ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis Economics of Markets and Organisations

The Effect of Track Availability on Track Underestimation and Enrolment in Dutch Secondary Education.

Name student: Karzan Schippers
Student ID number: 548390

Supervisor: Sacha Kapoor
Second Supervisor: Dinand Webbink

Date final version: 24/08/2023

Abstract

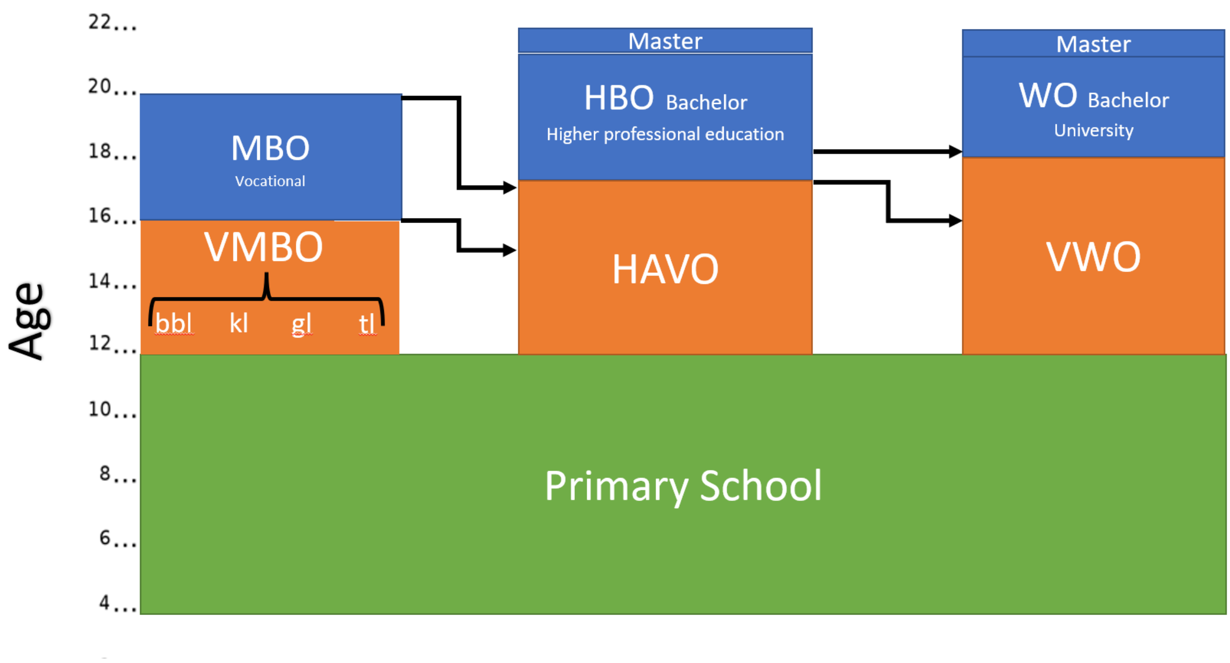
In the Netherlands, students finish primary education and are assigned a track at the end of sixth grade, based on teacher and test track recommendations. A lower final track recommendation than the test recommendation it is known as *track underestimation* and attending a lower track than the test recommendation is knowns as *underattendance*. A correlation has been established between the additional distance to the nearest secondary school offering the havo track and track underestimation, but there is no research exploring the causality between these concepts. This paper examines the causal effect of track availability on track underestimation and enrolment in the Netherlands. I estimate that a large increase in the additional distance to the nearest havo/vwo school increases the chance of track underestimation by 1.3 percentage points ($P < 0.01$) and find no evidence for a significant effect on underattendance. Different specifications and robustness checks are explored, which indicate a high chance of the results being biased.

Introduction

In the Netherlands, students finish primary education and are assigned a track at the end of sixth grade. Figure 1 depicts the main tracks in secondary education and how they relate to primary and tertiary education. Although students may be placed in mixed classes with multiple tracks present in the first year(s) of secondary education, Ree et al. (2023) and Schippers (2022) show that the track students end up in is primarily influenced by their final track recommendation at the end of primary education. This final track recommendation is influenced by an initial teacher track recommendation and a test track recommendation. If the test-based recommendation exceeds the initial recommendation, the primary school legally has to consider an upwards revision and must give reasoning if they do not proceed with such a revision (WPO, 2014).

Figure 1

The Dutch Education System

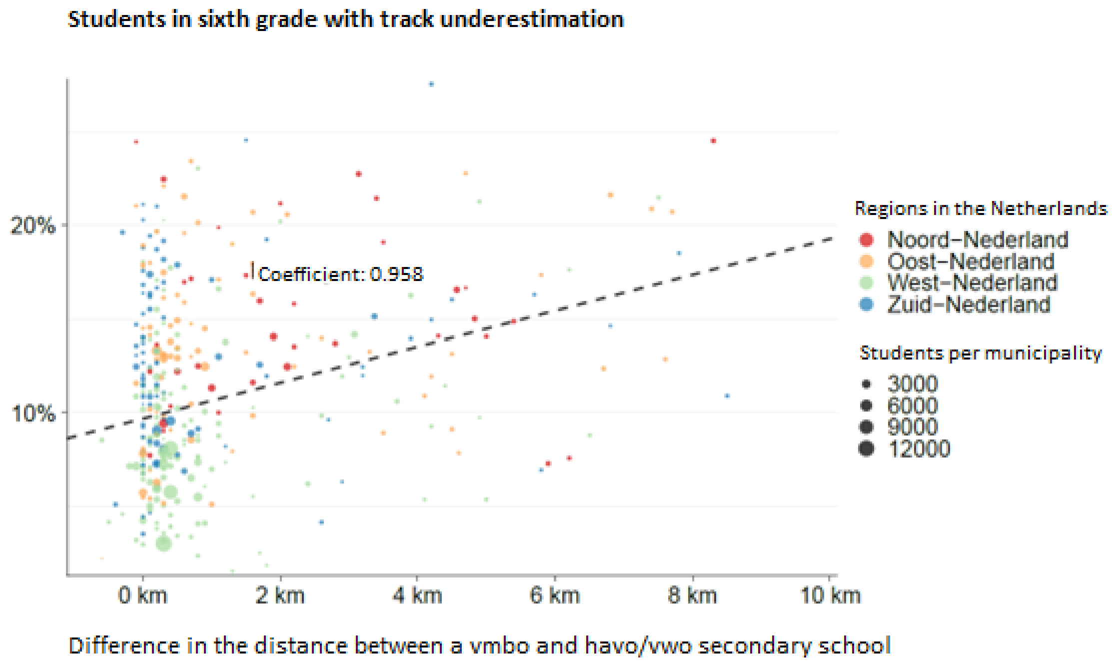


Note. The figure gives an overview of the Dutch education system and the expected ages of students at various points throughout. Once students graduate primary school, they enter secondary school and are sorted into separate tracks immediately or at some point in the next three years. These tracks are: vmbo, havo and vwo. Those who attend the vmbo track can be further subdivided into vmbo-bbl, vmbo-kl, vmbo-gl and vmbo-tl. The arrows indicate when students may generally switch tracks and the delay they incur. For example, one who graduates vmbo-tl at the end of the fourth year may move to havo, where they will take the fourth and fifth year of havo before graduating again.

Although mandated by law that the primary school has to consider revising a track recommendation upwards, the freedom granted to the supervising body to decline such a revision has led to a phenomenon the Dutch call *onderadvisering*, which translates to 'track underestimation'[1]. This occurs when the final track recommendation is lower than the track corresponding to their test score. Track underestimation is correlated with the additional distance students have to travel to the nearest secondary school that teaches the havo or vwo tracks as seen in Figure 2. Each additional kilometre that students have to traverse to reach the nearest havo/vwo school is associated with a 0.96 percentage point increase in the likelihood to receive a track underestimation (Lam et al., n.d).

Figure 2

The Correlation between track availability and track underestimation



Note. The figure pictures the percentage of students that receive a final track recommendation lower than the track recommendation corresponding to their test score per municipality on the y-axis and the difference in distance between a vmbo and havo/vwo secondary school.

Multiple explanations are possible for this correlation. First, some regions may be systematically overestimated by the standardized test due to some omitted variable which the teacher corrects for by revising recommendations upwards less often. Second, the culture in

[1] Although these terms may imply that one track recommendation is more valid than another, that is not the stance of this thesis. Teachers have additional information about students, which they may use as they see fit and because of this, not all track underestimation is something that needs to be 'fixed'.

these regions may find the havo and vwo tracks less important. This could for example manifest as municipalities and the people in them placing less importance in the presence of schools offering these tracks and teachers revising the track recommendation upwards less often due to the lowered perceived importance of the tracks. Third and last, the availability of schools in the region may be the cause.

Teachers may take into account the availability of secondary schools offering each track. Students may suffer academically due to sleeping less, socially by being placed in environments with less people they know and physically by having to travel larger distances at the relatively young ages of secondary school students (eleven and up). In extreme cases, students originating from the island of Terschelling must live with a foster family on the mainland throughout the week if they wish to attend secondary schools offering havo or vwo tracks. If teachers feel that this low availability of certain tracks is harmful to students, they may be less willing to revise track recommendations upwards when they would have done so if schools were simply placed closer.

In this paper, I explore if this relationship is causal. That is, what is the effect of track availability on track underestimation and underattendance? Underattendance is an adjacent phenomenon where students enrol into secondary school at a lower track than their final track recommendation. Administrative data from Statistics Netherlands (CBS) is used to gather information about individuals' track recommendations, their place of residence and parental characteristics regarding place of origin and education between 2009 and 2021. This is combined with public data from CBS on the average distance to the nearest vmbo and havo/vwo schools per neighbourhood in the Netherlands.

