CPB Document

No 182

The effect of early tracking on participation in higher education

Roel van Elk, Marc van der Steeg, Dinand Webbink

Centraal Planbureau
Van Stolkweg 14
PO Box 80510
2508 GM The Hague

Telephone +31-70-338 33 80
Telefax +31-70-338 33 50
Internet www.cpb.nl

ISBN 978-90-5833-398-8

Abstract in English

This paper examines the impact of early tracking on enrollment in and completion of higher education. We compare pupils that are directly tracked in lower general secondary education ('mavo') to pupils that postpone their choice of education level by entering secondary education in a combined first-grade class. Potential self-selection problems are addressed in two ways. First of all, using micro data allows us to control for a large set of individual background characteristics including tests of cognitive ability. Second, we exploit differences in regional supply of particular school types. The estimates show that early tracking has a detrimental effect on enrollment in and completion of higher education for pupils who leave primary education with a mavo advice. In addition, we find no evidence that pupils who leave primary education with a higher general secondary education ('havo') advice would be negatively affected by being in a comprehensive class together with the mavo advice pupils. Enrollment in and completion of higher education can be increased by stimulating participation in combined first-grade classes that keep pupils with a mavo or havo advice together for an additional one or two years.

Key words: early tracking

Abstract in Dutch

Dit paper onderzoekt het effect van vroege selectie op de deelname aan en het afronden van het hoger onderwijs. Hiervoor vergelijken we leerlingen die vroeg geselecteerd zijn (d.w.z. direct naar een categorale mavo gaan) met leerlingen die 1 of 2 jaar later geselecteerd worden door het volgen van een gecombineerde brugklas. Er wordt op twee manieren met mogelijke zelfselectie omgegaan. We controleren voor een grote verzameling individuele achtergrondkenmerken zoals testscores op cognitieve vaardigheden. Daarnaast maken we gebruik van verschillen in regionaal aanbod van bepaalde typen scholen. De schattingen laten zien dat vroege selectie een negatief effect heeft op de deelname aan en het afronden van het hoger onderwijs voor de leerlingen met een mavo-advies. We vinden bovendien geen aanwijzing dat leerlingen met een havo-advies een negatieve invloed zouden ondervinden van leerlingen met een mavo-advies in de gecombineerde brugklas. De deelname aan het hoger onderwijs en het aantal hoger opgeleiden in Nederland kunnen vergroot worden door het stimuleren van deelname aan gecombineerde brugklassen waarin leerlingen met een mavo- of havo-advies één of twee jaar bij elkaar gehouden worden.

Steekwoorden: vroege selectie

A comprehensive summary is available from www.cpb.nl.

Contents

Preface		7
Summary		9
1	Introduction	11
2	Empirical Strategy	17
2.1	Variation in the timing of tracking in Dutch education	17
2.2	Identification	19
3	Data description	23
4	Results	29
4.1	The effect of regional supply on school type choice	29
4.2	The effects of early tracking	31
5	Effects after the introduction of vmbo	35
6	Robustness checks	39
6.1	The role of entrance tests as proxy for ability	39
6.2	Various definitions of tracking	40
7	Effects for pupils with higher school advices	43
8	Conclusions and discussion	47
References		51
Appendix A		54
Appendix B		57
Appendix C		60

Preface

In Dutch education, pupils are tracked into different school types directly after leaving primary school, at age twelve. According to the OECD, this relatively early timing of tracking is a potential constraint for the growth of higher education participation. This study examines the impact of early tracking on enrollment in and completion of higher education. The research was executed as a part of the so-called 'experiment budget', which was raised by the Ministry of Education in 2007. This budget facilitates experimental studies in the field of education with the aim to contribute to more 'evidence-based policy'.

The document has been written by Roel van Elk, Marc van der Steeg and Dinand Webbink. First of all, the authors are grateful to the members of the advisory committee: Paul van Oijen, Vera Pieterman and Tim Schokker from the Ministry of Education, Niels Achterberg and Jurgen Geelhoed from the Ministry of Economic Affairs, and Maarten Cornet from the Ministry of Finance. Furthermore, we thank Jaap Roeleveld (SCO-Kohnstamm Institute), Hans Kuyper (GION), Maurice Krul (UvA) and Ib Watterreus (Education Council of the Netherlands) for their help and worthful comments. The study also gained from the valuable comments of Pierre Koning, Free Huizinga and George Gelauff.

Coen Teulings
Director

Summary

In the Netherlands, both the level and growth of participation in higher education is low compared to other rich countries (CPB, 2005). According to the OECD[1], the early tracking regime in the Netherlands causes a severe constraint for the growth of higher education participation. It states that 'In the end, postponement of the present early tracking regime seems inevitable, although this is a major change in the way Dutch society thinks of itself' (OECD, 2007, p.38).

Early tracking makes that pupils are grouped by ability in different school types and that they face different peers. This may have major consequences for future educational outcomes. This paper examines the impact of early tracking on enrollment in and completion of higher education.

We use a within country analysis in which we exploit a special feature of the Dutch education system. At the start of secondary education, schools in the Netherlands offer different types of first-grade classes. In some schools, pupils are directly tracked into categorial classes of a certain education level. Other schools offer one- or two-year combined classes, in which pupils are kept together before they are tracked in a particular school type. This causes variation in the timing of tracking which we exploit to analyze the effects of early tracking.

A concern regarding our analysis is that pupils may self-select into categorial or comprehensive classes. We address this potential problem first by including a large set of covariates. Using micro data of pupils allows us to control for various individual background characteristics including test scores on arithmetic, language and information processing. Second, we adopt an instrumental variables approach in which we exploit exogenous variation in tracking driven by the supply of different school types in a pupil's region.

The estimates show that early tracking has a negative effect on participation in higher education. Early tracking decreases the probability of both enrollment in and completion of higher education by around 4%- points for pupils with a mavo advice. In the cohorts of pupils that enter secondary education in 1989 and in 1993, average enrollment and completion of higher education for the tracked pupils is around 22 and 35%, respectively. Hence, pupils with a mavo advice that are in a categorial mavo can increase their probabilities to enroll and complete higher education to around 26 and 39%, respectively, by entering secondary education in a combined mavo-havo(-vwo) class. Additional analyses suggest that while pupils with a mavo advice gain from being in a comprehensive class, higher ability pupils do not experience

[1] Organisation for Economic Cooperation and Development.

negative consequences from this. This points to an inefficiency of the early tracking regime with respect to participation in higher education.

Our study supports the OECD conclusion on the Dutch education system that the early tracking regime has a negative effect on participation in higher education. Policies that stimulate participation in comprehensive classes for mavo pupils can raise enrollment and completion of higher education.

1 Introduction

In the Netherlands, both the level and growth of participation in higher education is low compared to other rich countries (CPB, 2005). According to the OECD, the early tracking regime in the Netherlands causes a severe constraint for the growth of higher education participation. It states that 'In the end, postponement of the present early tracking regime seems inevitable, although this is a major change in the way Dutch society thinks of itself' (OECD, 2007, p.38). Considering the commitment of the Dutch government to the Lisbon goal to raise participation in higher education, an evaluation of the effects of early tracking is highly relevant from a policy perspective.

We define tracking as the allocation of students into differing ability schools or classes. In Dutch education system, pupils are selected into different school types directly after leaving primary school, at age twelve. This is early compared to other countries like the United States, Canada, the United Kingdom and the Scandinavian countries who keep their entire lower secondary school system comprehensive. Early tracking makes that pupils are grouped by ability in different school types and that they face different peers. This may have major consequences for future educational outcomes. This paper examines the impact of early tracking on enrollment in (and completion of) higher education.

