the column. Threshold is a dummy that equals 1 if a student's test score was at least the threshold value needed to qualify for the track recommendation category on the CET. Test Score reflects the CET score which ranges from 501-550. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2. Absolute distribution of Final track recommendations

Initial Track recommendation	*Bbl*	*Bbl-kl*	*Kl*	*Kl-gl*	*Kl-tl*	*Gl*	*Gl-tl*	*Gl-havo*	*Tl*	*Tl-havo*	*Havo*	*Havo/vwo*	*Vwo*	*TOTAL*
Bbl	29,858	711	1,032	17	47	<10	20	<10	66	<10	<10	<10	0	31,769
Bbl-kl	0	13,055	515	39	142	33	106	<10	205	23	<10	<10	<10	14,132
Kl	0	0	47,816	133	817	305	434	22	3,279	175	38	<10	<10	53,025
Kl-gl	0	0	0	1,780	<10	37	61	<10	76	14	<10	<10	<10	1,993
Kl-tl	0	0	0	0	9,876	20	86	23	1,278	189	40	18	<10	11,530
Gl	0	0	0	0	0	4,372	50	42	134	33	36	<10	<10	4,669
Gl-tl	0	0	0	0	0	0	7,917	73	104	145	220	<10	<10	8,470
Gl-havo	0	0	0	0	0	0	0	1,574	<10	<10	133	42	<10	1,759
Tl	0	0	0	0	0	0	0	0	79,960	3,672	3,225	152	14	87,023
Tl-havo	0	0	0	0	0	0	0	0	0	28,027	3,455	531	18	32,031
Havo	0	0	0	0	0	0	0	0	0	0	92,636	3,784	1,850	98,270
Havo/vwo	0	0	0	0	0	0	0	0	0	0	0	36,608	2,496	39,104
Vwo	0	0	0	0	0	0	0	0	0	0	0	0	102,631	102,631
Total	29,858	13,766	49,363	1,969	10,890	4,780	8,674	1,747	85,100	32,287	99,802	41,155	107,016	486,406

Notes. This table depicts frequencies of initial track recommendations and how they are distributed along final track recommendations for students that took the Centrale Eindtoets (CET) from the 2014/2015 school year until the 2017/2018 school year. Values below 10 have been suppressed and some total values are rounded to the nearest tenth digit for privacy purposes.