A difference-in-difference design with differential treatment timing is executed with using neighbourhood fixed effects instead of individual fixed effects as track availability data is available only at the neighbourhood level. I operationalize track availability as the change in the distance to the nearest havo/vwo school minus the distance to the nearest vmbo school compared to the previous year. If this change exceeds one kilometre, a neighbourhood becomes 'treated'. I further use two separate binary variables, one for each direction the distance variable may change. As alternative specifications, I use the underlying continuous distance variable as the treatment, change the timespan to 2015-2021 and change the threshold for the binary treatment variables to 0.5 kilometres. Finally, I perform a placebo test on the most significant results and examine the validity of the parallel trends assumption.

When using treatment dummies, I find significant positive effects on track underestimation for both increases and decreases in the additional distance to the nearest havo/vwo school, although not economically significant. These results were robust to adding time-varying controls and large positive changes were robust to adjusting the threshold for the treatment dummy from one kilometre to 0.5 kilometres. Large negative changes were not robust to the usage of the different threshold and remained significant when using a placebo test where no effects were expected to be found. Neither treatment dummy was robust to adjusting the sample years from 2009-2021 to 2015-2021. No significant effects were found for underattendance.

When using a continuous treatment, no significant effects were found for track underestimation but significant effects were found for the effect on underattendance, which were not economically significant. This contrast with treatment dummies indicates that underattendance may be influenced primarily by smaller changes of the treatment variable, not captured by the treatment dummies. This result was robust to adding time-varying controls and the estimates failed to remain significant in the placebo test. When examining pre-treatment dynamics to assess the likelihood of the parallel trends assumption holding, it was found that treatment and control neighbourhoods differed significantly both before and after treatment. This indicates that it is unlikely that treatment neighbourhoods would have had the same trends in track underestimation if they had not been treated. Harming the internal validity of the results.

Previous research on track availability, track underestimation and underattendance is sparse. Some research does exist which explores the relation between track recommendations and concepts like minority status and social economic status (Geven et al., 2018). To my knowledge, this is the first time for anyone to attempt to estimate a causal link between these concepts. I suspect that this is the case due to two main reasons. First, the number of countries where these concepts are applicable are limited. Second, track underestimation and underattendance are relatively new terms, only gaining popularity inside the Netherlands in the last few years. The hope is that this paper aids policy makers in making decisions regarding track recommendations and the opening and closings of schools.

Literature Review

In the Netherlands, students who graduate from primary education receive a track recommendation which determines the most academically challenging track they may enrol in at the start of secondary school. Students may decide to enrol in a less challenging track and although it is technically possible to shift tracks up- and downwards during secondary education and beyond, Ree et al. (2023) shows that the track students end up in is primarily influenced by the final track recommendation they receive in primary school. Track-specific diplomas are a primary determinant of possibilities in tertiary education and thus likely influence careers and lives in the long term. Evidence for this leading to increased wages has been mixed (Borghans et al., 2019; Dustmann et al., 2017; Oosterbeek et al., 2021), but the data used in these studies are often multiple decades old and are therefore not necessarily applicable to the current versions of their respective educational systems.

Track recommendations are decided by the primary school teacher and go through a three-step process. First, the teacher decides on an initial recommendation based on their experience with the student and previous test results. Second, students take one of multiple standardized tests which grants a test-based recommendation. Third, if the test-based recommendation exceeds the initial recommendation, the teacher may choose to increase the initial recommendation up to the test-based recommendation. This is also known as 'upgrading' the track recommendation and has become more common since an administrative change in 2015 which mandated teachers to consider such an upgrade (WVO, 2014). When teachers decide against such upgrades it is colloquially known as "*onderadvisering*", which I will call *"track underestimation"* from now on.

As Ree et al. (2023) shows, upgrading students' track recommendation generally increases the chance that students graduate at that higher track. It may also improve their later-in-life outcomes. At the same time, the marginal student being upgraded will likely face an increased chance to experience grade retention (Ree et al., 2023; Schippers, 2022), higher levels of school-related stress and have less time left over for leisure activities when being assigned to a more academically challenging track. Such difficulties may be exponentially higher when students have to travel further to reach a school that offers their new track due to a lacking track availability in their area. Additional travel time may require students to relocate, spend more time and/or money traveling and may place additional burdens on students and their families. An extreme example can be found in students who live in Terschelling, one of the Dutch islands, who must often find a foster-family on the mainland to temporarily take them in if they want to attend a havo or vwo track.

Previous research has shown that Dutch primary school teachers' track recommendations favour students with a higher socio-economic status (SES) and female students (Boone & Houtte, 2011; Geven et al., 2018; Leest et al., 2020; Timmermans et al., 2015). Geven et al, (2018) relates such 'bias' to the perception of an educationally supportive home, which would indicate that these teachers are using information about students' social environments and factors outside academic ability that will aid them in school to shape their track recommendations.

Thus, adjusting students' track recommendation to equal their test recommendation is a careful choice for primary school teachers. Students will likely benefit in the long-run due to them earning more academically challenging degrees and the subsequent consequences of those degrees. However, such upgrades come with downsides. Higher chances of grade retention and more challenging coursework are obvious examples. Furthermore, these downsides may be compounded by a lack of schools that teach the appropriate track in the area.

A similar choice has to be made by the students and their families themselves if such an upgrade is indeed granted. The same up- and downsides apply to this choice as the one made by their teachers. However, students and their families likely understand a students' exact situation better than their teacher does. At the same time, they may be less certain about what exactly is needed or expected to succeed at specific track levels and may therefore defer to a teacher's choice. Students who choose to forego the track recommendation would then end up attending a school which is located closer to them but only offers tracks that are less academically challenging. In the rest of this paper, this phenomenon will be labelled *"underattendance"*.

From an economic perspective, both teachers and students must make a choice in an attempt to maximize the student's utility. Teachers balance the potential up- and downsides from an increased track recommendation, knowing that a student's family may accept or decline an upgrade. Students and their families get to make their choice only once the teacher gives the green light, but may choose to defer to the teacher's judgement due to their experience if the family is unsure about how their situation affects their child's odds at different tracks. Note that there are additional costs to a student dropping out of a track and/or having to repeat a year. Not only does the government pay for each child's secondary education, delays may end up with students missing out on some earnings. These costs may also increase the downsides of the problem the parties face.

Research on track availability is sparse and research on the effects it may have on track recommendations does not exist to my knowledge. If track availability is indeed a factor that affects *track underestimation* and *underattendance* it would open up additional dimensions of discussion for decisions regarding the opening and closings of schools. Furthermore, depending on how policy makers view *track underestimation* and *underattendance* it may encourage them to change how upgrading track recommendations take place. They may make it mandatory for example.

Data

The main data is sourced from Statistics Netherlands (CBS) and consists of non-public microdata. Using their data, I gathered information on students' test scores, track recommendations, location, how many of their parents were born in the Netherlands and their parents' educational background in 2009-2019 and 2021. 2009 refers to the 2008/2009 school year. 2020 is excluded after Covid-19 related quarantine policies caused the CET test to be cancelled. The sensitive nature of the microdata used makes it not possible to publish the dataset for peer review. A secondary source of data is public data published by CBS and consists of information on the average distance to the nearest vmbo and havo/vwo school and the degree of urbanization in all municipalities, districts and neighbourhoods in the Netherlands under the 2021 CBS definition. Municipalities are made up of districts, which are in turn made up of neighbourhoods. The creation and management are up to the overarching municipality, but the guideline for a neighbourhood's population is between 250 and 2,500 inhabitants (CBS, 2023).

The initial and final track recommendations can consist of either individual tracks or a mixed recommendation which includes two or three 'adjacent' tracks. See Appendix table 1 for an overview of the possible track recommendations. The CET test is taken in February before 2015 and in April in the remaining years. The possible test score recommendations vary across the sample. For example, in 2009 only five possible test recommendations existed, while 2015 and 2021 had eight and six respectively[2]. This may affect the occurrence of track underestimation throughout the sample as teachers who want to exclude havo from the final

[2] There exists an additional possible recommendation (*praktijkonderwijs*) which is for when the regular pre-vocational tracks are considered too challenging. It has been excluded from discussions in this paper as it is taken by a small minority of students and falls outside the regular secondary educational system.

track recommendation would have to deviate from the test track recommendations to differing degrees depending on the year. This is year-dependent however. Only students who take the CET test are included in the final sample. While the CET (or its predecessor, Cito) represent the greater majority of tests before 2015, by 2018 this had dropped down to 56% of the time and has likely declined more since then (CPB, 2019). The reason for this sample restriction is that information about which test scores correspond to which test track recommendations are unreliable and unclear for all tests except the CET before 2015. Track recommendations may also include an indicator that a student will receive additional educational support. Distinctions in these recommendations are not seen as track underestimation and these indicators do not exist for the test track recommendations. Thus, all recommendations with such an indicator are replaced by their regular counterpart.

Other educational variables observed in the data are the date of the test taken, students' standardized test score which are rounded to discrete values, the track students attend during their first year of secondary school and finally, the distance to the closest vmbo and havo/vwo school averaged over all the inhabitants of a municipality, district and neighbourhood in the Netherlands. Additionally, for all students, I observe in which aforementioned municipality, district and neighbourhood they live according to municipality registers and their parents' educational background. Students' migration background is observed and defined as the number of parents not born in the Netherlands. Finally, the degree of urbanization of each neighbourhood, district and municipality in the Netherlands is observed per year in the sample. The degree of urbanization is measured in average address density in a one-kilometre radius around an address as defined by CBS.

The final sample spans twelve years and consists of 489,577 observations. I also aggregate the data to the neighbourhood level. The sample then consists of 1,924 neighbourhoods and 23,088 total observations, where each neighbourhood has twelve observations. 10,015 neighbourhoods with missing data for any of the years were excluded to form the sample. Table 1 provides summary statistics at the individual level.