Various potential effects of early tracking on educational outcomes have been suggested.[2] On the one hand, proponents of tracking often use the argument that homogeneous classrooms, with a lower ability variance, allow a more focused curriculum. Teaching becomes easier as the course material can be better adapted to the needs of the pupils. An appropriate pace of instruction, without losing the lowest ability pupils or boring the best ability pupils in class, contributes to an optimal learning environment for all pupils.

On the other hand, critics of tracking argue that there is a positive relationship between average pupil ability in class and individual performance. According to this view, high ability pupils benefit from tracking at the expense of low ability pupils. This may be induced by either teacher quality effects or peer effects. If better teachers prefer to teach relatively high ability classes, teacher sorting may result in a higher quality of education in these classes.

Low ability pupils may benefit from the interaction with higher ability pupils. If peer effects are nonlinear, this gives rise to efficiency gains. In case of linear peer effects, however, tracking would not have an impact on educational outcomes (Argys et al., 1996). In the empirical literature there is no consensus yet on the existence and the size of peer group effects and the role of peers in the discussion on early tracking remains disputed.[3]

[2] An overview of the literature can be found in Van de Werfhorst and Mijs (2007). Epple et al. (2002) and Brunello and Giannini (2004) present some theoretical models of school tracking.

[3] Some recent studies on peer effects, which also address issues like the reflection problem and self-selection into groups of peers, are Hoxby (2000) and Sacerdote (2001).

Besides effects driven by the composition of classes, there can be effects due to institutional differences. In an early tracking regime it may be difficult for a pupil, once being assigned to a particular education level, to change tracks. A comprehensive system is more flexible as pupils keep the same opportunities until the time a selection is made. Tracking pupils at an early age implies more uncertainty with respect to the pupils' true capabilities and hence a higher risk of mistakenly sending them to the wrong school type.

It is difficult to empirically identify causal effects of early tracking. Firstly, there is often little variation in the timing of tracking within a particular education system. As a consequence, an analysis within a country or state is often not feasible. Pupils in a specific education system are generally tracked at a predetermined moment in their education career, at a particular age. Secondly, when using cross-country variation in tracking it is hard to control for other differences between countries.[4] It is difficult to disentangle the effect of tracking from the impact of other differences between education systems.

Ariga and Brunello (2007) study the effect of the experienced tracking length on performance of young adults in a standardized cognitive test. They use cross-country data and distinguish the total years of completed education in 'years spent in a comprehensive system' and 'years spent in a tracking system'. The years spent in a tracking system depend both on exogenous between-country variation and some within-country variation caused by age variation, pupils dropping out of school and curriculum differences. Self-selection problems are addressed by an instrumental variables approach in which the years of tracking are instrumented by dummy variables for pupils that stopped school because of financial constraints, family reasons, or because of personal constraints like illness. Their instrumental variables estimates suggest that there is a positive contribution of tracking to performance.

Hanushek and Woessman (2006) use a difference-in-differences approach in a cross-country setup to identify the effect of an early tracking regime on standardized test scores. They match international primary school tests to secondary school tests and compare achievement differences in test scores across countries. An education system is defined as a tracking regime if pupils are tracked before the age at which the secondary school test is performed. They find that early tracking increases inequality, while it does not have a clear impact on achievement.[5]

[4] Another potential problem in a cross-country approach is the endogeneity of the choice for a particular tracking regime itself.

[5] Hanushek and Woessman use the PIRLS/TIMMS tests (which are executed by the IEA) as well as the PISA test scores (executed by the OECD) and assume that differences between these tests can be neglected. Recently, Jakubowski (2007) argued that the differences in design of these tests, however, may affect the results. While PISA measures achievement of pupils aged 15 (independent of their grade), PIRLS/TIMMS measures achievement in specific grades (independent of age). Jakubowski concludes that there is no evidence for early tracking to increase inequality when comparing pupils of the same age and grade.

Figlio and Page (2002) investigate the effect of tracking on achievement growth in math test scores between the 8th and 10th grade for different ability groups of students. They divide the students according to the 8th grade test score into top, middle and bottom thirds of the distribution and estimate separate regressions for each of these subsets in which they exploit variation in tracking across schools.[6] They find no significant effect in each of these regressions and interpret this as evidence that tracking does not harm the low-ability students. The results from additional two-stage least squares estimations in which tracking is instrumented by several interaction terms of variables that provide incentives for schools to track (like state academic requirements for graduation and the number of schools in the county), suggest that low ability students may actually gain from tracking.

The discussion on inequality is often related to socio-economic background. Brunello and Checchi (2007) focus on the interaction between family background an tracking. They argue that outcome variables like test scores at secondary school may substantially over-estimate or under-estimate the overall impact of tracking on schooling and labor market outcomes if these effects are temporarily or cumulate over time. Therefore they use a variety of outcome variables including educational attainment and some labor market outcomes. They use cross country data for different cohorts of pupils and focus on the impact of an interaction variable of family background with tracking.[7] They do not use a difference-in-differences approach, but include many covariates including country-by-cohort dummies and interactions between family background and potential confounding factors.[8] They find some evidence that tracking reinforces the role of family background and thereby reduces quality of opportunities with respect to educational attainment and labor market outcomes.[9]

Schuetz et al. (2005) also find that early tracking accentuates the impact of family background and increases the dispersion of student achievement. They use a similar approach in which they investigate the effect of the interaction of family background with an indicator of school tracking (for which they take the age of selection into different tracks) on test scores of pupils in over 50 countries.

[6] Strictly speaking, the US has a comprehensive school system and employs streaming within schools, which is a milder form of ability grouping. While tracking implies that students are placed in different school types, streaming implies that only particular courses are taught at different levels of complexity. Often, this distinction is ignored in the existing literature. A problem in the US studies, however, is that there does not exist a unique definition of tracking, as streaming can be applied in various ways and to different extents. Figlio and Page include tracking as a dummy for whether the principal reports that the school tracks. In addition, they do some sensitivity tests in which they use several alternative definitions of tracking.

[7] Based on some theoretical considerations Brunello and Checchi use two indicators for tracking, which are the time spent in a tracked school and the share of pupils enrolled in a vocational track. This idea underlying this last indicator is that changes in the share of enrollment affect peer group effects through changes in average ability.

[8] The inclusion of the country-by-cohort dummies makes that the direct effect of tracking cannot be estimated.

[9] They do, however, not find this effect with respect to literacy and on-the-job training. Moreover, the empirical findings are less clear for the indicator 'share in vocational education'.

These empirical findings suggest that disadvantaged pupils experience relatively more negative effects from an early tracking regime.[10] According to the OECD (2006) the early tracking system in the Netherlands is a possible determinant of the low participation in higher education among immigrants.

This paper abstracts from studying equality issues and focuses solely on efficiency effects of early tracking with respect to educational attainment. The critique of the OECD on the early tracking system in the Netherlands serves as a starting point and we address the question whether there is a detrimental effect of early tracking on enrollment in higher education.

Our empirical strategy differs from the previously discussed studies. We use a within country analysis in which we exploit a special feature of the Dutch education system. At the start of secondary education schools in the Netherlands offer different types of first grade classes. In some schools, pupils are directly tracked into categorial classes of a certain education level. Other schools offer one- or two-year combined classes, in which pupils are kept together before they are tracked in a particular school type. Hence, after leaving primary school, pupils can either be subject to an early tracking regime (when they enter a categorial class) or to a comprehensive system (when they enter a combined class). This causes variation in the timing of tracking which we exploit to analyze the effects of early tracking within one country. This approach ensures that other features of the education system are equal among all pupils.