Table 1

Summary Statistics

Variable	Total	Never Treated	Ever Treated
Cito/CET Score	535.1633	535.228	534.913
	(10.576)	(10.067)	(12.347)
Track underestimation	.097	.093	.112
	(.296)	(.290)	(.316)
Havo Track underestimation	.056	.054	.066
	(.230)	(.226)	(.247)
Vmbo Track underestimation	.030	.029	.035
	(.171)	(.169)	(.183)
Enrolled into PRO	.005	.005	.005
	(.069)	(.069)	(.069)
Enrolled into Vmbo B/K	.802	.801	.802
	(.399)	(0.400)	(.398)
Enrolled into Vmbo G/T	.053	.053	.053
	(.224)	(.224)	(.223)
Enrolled into Havo/Vwo	.141	.141	.140
	(.348)	(.348)	(.347)
Underattendance	.582	.583	.578
	(.493)	(.493)	(.494)
Havo Underattendance	.006	.006	.007
	(.078)	(.077)	(.083)
Vmbo Underattendance	.222	.221	.227
	(.416)	(.415)	(.419)
Extra distance to a Havo/Vwo school in km	.536	.353	1.241
	(1.579)	(.991)	(2.776)
Fewer than 1000 addresses within 1 kilometre	.511	.557	.334
	(.500)	(.497)	(.472)
Male	.496	.496	.498
	(.500)	(.500)	(.500)
Parents born outside the Netherlands	.416	.433	.350
	(.745)	(.755)	(.703)
Parent with at least a Master degree	.153	.160	.125
	(.360)	(.367)	(.331)
Parent with at least a Bachelor degree	.690	.689	.695
	(.463)	(.689)	(.460)
Observations	489,577	389,016	100,561

Note. This table provides mean summary statistics for the sample. Track underestimation is defined as receiving a lower final track recommendation than the test track recommendation. Enrolment can be one of four tracks and is based off the year after students finish primary education. Underattendance is defined as enrolling in a lower track than the track recommendation from the end-of primary school test, while Havo (Vmbo) underattendance is defined as a student enrolling in vmbo GL-TL (BL-KL) but receiving at least a havo (vmbo GL but less than a havo) test recommendation. Standard deviations are in parentheses.

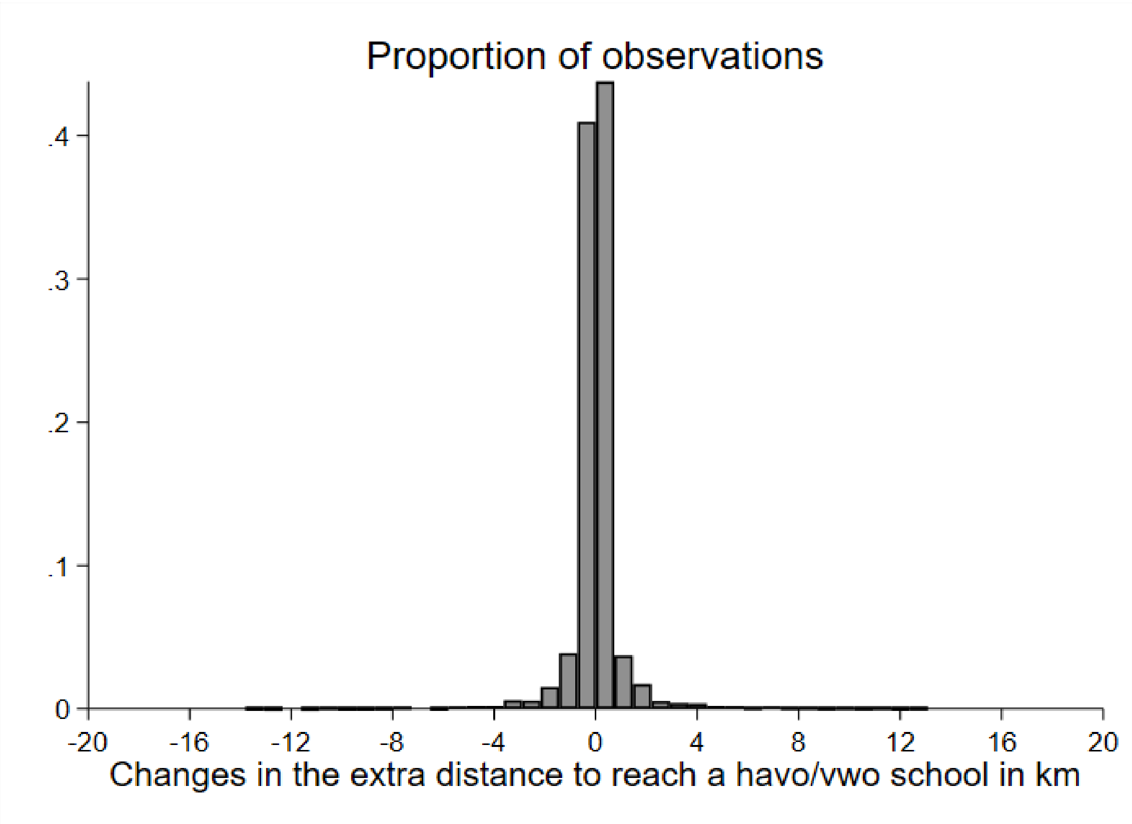
Around four-fifths of the sample lives in neighbourhoods that never experienced any change in the treatment variable. Which is the average additional distance to the nearest havo/vwo school compared to the nearest vmbo school for all inhabitants in a neighbourhood in the year of the observation. Notable differences between the never treated and ever treated are parental characteristics. Parents from children in the ever-treated group are less educated, make less money and are less likely to have been born outside the Netherlands. This lines up with the idea that this group lives in more rural environments. However, this group is substantially less likely to live in a neighbourhood with on average, fewer than one-thousand addresses in a one-kilometre radius around an address. This variable was supposed to indicate if a neighbourhood can be considered ‘rural’. It may thus be that this variable is not appropriate to Moreover, note that the outcome variables rarely occur. Track underestimation at the havo level occurs roughly six percent of the time. While underattendance occurs roughly only 0.6% of the time. This is contrasted by the occurrence of underattendance, which is 58.2% of the sample.

The reason for this contrast lies in the definition of these variables. Underattendance is defined in Appendix table 2. Havo underattendance is defined as a student enrolling into vmbo GL-TL and receiving at least a havo test recommendation. This means that those who receive at least a havo test recommendation but enrol into vmbo BL-KL or lower are counted for regular underattendance but not for havo underattendance. This occurs quite often in my sample at roughly 30% of the time. Underattendance and its subgroups are defined in this way to avoid overlap for placebo testing later on. Vmbo underattendance is defined as a student enrolling into vmbo BL-KL while receiving at least a vmbo GL but less than a havo test recommendation and will be used as a placebo test in this paper. Overlap between these two specific underattendance groups was thus avoided.

Figure 3 shows the distribution of the changes in the treatment variable compared to the previous year, using 2009 as the first observation. The treatment variable is measured in kilometres and is rounded to one decimal point. Observations with no change are excluded, which represents 72,26% or 15,293 neighbourhood-years of the full sample. The remaining 27,74% are distributed below. See figures 1-3 in the Appendix for more detail on the distribution of the treatment variable.

Figure 3

Additional Distance to Schools Offering a Havo/Vwo Track Compared to a Vmbo Track



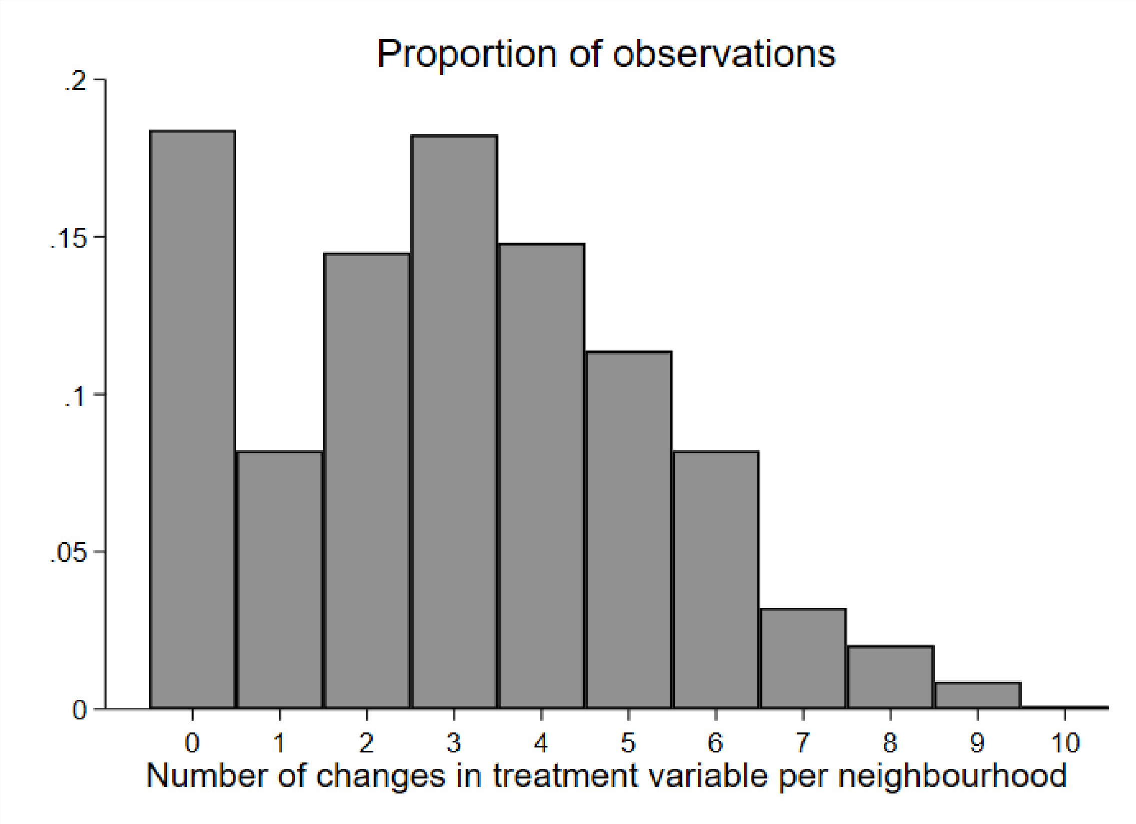
Note. This figure shows the distribution of the change in the average distance to the nearest havo/vwo school minus the average distance to the nearest vmbo school compared to the previous year per neighbourhood in the sample. Values of zero are excluded. Each bar represents 0.2 kilometres.

As can be seen in the figure, over eighty percent of changes are less than 0.2 kilometres since the previous year. Since changes would primarily be influenced by track openings and closings it may seem illogical for these to only influence our treatment variable in such a small way. Especially since neighbourhoods only occupy a small geological area, which would make it likely for any track openings or closings to have a large impact. However, the treatment variable only contains information of the neighbourhood as it existed at the time of the observation. Municipalities, districts and neighbourhoods are removed, added and changed geographically consistently throughout the time-span of the sample. Since each treatment datapoint is specific to a neighbourhood-year, small changes in the treatment over time are likely due to these geographic or demographic shifts instead of track openings and closings.

To address this issue, I single out larger changes in the treatment variable. Figure 4 depicts the number of changes in the treatment variable each neighbourhood experiences during the sample period. Over eighty percent of neighbourhoods experienced at least one change in the treatment variable. Three changes in a neighbourhood was the most common. Considering Figure 3, the great majority of these changes are less than even 0.2 kilometres.

Figure 4

The Rarity of Changes in the Treatment Variable

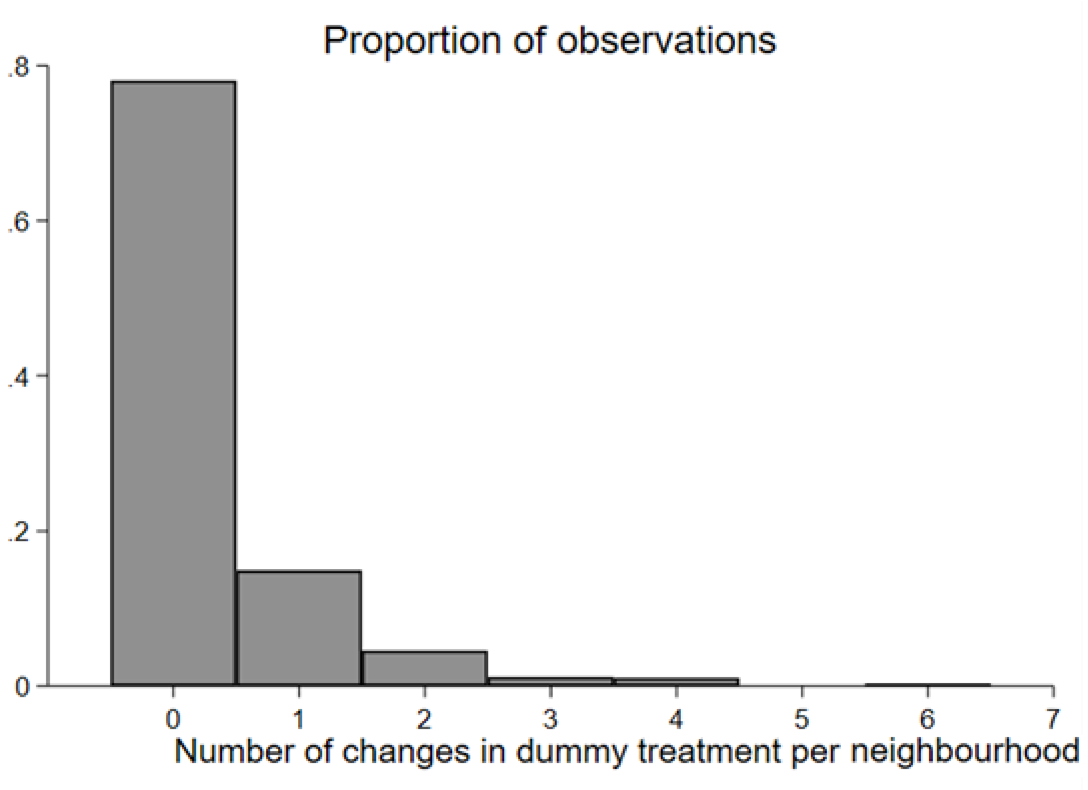


Note. This figure shows the distribution of how often the average distance to the nearest havo/vwo school minus the average distance to the nearest vmbo school compared to the previous year changes. The unit of observation is Dutch neighbourhoods.

Compare this to Figure 5, which is the same but requires an absolute change greater than a kilometre before a neighbourhood becomes 'treated'. Around eighty percent of neighbourhoods never experience any change in treatment status. Those that do usually only get treated once. More changes than two are quite rare and usually consist of a large positive change followed by a large negative change in the treatment variable or vice-versa. Throughout the rest of this paper, I maintain the requirement of a change greater than a kilometre needing to occur as the baseline for treatment status when using treatment dummies unless specified otherwise.

Figure 5

The Rarity of Large Changes in the Treatment Variable



Note. This figure shows the distribution of how often the average distance to the nearest havo/vwo school minus the average distance to the nearest vmbo school compared to the previous year changes more than a kilometre. The unit of observation is Dutch neighbourhoods, known as 'buurten'.

The original issue, that the definition of each unit of observation may shift demographically or geographically remains however. Although such changes are likely to be minor, they may affect a difference-in-difference design negatively since the method relies on being able to use unit fixed effects. Observed variation in variables that only occur due to an underlying change of a neighbourhood's population or definition may add noise to the final estimate, reducing my accuracy.

The, larger variation of the treatment variable may occur more often in rural areas where less schools are available to begin with. A new havo track opening in areas with fewer schools to begin with will likely affect the extra distance needed to get to a havo school more than if a new havo track is opened in an urban environment. This also brings into question if a reduction (increase) in the distance needed to get to a havo (vmbo) school has heterogeneous effects based on factors like urbanization, common methods of transport used and expectations that students and parents may have about the distance needed to travel to get to secondary school. If such heterogeneity is strong and significant effects are found anyway, it likely signals that the effects are even greater than estimated for the most relevant portion of those affected. Non-significant effects may accidentally ignore a portion of the sample where a strong effect does exist. The best way to estimate effects for these concepts would be to interact the treatment variable with relevant variable (For example, if a neighbourhood is rural or not). Unfortunately, I lack information about the above concepts to facilitate such an interaction except for the degree of urbanization. However, information about the density of addresses seems insufficient to capture the concept of 'urban vs rural'. This is because the density of addresses only captures information within a kilometre of an address and no other information that may be important such as the distance to large cities, the degree of trade or land usage. I thus avoid exploration on this topic throughout this paper and leave it for the future.

Empirical Specification

The empirical strategy used looks at changes in track underestimation in individuals affected by havo and vwo track openings relative to those not affected. This is estimated by a two-way fixed effects (TWFE) model with variation in treatment timing since not all track changes affect the same individuals. The main equation is:

$$Y_{ijt} = \alpha + \delta T_{jt} + \gamma W_{jt} + \sigma_j + \rho_t + \boldsymbol{X}_{ijt} + \epsilon_{ijt} \tag{1}$$

where Y_{ijt} reflects havo *track underestimation*, a dummy that equals one if student i in neighbourhood j has a final track recommendation below their test score recommendation, if that test recommendation was at least havo in year t and zero otherwise. T_{jt} is a second dummy that equals one if the additional distance that has to be travelled to the nearest havo or vwo track compared to the nearest vmbo track in student i 's neighbourhood j has increased with more than one kilometre in year t compared to year $t-1$. It also equals one for all subsequent years. W_{jt} functions similarly but is based off decreases of more than one kilometre instead. σ_j and ρ_t control for neighbourhood and year fixed effects respectively. Additionally, a second specification will be estimated where Y_{it} represents havo *underattendance*, a dummy that equals one if student i enrols in a track lower than their test score track recommendation if that test track recommendation was at least havo. Finally, $\boldsymbol{X}_{ijt}$ is a vector of time-varying controls.

Above I mentioned that both outcome variables have to be below the test recommendation. In practice, final and test track recommendations often include a range of multiple tracks. Furthermore, track enrolment as catalogued by the SOI2021 from CBS does not conform exactly with track recommendations. Overall, *track underestimation* is defined as receiving a final track recommendation that is a half-level or more below the test track recommendation. See Appendix table 1 and 2 for an overview of the conditions that result in *track underestimation* and *underattendance* respectively. I use the havo margin for both of these variables since the treatment variable also defines itself as the difference between havo/vwo and vmbo schools.

There are three assumptions that have to be met for equation 1 to be able to deliver causal estimates. First, variation in the treatment variable has to represent only shifts in track availability. This may seem simple, but as mentioned in the data chapter, demographic or geological changes to a neighbourhood may shift the additional distance to reach a havo or vwo track slightly from year to year. When observing students' location, I use the geological neighbourhood definitions from 2021. This keeps definitions for clusters consistent across the sample. Furthermore, because a dummy is used which only equals one at a minimum change compared to the previous year, small shifts in the underlying distance variable that may be attributed to demographic changes, recalculations or rounding issues are hopefully excluded.

The downside of this approach, is that potentially meaningful variation in the underlying continuous distance variable is reduced when converting that variable to a dummy. Beyond that, the effect of additional distance may have heterogenous effects for different groups of people, due to them using different methods of transport for example. This brings into question if a one-size-fits-all dummy is appropriate for causal inference in this context. I address this in the results section where the main results are compared with a different approach where I employ the additional distance to the nearest havo track as a continuous variable instead of converting it to two treatment dummies.