A concern regarding our analysis is self-selection of pupils into categorial or comprehensive classes. We address this potential problem first by including a large set of covariates. Using micro data of pupils allows us to control for various individual background characteristics including test scores on arithmetic, language and information processing that proxy ability. Second, we adopt an instrumental variable approach in which we exploit exogenous variation in tracking driven by the supply of different school types in a pupil's region.

Both ordinary least-squares (OLS) and instrumental variables (IV) estimations show that early tracking has a negative impact on participation in higher education. Early tracking decreases the probability of participation in higher education with around 4%- points for the relevant group.

The structure of this paper is as follows. Section 2 discusses our empirical strategy including the way of exploiting differences in the timing of tracking and the identification issue. Section 3 describes the data and section 4 presents the main estimation results. Section 5 presents an additional analysis on a more recent time period, after a change in Dutch education system. In section 6 we check the robustness of the main results. Section 7 presents some analyses for

[10] A related, but different issue is that admission to certain tracks is not only based on perceived student ability but also on socio-economic background characteristics like ethnicity (Driessen, 2006). This would imply that disadvantaged people are not offered the same opportunities a priori.

pupils with differing school advices to provide a view on the total efficiency consequences of early tracking. Finally, section 8 concludes and discusses potential policy implications.

2 Empirical Strategy

2.1 Variation in the timing of tracking in Dutch education

In order to analyze the effects of early tracking, we would like to compare the outcomes of pupils that are subject to an early tracking regime with the outcomes of those who are not. Although from an international perspective Dutch educational system as a whole is considered to be an early tracking regime,[11] there exists some variation in the timing at which pupils are placed in a certain track. At the start of secondary education, schools in the Netherlands offer different types of first-grade classes. In some schools, pupils of age twelve are directly tracked into categorial classes of a certain education level. Other schools offer one- or two-year combined classes, in which pupils are kept together before they are tracked in a particular school type. Hence, pupils entering secondary education in a combined class postpone their choice of school type with at least one year. This difference causes some variation in the timing of tracking, which we can exploit to analyze effects of early tracking by comparing educational outcomes of pupils who enter secondary education in a categorial class (the 'tracked' pupils) to those who enter in a combined class (the 'non-tracked'). Strictly speaking, we investigate the effect of being tracked one or two years earlier.

Secondary education in the Netherlands consisted of four tracks in 1989: 'lbo' (pre-vocational secondary education), 'mavo' (lower general secondary education), 'havo' (higher general secondary education) and 'vwo' (pre-university education). In 1999, 'lbo' and 'mavo' were combined into one track called 'vmbo'. Since our analyses will primarily concentrate on the cohorts of 1989 and 1993, we choose to use the terms 'lbo' and 'mavo' instead of 'vmbo' from now on.

Higher education in the Netherlands is offered in two types of institutions: research universities, which offer research oriented programmes ('wo') and universities of applied sciences, which offer programmes of higher professional education ('hbo') which prepare students for particular professions.

The minimum access requirements for higher education are a havo degree or a level 4 degree of 'mbo' (upper secondary vocational education).[12] Hence, pupils who have completed havo or vwo are qualified for access to higher education, while pupils with a mavo degree can not directly enroll.

[11] See for example Hanushek and Woessman (2006).

[12] More specifically, for access to wo students are required to have a vwo degree or to have completed the first year of an hbo programme. For access to hbo, students are required to have a havo degree or to have a (level 4) mbo degree. Mbo programmes are offered in four levels (1 to 4). Only completion of the programme at level 4 qualifies pupils for access to hbo.

We interpret all pupils who enter education in a stream that does not give direct access to higher education (mavo and lower) as being assigned to a low track. Similarly, all pupils who enter secondary education in a havo or higher are assigned to a high track. Both groups of pupils can be considered to be subject to an 'early tracking regime'.
The pupils who enter some combined class that consists of at least mavo-havo postpone their choice of being assigned to either a low or a high track.[13] Those can be considered to be subject to a 'comprehensive system'. The comprehensive classes keep pupils together for an additional one or two years.

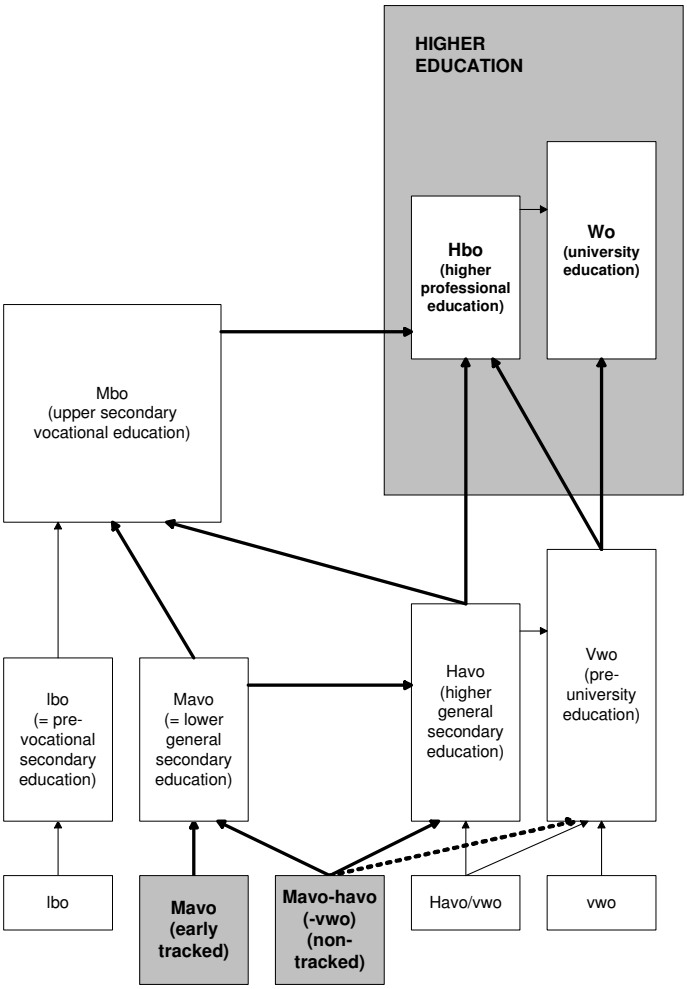
In our analysis, we primarily focus on pupils that leave primary education with a 'mavo' advice and compare educational outcomes for those who are immediately tracked in a categorial mavo to those who postpone their choice by entering a combined mavo-havo or mavo-havo-vwo class.[14] [15] Hence, we select a homogeneous group with respect to ability and compare the subsample of this group that flows into a low track to the subsample that flows into a comprehensive class. This group seems most interesting considering the critique of the OESO as the former subsample does not qualify for direct access to higher education. If early tracking would have a detrimental effect on participation in higher education, the mavo advice pupils who are tracked early are likely to be the group who experience the most negative effects of this.[16] Figure 2.1 graphically summarizes the above mentioned flow opportunities from the beginning of Dutch secondary education, that is, starting at age 12.

[13] It does not matter whether this is a mavo-havo, mavo-havo-vwo or another combined class that includes both mavo and havo. As long as the choice for both the low track and the high track is open it is considered to be a comprehensive class.
[14] At the end of primary school pupils get an advice which secondary school type they can best enroll in. This advice is not binding, but it is a strong recommendation taken seriously by the secondary schools. It can be interpreted as a proxy for perceived ability.
[15] In the period 1997-2004 comprehensive schools received extra funding from the government in order to stimulate scale increase. We have to be aware that differences in the amount of available resources (which is essentially not related to early tracking) might play a role in a comparison between pupils in categorial mavo and a mavo-havo-(vwo) class. However, this potential difference is not likely to have a large impact on our analyses. First of all, the extra funding has not affected the first two cohorts of 1989 and 1993. Secondly, available economic literature shows that additional resources do not guarantee better educational performance (see for example Leuven et al. (2007)).
[16] Hence we choose to use one advice group and, given this advice group, to take into account specific class types. Pupils that share the same advice after primary school can be considered to be homogeneous with respect to perceived ability. Taking into account more advice groups simultaneously in the analysis increases the risk of omitted variable bias as pupils with different advices are more likely to differ in unobserved characteristics. Therefore, we choose to focus our analysis on one advice group. The mavo advice group seems to be the most interesting group which potentially benefits most in terms of an increase in participation in higher education by implementing a comprehensive system.

Given the homogeneous advice group we selected, we can choose which class types to include in the 'low track' and the 'comprehensive system'. Again to decrease the risk of endogeneity problems we choose to include class types which are very much in accordance with the advice, as unobserved characteristics may be behind the choice of a pupil to enter a certain education level that is not in line with his ability.

Therefore, we consider the mavo advice group that is either tracked in a categorial mavo or enters a comprehensive mavo-havo or mavo-havo-vwo class. This already captures a very large part (around 85-90%) of the mavo-advice sample.

Figure 2.1 **Flow chart of Dutch education system: routes towards higher education for tracked versus non-tracked pupils in first year of secondary education.**



2.2 Identification

To analyze the effect of early tracking we estimate the following equation:

$$Y_i = \alpha X_i + \beta T_i + \varepsilon_i , \qquad (2.1)$$

where Y_i is the outcome variable of interest, X_i denotes a vector of background characteristics, T_i is a dummy variable which indicates whether a pupil is early tracked or not and ε_i is the error term. T_i takes value 1 if a pupil enters secondary education in a categorial mavo and takes value 0 if it enters a combined class. The parameter of interest is β.

In our analysis we consider two educational outcomes. First, the essential outcome of interest is participation in higher education. Second, we investigate the effect on completion of higher education. For these analyses we use dummy variables indicating whether a pupil has once entered or completed higher education.

Estimation of equation (2.1) by OLS may provide biased and inconsistent estimates if the error term is correlated with tracking. Hence, in case of unobserved heterogeneity we can no longer interpret β as the causal effect of early tracking. We address this problem in several ways. First, by restricting our estimation sample to the group of pupils that leave primary education with a mavo-advice, we select a homogeneous group of pupils with respect to ability which reduces potential endogeneity problems. Second, we include a large set of individual control variables like personal and socioeconomic background characteristics, which decreases the likelihood of an omitted variable bias. In addition, we are able to control for differences in ability by including test scores of the pupils.

Nevertheless, it is conceivable that unobservables exist that are correlated both with early tracking and the outcome variable. Motivated parents, for example, may rather have their child start secondary education in a combined class which gives better perspective to flow into a more advanced type of secondary education (havo or vwo) later on.

To solve for this potential endogeneity problem, we use an instrumental variables (IV) technique. The instrument we use is the relative supply of categorial schools in particular municipality types. The underlying idea is to exploit exogenous variation in tracking caused by regional differences in the supply of combined and tracked classes.[17]

The municipalities of residence of the pupils are classified in 12 categories based on a number of characteristics including the total number of residents and the percentage of the population active in agriculture. For each of these 12 categories the total number of 'tracked' and 'non-tracked' schools is known. From this we calculate a supply-ratio of tracked schools which is defined as the total number of categorial secondary schools divided by the total number of schools in that type of municipality. We use the supply-ratio as an instrument for early tracking and estimate equation (2.1) by two-stage least squares (2sls). The first stage, in which early tracking is regressed on the supply-ratio and all covariates, is

$$T_i = \gamma X_i + \delta S_i + u_i \,, \tag{2.2}$$

where S_i denotes the relative supply-ratio of categorial schools in the municipality of residence type of pupil i and u_i is the error term.

[17] Ideally, we would have had individual data on the supply of both categorial and combined classes in a pupil's neighborhood and used some distance measure (for example the distance between the nearest categorial and the nearest combined class) as an instrument. As these data are not available, we use more aggregated data on the supply of categorial and combined classes in different municipality types.

The supply-ratio is a legitimate instrument if it is correlated with early tracking and not with enrollment in (or completion of) higher education. Hence, the effect of early tracking is identified on the assumption that the supply-ratio does only affect the outcome variable of interest through early tracking. To further reduce the risk of an endogeneity bias we include an urbanization indicator in the analysis to control for possible effects of the municipality of residence on the outcome variable. This urbanization indicator is an additional measure for municipality of residence type.[18]

[18] In contrast to the classification in 12 categories, this indicator is only based on the number of residents.

3 Data description

In our empirical analysis, we use data from the Secondary Education Pupil Cohorts (VOCL) 1989, 1993 and 1999.[19] These are longitudinal data collected by Statistics Netherlands and GION.[20] Each cohort consists of a representative sample of around 20,000 pupils that enroll in the first grade of secondary education in a particular year.[21] These pupils are followed during their school careers until they leave the education system. Hence, for each pupil in each calendar year its corresponding school type and grade is known.

Approximately halfway the first year of secondary education, the pupils do tests in arithmetic, language and information processing. These so-called 'entrance tests' are comparable to the CITO test and results serve as an indicator for ability of the pupils.[22] Furthermore, the parents of the pupils were asked to fill in questionnaires which include questions on ethnicity, education and profession level of the parents, and family composition. This provides a large set of personal and socio-economic background characteristics.

The public files of the Secondary Education Pupil Cohorts follow the pupils until the school year 2003-2004. As the measurement of the final educational attainment is the key to our analysis, it is important to dispose of more recent years too. This makes it most plausible that for all pupils final educational attainment is observed. For this purpose, we received access to the private files of Statistics Netherlands with the most recent available data until the school year 2007-2008.

Moreover, we extended our data on participation in higher education with one additional school year (2007-2008) by matching it with the CRIHO data, which are register data on enrollment in higher education.[23] This provided us with the most up-to-date information on enrollment.

A comparison of our sample with the population shows that the shares of pupils that enter (around 40%) and complete (around 30%) higher education in the cohorts 1989 and 1993 are very much in line with the nationwide population figures.[24]

[19] Because of the change in education system, we discuss both the data and analyses of the 1999 cohort (after the introduction of vmbo) separately in section 5.

[20] GION is an institute for research on education that is part of the University of Groningen.

[21] For more details on representativeness of the sample we refer to Statistics Netherlands (1991) and Driessen and Van der Werf (1992).

[22] The CITO test is a nationwide test pupils do at the end of primary school which is used for the assignment of pupils into secondary schools. See section 6.1 for a discussion on potential problems in using these entrance tests as proxy for ability and their implications for the estimation results.

[23] In Dutch CRIHO is an abbreviation of 'Centraal Register Inschrijvingen Hoger Onderwijs'. We also used this data from CRIHO to check our VOCL data on enrollment in higher education. Both data sets turned out to coincide very well.

[24] In the Dutch Labor Force Survey 23.6% of the students aged 24 completed higher education in 2001. Furthermore, the share of pupils aged 25-34 with a tertiary education degree is 34% in 2004, while the share of the population aged 25-65 with a tertiary degree is 29% (CPB, 2007).

In 1989, the sample of pupils that leave primary education with a mavo-advice consists of 4912 pupils, of which 3123 (63.6 %) enter secondary education in a categorial mavo class and 1190 (24.2 %) in combined mavo-havo or mavo-havo-vwo class.[25]

The descriptive statistics of the variables we use in our analysis are shown in table 3.1. This table reports the descriptives for the estimation sample consisting of the pupils for which all covariates are available. The first column reports the sample means for the group of pupils that enter secondary education in a tracked class, while the second column reports them for the group of pupils that enter secondary education in a comprehensive class. The last column reports the *p*-value of the difference, calculated using a two-tailed *t*-test or a chi-squared test. Scores on each of the tests are between 0 (minimum score) and 20 (maximum score). The highest education level of parents, highest profession level of parents, ethnicity, household composition and urbanization of the city of residence, are all categorical variables.