Second, the parallel trends assumption requires that in the absence of these track availability changes, individuals who had their average additional distance to the nearest school offering a havo or vwo track compared to a vmbo track changed would have had similar changes in *track underestimation* and *underattendance* as those who did not experience such changes. In TWFE models with variation in treatment timing this assumption instead becomes that of variance weighted common trends (VWCT). Goodman-Bacon (2019) shows that this assumption is technically a weaker one compared to the regular parallel trends assumption. This is because in TWFE models with differential treatment timing the estimated coefficient is made up of a weighted average of all individual 2×2 difference-in-differences in the data. These weights can technically correct for a situation of non-identical trends although the usefulness of this relaxation is questionable. However, the presence of time-varying treatment effects may generate bias. This occurs due to later 2×2s using earlier 2×2s as controls and vice versa. Therefore, if treatment timing affects the underlying individual ATTs, then this will cascade and end up biasing the final ATT.

These concerns are examined in two ways. First, regular parallel trends are tested by including leads and lags of treatment timing and regressing *Track underestimation* in the standard TWFE model. Note that this is of limited value in this context. As stated before, treatment neighbourhoods are not only compared to control neighbourhoods but also to other treated neighbourhoods. However, this test will hopefully provide useful information which can be used to influence the expectations for how likely it is that the VWCT holds. Second, multiple regressions are run where the sample does not run from 2009-2021 but starting from 2015 instead. How these point estimates compare to the main estimates will give some small indication on how important treatment timing is for this setting. Which in turn sheds light on the likelihood of bias originating from differential treatment timing.

Finally, the Stable Unit Treatment Value Assumption (SUTVA) as defined by Angrist, Imbens and Rubin (1996) states that potential outcomes for any given observation respond only to their own treatment status; potential outcomes are invariant to random assignment of others. In the context of this paper, it requires that teachers only change their track underestimation methods based off the neighbourhood-specific additional distance students have to travel to a havo/vwo school. It seems unlikely for teachers to internalize or even know their students' location fully. Likewise, teachers may have their track recommendation behaviour influenced not only by the closest schools but also by the second or third closest schools. Finally, since elementary schools enrol students from multiple neighbourhoods and since teachers likely base their track underestimation strategies off either a school wide-average distance to the closest secondary schools or the distance between the elementary school where they are employed to the closest secondary schools, it seems a foregone conclusion that teachers' behaviour likely changes and affects multiple neighbourhoods at once. This includes cases when some of those neighbourhoods are not technically treated.

This is a key issue if true and would invalidate the SUTVA assumption. However, large shocks to the additional distance to havo/vwo schools would be less likely to run into this problem as they are more likely to affect which schools are the closest to all the students in an elementary school at once. Since this paper only considers changes in the distance measure larger than a kilometre to be a treatment it partially corrects for this problem in the SUTVA assumption. The main problem remains however and would be better tackled if distance measures from elementary schools were used instead. Such data is not currently available and will have to be computed first. This was not deemed feasible in the timeframe of this paper and is thus not done despite its advantages.

Returning to the SUTVA, in repeated cross-sections (which this paper uses), it requires the composition of the underlying sample to not change over time. To test for this assumption, most regressions include several time-varying controls consisting of: the number of parents not born in the Netherlands, the degree of urbanization of a neighbourhood and if either of a students' parents have obtained a bachelor or master degree. It is a good sign for the chance that the SUTVA holds if the main point estimates do not change much compared to the regressions which include these time-varying controls.

Lastly, the level of clustering used in the standard errors needs to be discussed. Incorrect clustering of standard errors may lead to bias (Bertrand et al., 2004). It is not entirely clear at what level such clustering should take place as track availability changes are unlikely to fully line up with the geological boundaries of neighbourhoods, districts or even municipalities. I use clustering at the neighbourhood-year level since that is the source of variation in the treatment variable. I however acknowledge that this is not a perfect solution and it would be better to test for multiple levels of clustering.

Main Results

Table 2 presents the main results for the effects of track availability on havo track underestimation. The second column controls for the following time-varying variables: the number of parents not born in the Netherlands, a dummy that equals one if the average address density in a one-kilometre radius around an address is less than one-thousand and zero otherwise and two dummies that equal one if at least one parent obtained a bachelor or master degree respectively. The third column restricts the sample to 2015-2021 and the fourth column changes the threshold for the two treatment dummies to be greater than 0.5 kilometres compared to the default one kilometre. All columns restrict the sample to neighbourhoods that have received either one or zero absolute changes larger than one kilometre.

The estimated effect for a positive treatment on Havo track underestimation is an increase of 0.013 or 1.3 percentage points and is significant at a level of 1%. This estimate remains robust to other specifications except for when the sample is restricted to 2015-2021. The estimated effect for a negative treatment is an increase of 0.008 or 0.8 percentage points and is significant at a level of 5%. This estimate is robust to the addition of time-varying controls but becomes significant under the sample restriction to 2015-2021 or when a smaller threshold for the dummy variables is used. The robustness of the results to the addition of time-varying controls is a good sign regarding the potential problems of remaining time-varying omitted variable bias (OVB) and underlying compositional changes in the cross-sectional data. Although the estimates are statistically significant, their economic significance is weak. A 1.3 or 0.8 percentage point increase to the chance to receive a havo track underestimation is on its own negligible. It is possible that the estimates will increase when looking at even larger changes than one kilometre, which increases the economic significance as well. Overall, the results do not indicate that these changes in track availability are a major driver of track underestimation at the havo level.

Note that although the estimate for the negative treatment is significant, the expectation for this estimate was that its sign would be negative. A reduction in the additional distance to a havo/vwo school leads to an expectation of either a drop in havo track underestimation or alternatively, no effects at all. Furthermore, this estimate becomes negative (although also statistically indistinguishable from zero) when adjusting the threshold for the dummy treatments to 0.5 kilometres. It is unclear why the estimated coefficient is positive, but a possible cause may be that primary school teachers react negatively to both types of treatments but for different reasons. A positive change may result in an uptick to track underestimation due to the nearest havo/vwo school closing or a new vmbo school opening nearby, as outlined in the framework developed in the literature review. Simultaneously, if variation in negative treatments is mostly caused by new havo/vwo schools opening nearby, and primary school teachers are more hesitant to sending their students to these new schools, for example because these new schools still lack some staff. Then they may end up increasing the rate of track underestimation temporarily.

Table 2

Effects of Track Availability on Havo Track Underestimation

VARIABLES	(1) Havo Track underestimation	(2) Havo Track underestimation	(3) Havo Track underestimation	(4) Havo Track underestimation
Positive Treatment	0.013***	0.012***	0.000	0.011***
	(0.004)	(0.004)	(0.004)	(0.003)
Negative Treatment	0.008**	0.008**	-0.003	-0.002
	(0.004)	(0.004)	(0.005)	(0.003)
Constant	0.013	0.011	0.151	0.018
	(0.025)	(0.026)	(0.093)	(0.025)
Observations	457,558	457,558	422,078	414,015
R-squared	0.027	0.027	0.043	0.027
Neighbourhood FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Time Varying Controls	NO	YES	YES	YES
Sample	2009-2021	2009-2021	2015-2021	2009-2021
Threshold for Treatment	>1km	>1km	>1km	>0.5km

Note. This table shows the effects of large changes in the distance to the nearest havo/vwo school minus the distance to the nearest vmbo school per Dutch neighbourhood. Each student may receive a track underestimation individually and their neighbourhood is observed. Havo track underestimation is defined as a student receiving at least a havo track recommendation in the final primary school test, but receiving a final track recommendation below havo. Time-varying controls are the number of parents not born in the Netherlands, a dummy that equals one if the average address density in a one-kilometre radius around an address is less than one-thousand and zero otherwise and two dummies that equal one if at least one parent obtained a bachelor or master degree respectively Robust standard errors in parentheses and are clustered at the neighbourhood-year level. *** p<0.01, ** p<0.05, * p<0.

One may bring up the counter-argument of there not being any need for such an increase in track underestimation, after all, the nearest havo/vwo school from the previous year may still be an option and the teachers may send their students there. However, consider the possibility that teachers primarily influence the track students attend and thus the school they attend through their track recommendation as shown in Ree et al. (2023). If granting more havo track recommendations will result in the students going to their nearest havo school instead of the old 'better' option as proposed. Then teachers may feel the need to increase their rate of track underestimation to prevent their weakest students from attending this new, unstable school. Either way, the significant negative effects indicate that the theoretical framework behind the causal effects of track availability may be incomplete.

Column 3 limits the sample to 2015-2021. Note that this includes all neighbourhoods which had non-missing observations between 2015-2021 for the distance variables to the nearest havo/vwo and vmbo schools. Due to this, some neighbourhoods who are included in 2015-2021 do not exist in the 2009-2021 sample, which is the reason for the sample size remaining relatively similar between columns. This may be one reason for the lack of significant estimates. Other reasons may be that changing the panel length changes the estimate for no other reason than that it does, as shown by Goodman-Bacon (2019) and that there may simply be no effect since the Dutch education reform in 2015 where primary school teachers became more encouraged to increase track recommendations after high test scores.

Table 3 presents the main results for the effects of track availability on havo underattendance. This table mirrors Table 2 but uses havo underattendance as the outcome variable instead. No significant effects were found for the positive treatment. As in Table 2, the estimates were robust to adding time-varying controls and got closer to zero when using a threshold of greater than 0.5 kilometres as in column 3. The estimated effect for a negative treatment is -0.003 or a reduction in the chance for havo underattendance of 0.3 percentage points. This is statistically significant at the 5% level, but not economically significant considering the small effect size.

Table 3

Effects of Track Availability on Havo Track Underattendance

VARIABLES	(1) Havo Underattendance	(2) Havo Underattendance	(3) Havo Underattendance
Positive Treatment	0.000	0.000	-0.001
	(0.001)	(0.001)	(0.001)
Negative Treatment	-0.003**	-0.003**	-0.001
	(0.001)	(0.001)	(0.001)
Constant	0.000	-0.001	-0.002
	(0.001)	(0.002)	(0.002)
Observations	457,558	457,558	414,015
R-squared	0.021	0.021	0.021
Neighbourhood FE	YES	YES	YES
Year FE	YES	YES	YES
Time Varying Controls	NO	YES	YES
Sample	2009-2021	2009-2021	2009-2021
Threshold for treatment	>1km	>1km	>0.5km

Note. This table shows the effects of the distance to the nearest havo/vwo school minus the distance to the nearest vmbo school per Dutch neighbourhood. The resulting distance variable is the treatment. In column 3, Treatment is increased by 10 kilometres for everyone to avoid any negative values. Each student may receive a track underestimation individually and their neighbourhood is observed. Havo underattendance is defined as a student receiving at least a havo track recommendation in the final primary school test, but enrolling in a track below havo in the first year of secondary school. Time-varying controls are the number of parents not born in the Netherlands, a dummy that equals one if the average address density in a one-kilometre radius around an address is less than one-thousand and zero otherwise and two dummies that equal one if at least one parent obtained a bachelor or master degree respectively Robust standard errors in parentheses and are clustered at the neighbourhood-year level. *** p<0.01, ** p<0.05, * p<0.1

The main differences between Table 2 and Table 3 are twofold. First, the estimates for a positive treatment are no longer statistically significant. Second, the sign for the estimates for a negative treatment are now negative. In the literature review I outlined that underattendance is the result of a decision problem that a students' family faces *after* the teacher decides to not underestimate a students track. It makes sense therefore that the estimates in Table 3 are overall smaller compared to Table 2.

Assuming that these results show causal effects, they indicate that students do not often decide to enrol into a track below their final track recommendation. Furthermore, the reaction from students and their parents to under-attend less often when a havo school gets relatively closer is in line with the expectations set up in the theoretical framework. This lines up with the theory that teachers become more likely to underestimate a students' track when a new havo/vwo school appears in the area. After all, if students and their parents care the most about the track the school offers and do not look much further, then teachers may 'protect' their students through track underestimation. Making it more likely they attend a vmbo school instead.

Table 4 estimates the effect of track availability on havo track underestimation. This time a continuous treatment variable is used instead of two separate treatment dummies. These estimates are largely insignificant, except for in column 3, where polynomials[3] of the treatment variable were used. There the cubed version of the treatment variable is significant at the 10% level. These results are consistent with the idea that a continuous treatment variable does not work well due to shifts in geographic neighbourhood definitions, resulting in a large amount of noise.

[3] The treatment variable was increased by 10 in column 3 in Tables 4 and 5 to avoid any negative values.

Table 4

Effects of Track Availability on Havo Track Underestimation Using a Continuous Treatment

VARIABLES	(1) Havo Track underestimation	(2) Havo Track underestimation	(3) Havo Track underestimation	(4) Havo Track underestimation
Treatment	0.000 (0.001)	0.000 (0.001)	-0.005 (0.006)	-0.000 (0.001)
$\mathrm{Treatment}^2$			0.001 (0.000)	
$\mathrm{Treatment}^3$			-0.000* (0.000)	
Constant	0.016 (0.024)	0.013 (0.025)	0.014 (0.033)	0.143 (0.092)
Observations	489,577	489,577	489,577	452,415
R-squared	0.027	0.027	0.027	0.043
Neighbourhood FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Time varying Controls	NO	YES	YES	NO
Sample	2009-2021	2009-2021	2009-2021	2015-2021

Note. This table shows the effects of the distance to the nearest havo/vwo school minus the distance to the nearest vmbo school per Dutch neighbourhood. The resulting distance variable is the treatment. In column 3, Treatment is increased by 10 kilometres for everyone to avoid any negative values. Each student may receive a track underestimation individually and their neighbourhood is observed. Havo track underestimation is defined as a student receiving at least a havo track recommendation in the final primary school test, but receiving a final track recommendation below havo. Time-varying controls are the number of parents not born in the Netherlands, a dummy that equals one if the average address density in a one-kilometre radius around an address is less than one-thousand and zero otherwise and two dummies that equal one if at least one parent obtained a bachelor or master degree respectively Robust standard errors in parentheses and are clustered at the neighbourhood-year level. *** p<0.01, ** p<0.05, * p<0.1

Table 5

Effects of Track Availability on Havo Underattendance Using a Continuous Treatment

VARIABLES	(1) Havo Underattendance	(2) Havo Underattendance	(3) Havo Underattendance
Treatment	0.063***	0.063***	-0.233***
	(0.013)	(0.013)	(0.085)
Treatment^2			0.021**
			(0.008)
Treatment^3			-0.000**
			(0.000)
Constant	-0.153*	-0.296*	0.370
	(0.091)	(0.156)	(0.342)
Observations	489,577	489,577	489,577
R-squared	0.020	0.021	0.021
Neighbourhood FE	YES	YES	YES
Year FE	YES	YES	YES
Time Varying Controls	NO	YES	YES
Sample	2009-2021	2009-2021	2009-2021

Note. This table shows the effects of the distance to the nearest havo/vwo school minus the distance to the nearest vmbo school per Dutch neighbourhood. The resulting distance variable is the treatment. In column 3, Treatment is increased by 10 kilometres for everyone to avoid any negative values. Each student may receive a track underestimation individually and their neighbourhood is observed. Havo track underattendance is defined as a student receiving at least a havo track recommendation in the final primary school test, but enrolling in a track below havo in the first year of secondary school. Time-varying controls are the number of parents not born in the Netherlands, a dummy that equals one if the average address density in a one-kilometre radius around an address is less than one-thousand and zero otherwise and two dummies that equal one if at least one parent obtained a bachelor or master degree respectively Robust standard errors in parentheses and are clustered at the neighbourhood-year level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5 estimates the effect of track availability on havo underattendance with a continuous treatment variable. Havo underattendance has been multiplied by a factor of 100 to get clearer estimates. Although the noise from geographic shifts likely remains. I estimate a statistically significant positive effect of track availability on havo underattendance at the 1% significance level. Columns 1 and 2 estimate that for each additional kilometre a student has to travel to a havo school beyond a vmbo school they are 0.063 percentage points more likely to experience underattendance. Column 3 uses polynomials of the treatment variable and estimates a positive significant effect once the nearest havo school is located two or more kilometres further away than the nearest vmbo school. As before, the estimates are robust to the addition of time-varying controls and are not economically significant.

A curious pattern emerges from these results. Table 4 estimates the same effect as Table 2 but uses a continuous treatment. The same is true for Table 5 and Table 3. In both cases do I find opposite results in terms of statistical significance whenever estimating effects using a continuous treatment compared to the dummy treatments. It is unclear why this is the case, but it is a poor sign for the trustworthiness of the results.

Robustness Checks

In this chapter I attempt to test the parallel trends assumption and perform placebo tests for the results which yielded significant estimates in the previous chapter.

First, I attempt to test the parallel trends assumption by adding leads and lags to the main regression for estimating the effect of track availability on track underestimation. These replace the year fixed effects. Instead, five leads and five lag dummies are added which depend on the relative time of treatment. The results can be found in Table 6. Ideally, all the lead variables are insignificant, while the lag variables are significant, showing that before treatment took place, treatment and control neighbourhoods are quite similar, and diverge post treatment. Table 6 examines pre-treatment dynamics, and finds that regarding positive treatment changes, treated and control neighbourhood differed significantly before the treatment was implemented. This difference fizzles out in the years after the treatment. The results are thus exactly the opposite of what would be a good sign for the validity of the main results. Note that a similar regression was done, excluding the positive treatment variable, which yielded no different qualitative conclusions.

Second, Table 7 depicts the results of three separate placebo tests. Instead of using havo track underestimation or underattendance. The placebo tests use vmbo track underestimation or underattendance. The idea is simple. Treatment status is based on the difference in distance to the nearest havo/vwo school and the nearest vmbo school and how that changes year-to-year. Variation in the treatment variable should thus not affect vmbo track underestimation or underattendance. These outcome variables equal one when a student receives a final vmbo track recommendation lower than their vmbo test track recommendation or enrols in a lower vmbo track compared to their vmbo test track recommendation. See Appendix tables 2 and 3 for a detailed overview of the different types of vmbo tracks. Since the outcome variables looks at differences between subtypes of vmbo and the treatment variable does not distinguish between different vmbo tracks we should largely find no effect.

Table 7 has three columns. The first column runs a placebo test based off the significant effects found in Table 2, column 2. The second column functions as a placebo for Table 5, column 2. Lastly, the third column is a placebo test for Table 5, column 3. These three results were chosen to be tested since they all yielded significant results. The second and third placebo tests find no significant estimates. The first placebo test finds no significant effects for the positive treatment which is a good sign. The negative treatment however remains significant and has a negative sign. It estimates the effect of a negative treatment to be -0.005 or a reduction in the chance to receive track underestimation of 0.5 percentage points and is significant at the 5% level.

It is possible for this coefficient to be significant and the underlying design to be solid. After all, if a havo school that also offered vmbo-Tl and vmbo-Gl tracks opened nearby, it may change the distance to the nearest havo school compared to the nearest vmbo school from a number larger than one to zero. Then that would satisfy the requirement for the dummy and possibly cause vmbo track underestimation to drop. However, such events are likely to be rare and it raises the question why the other coefficients are not significant. Furthermore, this same treatment dummy yielded a surprising result when estimating the effect of track availability on havo track underestimation. It may thus also be a possibility that such negative shocks to the additional distance to the nearest havo school are somehow more complex and yield significant results for an unknown reason.