There appears to be no significant difference in test scores between the tracked and the non-tracked pupils. Hence, both groups of pupils seem to be comparable with respect to ability. Furthermore, no significant differences in gender and age are observed. The parents of tracked pupils have a significant lower level of education and a lower profession level. Furthermore, significant differences are observed for urbanization degree, ethnicity and family composition.[26]

[25] Since we do not observe all classes a school offers, we cannot make an explicit distinction between a categorial mavo school that only offers mavo classes and a comprehensive school that also offers a categorial mavo track.

[26] An additional analysis in which we regress tracking on all covariates simultaneously, yields no significant effects. This suggests that the covariates are very much correlated.

Table 3.1 Descriptive statistics for mavo advice estimation sample, cohort 1989, early tracked (mavo) versus later tracked (mavo-havo)-vwo)[a]

Variables	Early tracked: mavo	Non- tracked: mavo-havo (-vwo)	P-value of difference
Ability			
Test score arithmetic[b]	10.6	10.6	0.864
Test score language[b]	11.7	11.7	0.620
Test score information processing[b]	11.7	11.7	0.983
Personal and SES variables			
Female	56	55	0.574
Age	12.6	12.6	0.410
Highest education level parents			0.000
No primary education	1	2	
Primary education	12	13	
Secondary education low	30	24	
Secondary education high	41	39	
Higher education first phase	14	17	
Higher education second phase	2	6	
Higher education third phase	0	0	
Profession level parents			0.000
Worker	29	22	
Self-employed without personnel	5	5	
Self-employed with personnel	5	4	
Lower employee	11	12	
Intermediate employee	22	21	
Higher profession	12	16	
Other	17	19	
Ethnicity			0.000
The Netherlands	89	81	
Marocco	1	1	
Antilles, Surinam, Aruba	2	5	
Turkey	2	2	
Other	7	10	
Family Composition			0.000
None of parents	0	1	
Only mother	1	8	
Mother and others	2	2	
Only father	1	1	
Father and others	0	0	
Father and mother	87	77	
Father and mother and others	0	1	
Other	5	11	

Table 3.2 Descriptive statistics for mavo advice estimation sample, cohort 1989, early tracked (mavo) versus later tracked (mavo-havo)-vwo)[a] (continued)

Variables	Early tracked: mavo	Non-tracked: mavo-havo (-vwo)	P-value of difference
Urbanization city of residence			0.000
Very high	9	18	
High	16	30	
Median	19	13	
Modest	29	23	
Low	26	17	
Education outcomes			
Ever participated in higher education	33.4	38.3	0.005
Completed higher education	21.3	26.8	0.001
Time to enrolment in higher education in years	7.94	7.68	0.030
Number of pupils	2905	1031	

[a] All numbers represent percentages, unless stated otherwise.
[b] Test scores are of entrance tests taken in first year of secondary education. The maximum score for each test is 20.

Both enrollment and completion of higher education are significantly lower for tracked pupils. This lower participation in higher education among the tracked pupils can be (partly) traced back to differences in flow opportunities after the first year of secondary education. Pupils that entered secondary education in a comprehensive class still have an opportunity to move on to a higher track (i.e. havo or vwo) in subsequent years, which provides direct access to higher education. Table 3.2 gives an overview of the education careers in secondary education of both tracked and non-tracked pupils.[27] We indeed observe that the non-tracked pupils are more likely to move on directly to the higher tracks. In the second cohort year, for example, participation in havo or vwo is 14 percentage points higher for the non-tracked students. In the third and fourth cohort years, this difference increases to 18 percentage points. Later on, these differences in education careers translate into differences in higher education enrollment in favor of non-tracked students.

[27] Similar overviews of the education careers in the other cohorts can be found in appendix A.

Table 3.2 **Routes through education in first nine years after entering secondary education, tracked (=mavo) versus non-tracked (=mavo/havo or mavo/havo/vwo); mavo-advice estimation sample, cohort 1989**

	< Mavo	Mavo	Comprehen-sive: (l)mh or (l)mhv	Havo	Havo/vwo	Vwo	Mbo	Higher education	Left [a]
Year 1									
Non-tracked	0	0	100	0	0	0	0	0	0
Tracked	0	100	0	0	0	0	0	0	0
Year 2									
Non-tracked	8	42	32	7	9	0	0	0	1
Tracked	3	94	0	1	1	0	0	0	0
Year 3									
Non-tracked	11	58	9	12	5	3	0	0	2
Tracked	7	90	0	2	0	0	0	0	1
Year 4									
Non-tracked	14	61	1	16	1	3	0	0	4
Tracked	9	87	0	2	0	0	0	0	2
Year 5									
Non-tracked	9	24	0	23	0	3	31	0	10
Tracked	6	21	0	12	0	0	52	0	9
Year 6									
Non-tracked	2	3	0	16	0	4	49	3	22
Tracked	2	2	0	11	0	1	64	0	20
Year 7									
Non-tracked	0	0	0	6	0	3	50	9	31
Tracked	0	0	0	3	0	1	62	5	29
Year 8									
Non-tracked	0	0	0	1	0	0	32	20	47
Tracked	0	0	0	0	0	0	37	16	47
Year 9									
Non-tracked	0	0	0	0	0	0	16	27	57
Tracked	0	0	0	0	0	0	16	22	62

[a] The percentages shown in this column seem to overstate real shares of people that left education. This is because the cohort may have lost track of some students that moved to other schools (in other cities) or that have temporarily left education and returned after some time. This also explains why the reported percentages active in higher education are somewhat lower than the ones we found after having matched the students from the VOCL cohorts with a central register of all higher education participants over time.

In addition, early tracked pupils need more time to enter higher education. This finding is intuitive as a comprehensive class provides an institutionally shorter way to higher education. Table 3.3 indeed shows that early tracked pupils more often take the longer route via mbo towards higher education, whereas non-tracked students predominantly take the shorter route via havo.[28]

[28] Appendix A shows similar patterns of education careers for the other cohorts.

Table 3.3 Routes towards higher education: percentages of last completed education level before entering higher education (cohort 1989)

Education level	Early tracked: mavo	Non-tracked: mavo-havo(-vwo)
Mbo	61	42
Havo	26	36
Vwo	4	12
Other	9	10

The descriptive statistics for the estimation sample of the cohorts 1993 can be found in table B.1 of appendix B.

In 1993, the sample of pupils that leave primary education with a mavo-advice consists of 4367 pupils, of which 2214 (50.1 %) enter secondary education in a categorial mavo class and 1373 (31.4 %) in a combined mavo-havo or mavo-havo-vwo class. The covariates are similar to those in the 1989 cohort, except for the variables 'family composition' and 'ethnicity' which are not available. As a proxy for ethnicity we use the variable 'father's country of birth' instead.[29]

In contrast to the 1989 cohort, the results suggest a difference in ability between the tracked and non-tracked pupils. In 1993 pupils in categorial mavo do significantly worse on all test scores. In addition, significant differences are observed for ethnicity and urbanization. Differences with respect to the education and profession level of the parents are not (clearly) observed now. The early tracked enroll significantly less in higher education and a lower fraction of the pupils completes higher education.

[29] In the 1989 cohort 'ethnicity' was based both the parents' and the pupil's country of birth.