Table 6

Testing the Parallel Trends Assumption

VARIABLES	(1) Havo Track Underestimation
Lead5	0.015**
	(0.007)
Lead4	0.012*
	(0.006)
Lead3	0.043***
	(0.007)
Lead2	0.028***
	(0.006)
Lead1	0.022***
	(0.007)
Lag1	-0.000
	(0.008)
Lag2	0.023**
	(0.011)
Lag3	-0.004
	(0.008)
Lag4	0.011
	(0.008)
Lag5	0.039***
	(0.014)
Positive Treatment	0.025***
	(0.006)
Constant	0.028
	(0.017)
Observations	426,634
R-squared	0.025
Neighbourhood FE	YES
Year FE	YES
Time Varying Controls	YES
Sample	2009-2021

Note. This table examines pre-treatment dynamics by adding leads and lags to the regular Diff-in-Diff regressions. A treatment occurs when the distance to the nearest havo/vwo school minus the difference to the nearest vmbo school increases with more than a kilometre compared to the previous year. Distances are measured per Dutch neighbourhood. Havo track underestimation occurs when a student scores at least havo on the final primary school test but receives a final track recommendation below havo. Time-varying controls are the number of parents not born in the Netherlands, a dummy that equals one if the average address density in a one-kilometre radius around an address is less than one-thousand and two dummies that equal one if at least one parent obtained a bachelor or master degree respectively Robust standard errors in parentheses and are clustered at the neighbourhood-year level. *** p<0.01, ** p<0.05, * p<0.1

Table 7

Placebo Tests

VARIABLES	(1) Vmbo Track Underestimation	(2) Vmbo Underattendance	(3) Vmbo Underattendance
Positive Treatment	-0.002		
	(0.003)		
Negative Treatment	-0.005**		
	(0.003)		
Treatment		-0.001	1.118
		(0.001)	(2.834)
Treatment^2			-0.112
			(0.208)
Treatment^3			0.003
			(0.005)
Constant	0.043*	0.241***	21.106
	(0.026)	(0.041)	(13.214)
Observations	457,558	457,558	457,558
R-squared	0.023	0.021	0.021
Neighbourhood FE	YES	YES	YES
Year FE	YES	YES	YES
Controls	YES	YES	YES
Sample	2009-2021	2009-2021	2009-2021

Note. This table performs placebo tests by regressing the vmbo track underestimation and vmbo underattendance on the additional distance to the nearest havo/vwo school minus the distance to the nearest vmbo school. Positive (negative) treatment indicates a positive (negative) change in that distance greater than one kilometre. Treatment is the underlying continuous distance variable. Column 3 increases the treatment variable by 10 kilometres to ensure only positive values exist. Distances are measured per Dutch neighbourhood. vmbo track underestimation occurs when a student scores one of the vmbo tracks on the final primary school test but receives a final track recommendation below that test recommendation. Time-varying controls are the number of parents not born in the Netherlands, a dummy that equals one if the average address density in a one-kilometre radius around an address is less than one-thousand and two dummies that equal one if at least one parent obtained a bachelor or master degree respectively Robust standard errors in parentheses and are clustered at the neighbourhood-year level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Overall, the robustness checks indicate that the design itself is likely solid. After all, the placebo tests mostly went well. No significant effects were found except for one variable which had already yielded weird results. The placebo test thus improved the trust I have in the most significant estimates from the main results chapter. The regression where leads and lags were added to examine pre-treatment dynamics however, showed that the treatment and control neighbourhoods differed greatly from each other in the years leading up to the treatment. Even with neighbourhood fixed effects. This indicates that it is unlikely for the parallel trends assumption to hold. After all, if the neighbourhoods were very different from each other before treatment, why would they follow the same trendline in outcome variables post treatment?

Conclusion

This thesis has attempted to examine the causal effect of track availability on track underestimation and underattendance using a difference in difference design with differential treatment timing. When using treatment dummies based on large changes in the distance to the nearest havo/vwo school and vmbo school, I find significant positive effects on track underestimation for both large negative and positive changes. These results were robust to adding time-varying controls and large positive changes were robust to adjusting the threshold for the treatment dummy from one kilometre to 0.5 kilometres. Large negative changes were not robust to the usage of the different threshold and remained significant when using a placebo test where no effects were expected to be found. Neither treatment dummy was robust to adjusting the sample years from 2009-2021 to 2015-2021. No significant effects were found for underattendance.

When using a continuous treatment, no significant effects were found for track underestimation but significant effects were found for the effect on underattendance. This contrast with treatment dummies indicates that underattendance may be influenced primarily by smaller changes of the treatment variable, not captured by the treatment dummies. This result was robust to adding time-varying controls and the estimates failed to remain significant in the placebo test.

The main limiting factor in how I examine the causal relationship between track availability and track underestimation and underattendance is the data regarding distances to the nearest schools. I use public data at the neighbourhood level and the only distance data available is the average distance to the nearest havo/vwo school and vmbo school. This caused many issues. Geographic shifts in neighbourhood definitions caused most of the variation to be noise. Treatment dummies based of large changes were used to operationalize track availability, but it would have been best if individual distances or distances from elementary schools to the nearest secondary schools could have been used. It would have greatly expanded the data to include much more people, who are currently excluded because their neighbourhood did not exist during all twelve years the sample spans. It would also have improved the accuracy of my estimates and allowed me to examine more than the margin between havo and vmbo tracks. Unfortunately, this was not feasible to execute in the timespan allotted for this paper. I thus highly recommend future research on this topic to use those measures to improve the research quality of their design.

That same limitation was the main reason behind the lack of research on heterogeneity. Adjacent research regarding topics like heterogeneous effects based on differing distances, methods of transport and how rural an area is would have also been valuable to examine. If one adjusts the Difference-in-Difference regressions to be logit instead of OLS, they could examine the heterogeneous effects of distance through for example, investigating marginal effects at different distances.

Overall, the internal validity of the research is not terrific, the largest problem is the likely failure of the parallel trends assumption as treatment and control neighbourhoods differed significantly both before and after treatment. This indicates that it is unlikely that treatment neighbourhoods would have had the same trends in track underestimation if they had not been treated. That being said, this is the first research attempting to identify causal effects of track underestimation and underattendance. The hope is that this paper functions both as additional information to inform policy and as a roadmap of issues to avoid and additional avenues to explore in future research.

The external validity of the sample to the rest of the Netherlands is likely to be poor. It would not be unreasonable to assume that newly build and defined neighbourhoods differ from the neighbourhoods which remained in existence during the sample period. Beyond the Netherlands the external validity reduces even more. However, concepts like track underestimation and underattendance are often not applicable in other countries, making this issue not very relevant for most countries.

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249-275.

Boone, S., & Van Houtte, M. (2013). Why are teacher recommendations at the transition from primary to secondary education socially biased? A mixed-methods research. *British Journal of Sociology of Education*, *34*(1), 20-38.

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

Centraal Bureau Statistiek, (2023). *Richtlijnen voor gemeenten bij het vastellen van een indeling naar wijken en buurten.* Retrieved from https://www.cbs.nl/-/media/cbs/dossiers/nederland-regionaal/gemeente/gemeente-en-regionale-indelingen/richtlijnen-voor-gemeenten-vaststellen-indeling-naar-wijken-en-buurten.pdf

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

Dustmann, C., Puhani, P. A., & Schönberg, U. (2017). The long-term effects of early track choice. *The Economic Journal*, *127*(603), 1348-1380.

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

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254-277.

Lam, H., Jansen, M., van de Kraats, C., & Ravesteijn, B. (mimeo).

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.

Oosterbeek, H., ter Meulen, S., & van Der Klaauw, B. (2021). Long-term effects of school-starting-age rules. *Economics of Education Review*, *84*, 102144.

de Ree, J., Oosterveen, M., & Webbink, D. (2023). The quality of school track assignment decisions by teachers. *arXiv preprint arXiv:2304.10636*.

Schippers, K. (2022). The effects of Track Recommendations on the secondary school career. *Business Economics*. Retrieved from: http://hdl.handle.net/2105/62903

Timmermans, A. C., de Boer, H., Amsing, H. T., & 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), 847-874.

WVO (2014). Artikel 3, lid 2, Inrichtingsbesluit Wet op het voortgezet onderwijs. Datum inwerktreding: 01/08/2014. Wetten.overheid.nl

Appendix

Table 1

Criteria for track underestimation

Test track recommendation → Final Track recommendation ↓	10 Praktijkonderwijs	20 VMBO BL	22 VMBO BL t/m VMBO KL	30 VMBO KL	34 VMBO KL t/m VMBO TL	40 VMBO GL	42 VMBO GL t/m VMBO TL	44 VMBO GL t/m HAVO	60 HAVO	61 HAVO -VWO	70 VWO
01 VSO											
10 Praktijkonderwijs											
20 VMBO BL											
22 VMBO BL t/m VMBO KL											
30 VMBO KL											
32 VMBO KL t/m VMBO GL											
34 VMBO KL t/m VMBO TL											
40 VMBO GL											
42 VMBO GL t/m VMBO TL											
44 VMBO GL t/m HAVO											
50 VMBO TL											
52 VMBO TL t/m HAVO											
60 HAVO											
61 HAVO-VWO											
70 VWO											

Note. This table depicts which combinations of final track recommendations and test track recommendations result in track underestimation. Any individuals with a combination of the two track recommendations marked in orange are registered as being track underestimated. The numbers reference the way track recommendations are indexed in the source data from the Central Bureau of Statistics in the Netherlands.