4 Results

In this section, we first present the estimation results of the first-stage regressions and discuss the impact of the relative supply-ratio of categorial schools on school type choice. In particular, we investigate whether the estimations give rise to concern for weak instruments. After that we present our main results on the effects of early tracking for all cohorts, which include both OLS and IV estimations.

4.1 The effect of regional supply on school type choice

The regional supply of school types may cause exogenous variation in actual school type choice. Table 4.1 presents the supply-ratios of categorial schools for each of the municipality types. The first column reports the twelve different municipality types. The second and third column report the supply-ratios for each of these municipality types for the cohorts 1989 and 1993, respectively.[30] There are no supply-ratios reported for the municipality types 'countryside A1' and 'countryside A2', as we observe no schools in our estimation sample for these regions.

In 1989, the relative supply-ratios are substantially smaller in the larger cities and specific commuter municipalities, compared to the countryside, rural and small cities. This implies that pupils living in these last types of municipalities are more restricted in their choice for combined classes. In 1993, the supply-ratios are lower for most municipality types compared to those in 1989. The average nationwide supply-ratio of categorial schools has decreased from 0.75 to 0.57 between 1989 and 1993.

As discussed before, we use the supply-ratio as an instrument for early tracking and estimate equation (2.1) by two-stage least squares. Table 4.2 presents the estimation results of regression (2.2) for different cohort samples. The robust standard error, corrected for clustering at the school level, is shown below each estimate in parentheses. The number of observations is denoted by N.

A well-known concern in using the IV-approach is the problem of weak instruments.[31] Staiger and Stock (1997) proposed to use a cut-off value of 10 for the F-value of the excluded instrument. If the F-value in the first-stage is above this cut-off value, correlation of the instrument with the endogenous variable is considered to be sufficiently strong.

[30] The supply-ratio is defined as the number of categorial schools divided by the total number of schools in each municipality type. See appendix C for an overview of the number of categorial and combined schools in our estimation sample for each of the municipality types in all cohorts.

[31] Bound, Jaeger and Baker (1995) address the finite sample problem of weak instruments, in which case IV estimates are biased towards OLS even in fairly large samples.

Table 4.1 Supply-ratios of categorial schools

Municipality type	1989	1993
Countryside A1		
Countryside A2		
Countryside A3	0.80	0.75
Countryside A4	1.00	0.57
Urbanized countryside B1	0.92	0.90
Urbanized countryside B2	0.85	0.77
Specific commuter municipality B3	0.59	0.60
Rural cities C1	0.88	0.78
Small cities C2	0.85	0.44
Medium-sized cities C3	0.56	0.58
Medium-sized cities C4	0.52	0.37
Big cities C5	0.74	0.41

Table 4.2 The impact of the relative supply-ratio on early tracking

	1989	1993	Pooled sample 1989-1993
Relative supply ratio	0.878***	0.324	0.589***
	(0.255)	(0.258)	(0.151)
Socio-economic status variables (SES)	yes	yes	yes
Test scores	yes	yes	yes
N	3936	2954	6577

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

In 1989, we find an F-value of 11.9. However, the F-value of the effect of the supply-ratio on early tracking equals 1.58 in 1993, which raises concerns for weak instruments. Apparently, the differences in relative supply of categorial classes between different types of municipalities do not explain much variation in early tracking in 1993. A possible explanation for this may be the overall decrease in relative supply of categorial schools over time.[32] Whereas pupils in 1989 had limited choice of entering a combined class in particular regions, pupils in 1993 faced less restrictions in choosing class type. The weak statistical relationship between supply-ratio and school choice indicates that we may have weak instruments for the cohort 1993, and hence we will not pursue with this approach for this cohort.

In a pooled sample of the 1989 and 1993 cohort, the F-value equals 15.3. In order to obtain a pooled sample in which identical control variables are available for all individuals, we had to make two adjustments to the 1989 cohort. First, we left family composition out of the set of covariates as this variable is not available for the 1993 cohort. Second, in the 1993 cohort we

[32] In the early nineties, there was a tendency to scale increase in Dutch schools which resulted in larger schools offering various education levels and presumably more first-grade combined classes.

only observe the country of birth of the father. Hence, we have substituted the variable 'ethnicity' for the variable 'country of birth of the father', which is also available in the 1989 cohort.[33] Apparently, pooling the data of the cohorts 1989 and 1993 decreases standard errors sufficiently to obtain a fairly good first-stage.

4.2 The effects of early tracking

Tables 4.3 and 4.4 present the main estimation results. Table 4.3 reports the estimation results of various regression analyses on participation in higher education for different cohort samples. The first model regresses a dummy variable for participation in higher education on only a constant and early tracking. In the second model personal and socio-economic control variables like age, gender, ethnicity, education and profession level of parents and family composition (if available) are included, while the third model additionally includes test scores.[34] The fourth model specification is identical to the third model and is estimated by 2sls, using the supply-ratio of categorial classes as an instrument for early tracking.
The first and second rows present the results for the cohorts 1989 and 1993, respectively.
The third row shows the estimation results for the pooled sample of the cohorts 1989 and 1993.

Table 4.4 is similarly structured and presents the estimation results of the regression analyses on completion of higher education.
In the presentation of the results, we only show the estimated coefficients for early tracking.[35] As discussed above, we do not present the IV results for the 1993 cohort, as we have serious concern for weak instruments.

[33] As there are more missings on this variable in 1989 compared to ethnicity, the total number of observations in the pooled cohort is smaller than the sum of observations of the two separate cohorts.

[34] In addition to test scores which proxy unobserved ability, we dispose of a variable that proxies unobserved motivation. This is a measure for motivation on a scale from 1 to 5, based on a questionnaire pupils fill in after a couple of months in the first grade of secondary education. Nevertheless, we have chosen not to include this variable in our main specifications as we have serious concern that this variable depends on school choice. As the questionnaires were not administered before the start of the school year, the variable can be very well affected by the particular school and class, as well as the choice of tracking. Inclusion of this variable in the set of covariates turns out to hardly change the estimated effect of tracking.

[35] The estimated coefficients for other covariates are available upon request.

Table 4.3 The impact of early tracking on participation in higher education

	(1) OLS	(2) OLS	(3) OLS	(4) IV
1989	− 0.049**	− 0.042*	− 0.034	− 0.111
	(0.023)	(0.022)	(0.023)	(0.084)
	(N=3936)	(N=3936)	(N=3936)	(N=3936)
1993	− 0.067***	− 0.057***	− 0.044**	
	(0.023)	(0.021)	(0.020)	
	(N=2954)	(N=2954)	(N=2954)	
Pooled sample 1989-1993	− 0.061***	− 0.052***	− 0.042***	− 0.298***
	(0.016)	(0.019)	(0.016)	(0.123)
	(N=6577)	(N=6577)	(N=6577)	(N=6577)
SES	no	yes	yes	yes
Test Score	no	no	yes	yes

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

Table 4.4 The impact of early tracking on completion of higher education

	(1) OLS	(2) OLS	(3) OLS	(4) IV
1989	− 0.054***	− 0.051***	− 0.045**	− 0.129*
	(0.019)	(0.018)	(0.018)	(0.068)
	(N=3936)	(N=3936)	(N=3936)	(N=3936)
1993	− 0.050**	− 0.044**	− 0.035*	
	(0.021)	(0.020)	(0.020)	
	(N=2954)	(N=2954)	(N=2954)	
Pooled sample 1989-1993	− 0.056***	− 0.049***	− 0.043***	− 0.159*
	(0.015)	(0.014)	(0.014)	(0.082)
	(N=6577)	(N=6577)	(N=6577)	(N=6577)
SES	no	yes	yes	yes
Test Score	no	no	yes	yes

* Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level. Robust standard errors in parentheses.

The OLS analyses all yield a negative effect of early tracking on participation in higher education. The estimated coefficients decrease (in absolute value) when additional control variables are included in the model. The IV estimations also yield negative, but larger (absolute) point estimates. Except for the full model OLS and IV estimations in 1989, all estimated coefficients are statistically significant.

The estimated coefficients of the cohorts 1989 and 1993 are reasonably in line which each other. To contribute to the precision of out estimations we also perform analyses on the pooled sample of the 1989 and the 1993 cohort pupils. As discussed before, we had to make some adjustments to the 1989 cohort to obtain identical control variables for all pupils. In addition, we included a dummy variable for being in the 1993 cohort in all models.[36] The results are similar to the basic estimations, but have smaller standard errors. The analyses on the pooled model all yield significant negative effects of early tracking.

The interpretation of the IV estimates differs from that of the OLS estimates. Whereas the OLS estimations give the average treatment effect (ATE), the estimated IV coefficients should be interpreted as the ATE for the subpopulation of pupils whose choice for being early tracked or not is affected by the instrument.[37] In our case, this is the subpopulation that is more likely to choose the comprehensive class if the supply of comprehensive classes increases. Hence, the IV effect is only applicable to the pupils whose choice of class type is affected if the supply of comprehensive classes in their region changes. Although the interpretations differ, both OLS and IV estimates clearly point to the same direction. Hence, we consider the OLS estimates to be supported by the IV estimates.

The analyses on completion of higher education also yield negative and significant effects of early tracking. The estimated effects on enrollment and on completion of higher education are almost identical.[38] The IV estimations on completion of higher education again yield larger (absolute) effects of early tracking.

To summarize, all presented estimation results yield a negative sign for early tracking and (besides the OLS and IV estimates on participation for the 1989 cohort) are statistically significant. We find that early tracking decreases the probability of both enrollment and completion of higher education by around 4%-points. The OLS estimates in the 1989 cohort and the pooled cohort are supported by the IV estimates. These results provide evidence of a detrimental effect of early tracking on higher educational outcomes.

[36] This estimated coefficient of this dummy variable is close to zero and insignificant.

[37] This is the local average treatment effect (LATE) which does not automatically apply to a random chosen pupil (Imbens and Angrist, 1994). Stated differently, it is the effect of the treatment on the subpopulation of compliers, who have changed their behavior as a result of the instrument. The local average treatment effect (LATE) can differ substantially from the average treatment effect (ATE) if the fraction of the population that is affected is small (Oreopoulous, 2006).

[38] Hence, early tracking decreases both enrollment in and completion of higher education by around 4 percentage points. This implies that, measured in percentages, early tracking has a larger negative impact on completion of higher education.

5 Effects after the introduction of vmbo

In 1999, 'lbo' and 'mavo' were combined into one track called 'vmbo'. Vmbo (preparatory middle-level vocational education) lasts four years and combines vocational with theoretical education. Vmbo itself consists of four levels, in each of which a different mix of practical vocational education training and theoretical education is combined. The track vmbo-tl ('theoretical learning path') is the most theoretical of the four and best comparable to the former mavo. This is the only level which provides direct access to havo.

The change in the education system has led us to treat the 1999 cohort separately from the other ones. The introduction of vmbo has caused quite some public debate, which may have contributed to a negative image of vmbo schools. As a consequence, we expect that selectivity has increased in the 1999 cohort. This section presents regression analyses for the 1999 cohort to provide some insight in the effect of early tracking in the current education system in the Netherlands.

The estimation sample for the cohort 1999 differs in some respects from the other two cohorts. In order to stay as close as possible to the analyses in the other cohorts, we selected pupils who leave primary education with a mavo or a vmbo-tl advice.[39] This group consists of 2452 pupils of which 1148 (46.8%) enter secondary education in vmbo (the 'tracked') and 1268 (51.7%) enter in a combined vmbo-havo or vmbo-havo-vwo class (the 'non-tracked').

In 1999, there is no variable for the profession level of the parents available. With respect to education level of the parents, we now dispose of two variables, one for the mother and one for the father. The descriptive statistics for the estimation samples of the cohort 1999 can be found in table B.2 of appendix B. There is a significant difference between the tracked and non-tracked pupils with respect to ability: the non-tracked pupils perform better on all test scores. Furthermore, there appears to be a significant difference in age, highest education level of both father and mother, ethnicity and urbanization degree.

As we observe the pupils up to the school year 2007-2008, we measure enrollment in higher education after nine years of education. At that time, enrollment is significantly higher for the non-tracked pupils. As none of the pupils has completed higher education after nine years, the time span over which we observe the pupils is too short to obtain information on this variable.

Table 5.1 reports the OLS estimation results of several regression analyses on participation in higher education.[40] The first model regresses a dummy variable for participation in higher education on only a constant and early tracking. In the second model, the available personal and

[39] As vmbo was to start in August 1999, some schools still used the 'old' mavo advices while others used the 'new' vmbo-tl advices at the time pupils left primary education.

[40] We do not carry out IV analyses for the 1999 cohort as, similarly to the 1993 cohort, there is serious concern for weak instruments. The average nationwide supply-ratio of categorial schools is 0.54 and the F-value of the effect of the supply-ratio on tracking in the first-stage regression is 4.61 in 1999.

socio-economic control variables are included, while the third model additionally includes test scores.

Table 5.1 The impact of early tracking on participation in higher education (cohort 1999)

	(1) OLS	(2) OLS	(3) OLS
Early Tracking	– 0.126***	– 0.112***	– 0.092***
	(0.034)	(0.034)	(0.029)
SES	no	no	yes
Test score	no	yes	yes
N	2153	2153	2153

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

The OLS analyses yield significant negative effects of early tracking on participation in higher education, which is in line with our previous findings. The estimated coefficient of early tracking is larger in absolute value compared to the 1989 and 1993 cohorts. It is, however, important to stress that the we cannot directly interpret this difference in magnitude of the effect as the results of the 1999 cohort are not easily comparable to those in the other cohorts. After all, we can think of many potential confounders originating from the change in the education system that may have affected the estimated coefficient.

First of all, the impact of the mechanisms underlying the effect of early tracking may have changed by the introduction of vmbo, for example because of differences in first-grade class composition. Our available data do not distinguish between vmbo-tl and other vmbo pupils and hence we do not observe the composition of first-grade classes in vmbo. However, it is well conceivable that there is more variation in ability compared to a categorial mavo, which may imply different peer effects or different opportunities for teachers to focus course material to all pupils. Besides, effects from teacher sorting may have changed if teacher preferences for teaching a vmbo class differ from teaching a categorial mavo class.

Secondly, it may be that the change in the education system has caused different opportunities to flow through various education levels. After the introduction of vmbo it may have become more difficult to switch to havo, either because of institutional reasons or because the education content at vmbo fits worse to havo compared to the education content at the former mavo.

Thirdly, there may have been some behavioral effects after the change of the education system. It might be that advices at primary school are somewhat adjusted, more directed at a clear recommendation for either a vmbo school or not. Moreover, pupils and their parents' motives for choosing a tracked or comprehensive class may have changed, because of a different image of vmbo compared to the former mavo.

Apart from the difficulties in comparing the results in the 1999 cohort to those in the other cohorts, we have to be careful with the estimation results themselves as the 1999 cohort is less suitable for studying the effects of early tracking in general. It is well conceivable that the introduction of vmbo has increased self-selection into either a vmbo track or a comprehensive class. The descriptive statistics indeed show that tracked and non-tracked pupils differ significantly on all covariates, including the cognitive test scores. There seems to be an increase over time in observable differences between tracked and non-tracked pupils which may also indicate an increase in unobservable differences. Because of this apparently larger self-selection problem, and the lack of a strong instrument that is needed to use an instrumental variables method to address this problem, the estimated effect may be biased.