Table 2

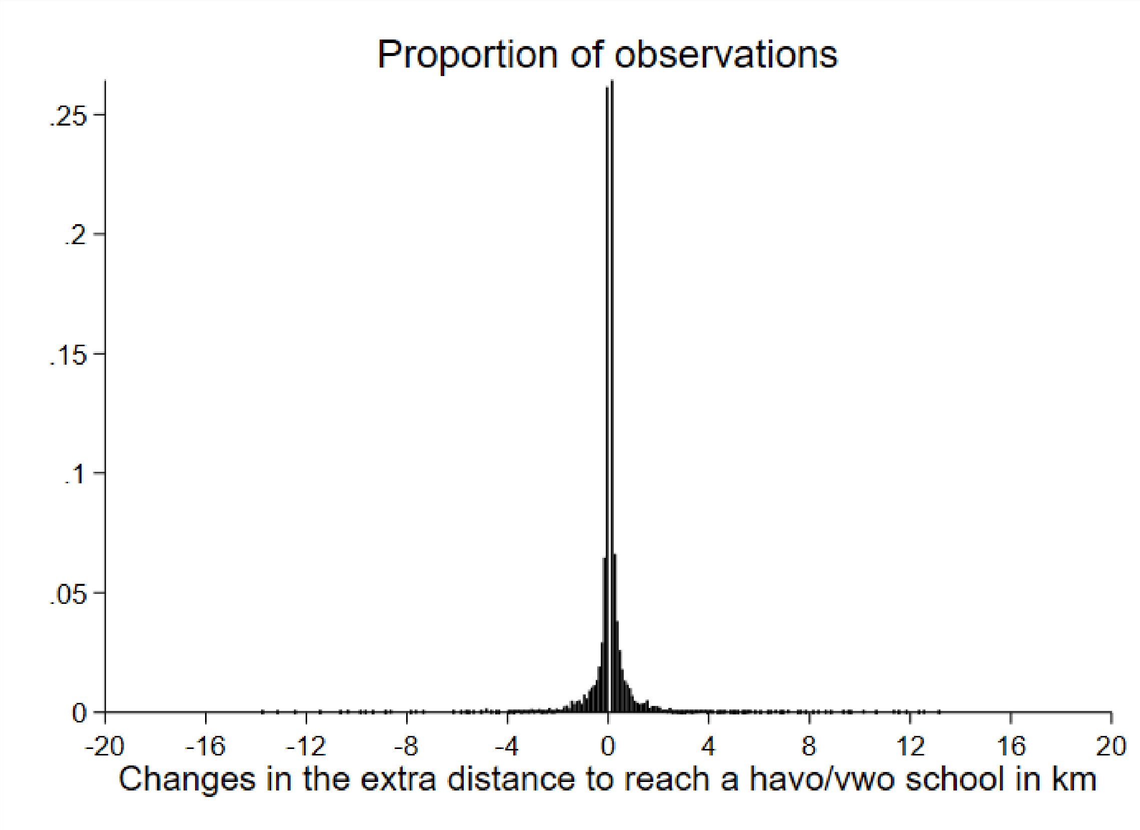
Criteria for underattendance

Test track recommendations → Track students enrol into ↓	10 Praktijkonderwijs	20 VMBO BL	22 VMBO BL t/m VMBO KL	30 VMBO KL	34 VMBO KL t/m VMBO TL	40 VMBO GL	42 VMBO GL t/m VMBO TL	44 VMBO GL t/m HAVO	60 HAVO	61 HAVO -VWO	70 VWO
1211 Praktijkonderwijs											
1212 VMBO BL-KL											
1221 VMBO GL-TL											
1222 HAVO-VWO onderbouw											

Note. This table depicts when underattendance takes place. Column numbers reference the way track recommendations are indexed in the source data from the Central Bureau of Statistics in the Netherlands. The track students enrol into are one of four possible categories relevant to first year secondary school students. The categories originate from the Standaard Onderwijsindeling (SOI) 2021.

Figure 1

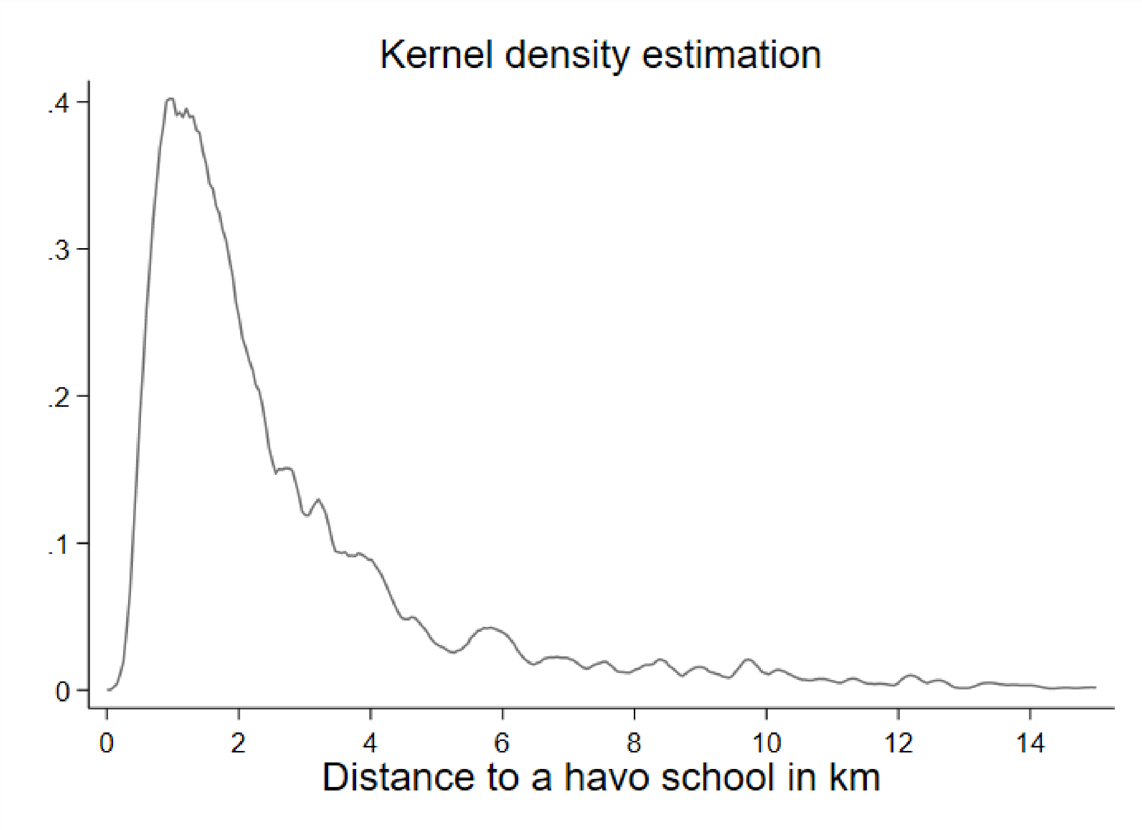
Additional Distance to Schools Offering a Havo/Vwo Track Compared to a Vmbo Track



Note. This figure shows the distribution of the change in the average distance to the nearest havo/vwo school minus the average distance to the nearest vmbo school compared to the previous year per neighbourhood in the sample. Values of zero are excluded. Each bar represents 0.1 kilometres.

Figure 2

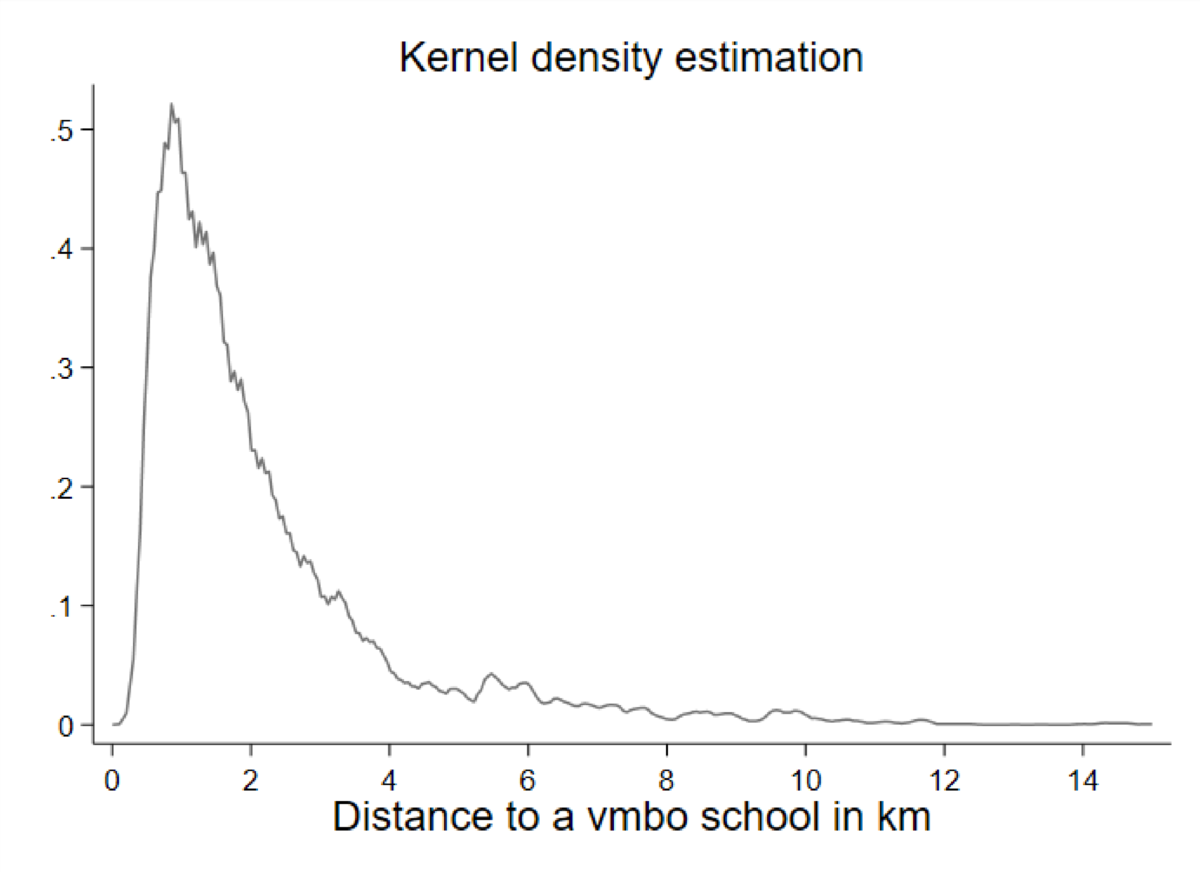
Distribution of Changes in the Distance to the nearest Havo/Vwo School



Note. This figure shows the distribution of average distance to the nearest havo/vwo school per individual in the sample. Average refers to the neighbourhood of each individual.

Figure 3

Distribution of Changes in the Distance to the nearest Vmbo School



Note. This figure shows the distribution of average distance to the nearest vmbo school per individual in the sample. Average refers to the neighbourhood of each individual.