In addition, as we only observe enrollment after nine years, this leaves us with less reliable estimates compared to the other cohorts, in which we observe the pupils over a longer time span.

Therefore, although this analysis may be most interesting from a policy perspective as it focuses on the most actual situation in Dutch education, it provides less reliable information about the effect the of postponement of track assignment in general. Consequently, the estimation results, which suggest a negative effect of early tracking for the vmbo-tl pupils, should be treated with some reservation.

6 Robustness checks

This section discusses some additional analyses to examine the robustness of the main results of this paper. We concentrate on two issues.

First, we address the role of the entrance test as proxy for ability including its potential drawbacks and implications for the estimated effects.

Second, we investigate the sensitivity of the results to the definitions of a low track and a comprehensive class we used in our analysis. In order to do this, we present additional regression analyses in which we vary the class types included in the low track and comprehensive classes.

6.1 The role of entrance tests as proxy for ability

In our main estimations, we use entrance test scores on arithmetic, language and information processing as proxy for unobserved ability. These tests are short versions of the CITO test pupils make around halfway the first year of secondary education.
A first criticism on this measure may be that it is not a high-stake test like the real CITO test which is explicitly used to assign pupils to different levels of secondary education. A second criticism may be that the tests are not made before the start of secondary education, which would rule out that test scores already pick up some of the learning effect in secondary education. Because entrance tests are made around halfway the first year, we might conjecture that our estimated coefficients of early tracking would be a lower bound (in absolute value) of the true effect.

CITO test scores of all pupils at the end of primary education would not suffer from these drawbacks, and hence would be a better measure for ability.

Unlike the other cohorts, the cohort 1999 provides us with information on individual CITO test scores for around two third of the sample. Within our estimation sample of the 1999 cohort, the CITO test scores are available for 1382 pupils. We can use this information to obtain some insight in the implications of using entrance test scores instead of CITO scores for the estimated effects. Table 6.1 presents the average total score on entrance tests (which is the sum of the test scores on arithmetic, language and information processing) and the average CITO test scores for this sample.

Table 6.1 Entrance and CITO test scores (cohort 1999)

	Vmbo	Vmbo-havo(-vwo)	p-value of difference
Total score entrance tests	33.5	34.9	0.00
CITO test score	532.9	533.3	0.21
N	741	641	

Both the test scores indicate that pupils in combined vmbo-havo(-vwo) classes have on average a higher ability. However, in contrast to the entrance tests, the difference in CITO test scores is not statistically significant. Hence, the difference in entrance tests might (partly) be the result of tracking.

Table 6.2 presents the estimation results of two models. In the first model, enrollment in higher education is regressed on tracking, entrance test scores as proxy for ability, and all other personal and socio-economic covariates. The second model is identically specified, except for the fact that entrance tests are now replaced by the CITO test score.

Table 6.2 The impact of early tracking on participation in higher education (cohort 1999)

	(1)	(2)
	OLS	OLS
Early tracking	– 0.036**	– 0.042**
	(0.018)	(0.018)
SES	yes	yes
Entrance Test Score	yes	no
CITO test score	no	yes
N	1382	1382

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

The estimated effect with entrance test scores included is lower (in absolute value) compared to the effect with CITO test scores included.[41]

Moreover, the estimation results in tables 4.3 and 4.4 already show that leaving out the entrance test scores in the model specification yields an even larger (absolute) effect of tracking.

These findings indicate that, if anything, our estimates provide a lower bound of the (absolute) actual effect, which confirms our conjecture. Hence, it does not seem plausible that potential problems with the entrance test scores would be behind the estimated negative effect of early tracking.

6.2 Various definitions of tracking

In the previous sections, we focused on pupils with a mavo advice that either entered secondary education in a categorial mavo (the low track) or in a combined mavo-havo or mavo-havo-vwo class (the comprehensive class).[42] Although this captures already a large fraction of the pupils

[41] In addition, point estimates for this subsample are now smaller (in absolute value) than in the full estimation sample, which can be due to the fact that this subsample of pupils has a lower average ability as can be seen from tables B.2 and 6.1. As the probability of enrollment in higher education is lower it may be that the difference in tracking is less important in this subsample.

[42] For a justification of the choice of this subsample of classes see section 2.1 including footnote 16.

with a mavo advice, it is worthful to check whether results change if we do include pupils that choose to enter other class types.
Including different types of classes in a certain track may for example change the effect because of different peer effects pupils experience.

Tables 6.3 and 6.4 present the OLS estimation results for the cohort 1989 of four models that are identically specified except for the definition of tracking and non-tracking.[43] Each of the models includes all available control variables. The definitions used in each of the models are as follows:

Model 1: tracking = mavo and comprehensive = mavo-havo, mavo-havo-vwo (the basic model)
Model 2: tracking = lbo-mavo, mavo and comprehensive = mavo-havo, mavo-havo-vwo
Model 3: tracking = lbo, lbo-mavo, mavo and comprehensive = mavo-havo, mavo-havo-vwo
Model 4: tracking = lbo, lbo-mavo, mavo and comprehensive = lbo-mavo-havo, lbo-mavo-havo-vwo, mavo-havo, mavo-havo-vwo.

Table 6.3 The impact of early tracking on participation in higher education (sensitivity analysis)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Early tracking	– 0.034	– 0.039*	– 0.044**	– 0.038*
	(0.023)	(0.022)	(0.022)	(0.022)
SES	yes	yes	yes	yes
Test score	yes	yes	yes	yes
N	3936	4205	4296	4351

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

Table 6.4 The impact of early tracking on completion of higher education (sensitivity analysis)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Early tracking	– 0.045**	– 0.048***	– 0.051***	– 0.042**
	(0.018)	(0.018)	(0.018)	(0.019)
SES	yes	yes	yes	yes
Test score	yes	yes	yes	yes
N	3936	4205	4293	4351

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

[43] For space reasons, we only show the sensitivity checks for the 1989 cohort. In the other cohorts, we obtain similar results.

The results show that the estimated effect of early tracking is fairly robust against including various class types in the definitions. Including some lower class types to the low track in the models (2) and (3) slightly enlarges the impact of tracking (in absolute value). Generally, we conclude that varying the class types does not substantially change the results.

7 Effects for pupils with higher school advices

The results presented so far suggest that lower ability pupils can substantially gain from being in a comprehensive class. It is important to note that this is the result of a partial analysis, which focuses on a single advice group and compares only assignment in a low track to assignment in a comprehensive class. These results are not sufficient to judge the efficiency of an early tracking regime as a whole. After all, higher ability pupils may be negatively affected by the lower ability ones in a comprehensive class.

This section investigates the effects of tracking for other advice groups in which we also compare allocation in a high track to allocation in a comprehensive class. Focusing on the effects of tracking for pupils with different abilities contributes to obtaining a more general view on the efficiency effects of early tracking.[44]

Firstly, we concentrate on the pupils who leave primary education with a havo advice. We use this group to analyze the effect of being in a high track compared to being in a comprehensive class. The non-tracked pupils are as usual defined as those who enter secondary education in a mavo-havo or mavo-havo-vwo class. The tracked pupils are defined as those who enter secondary education in a havo or havo-vwo class.[45] This is the mirror image of the previous analysis on the group of mavo-advice pupils, which compared assignment in a low track to assignment in a comprehensive class.[46] Table 7.1 presents the full model OLS estimations of early tracking on both participation in and completion of higher education for the 1989, 1993 and pooled cohorts. In 1989, the estimation sample consists of 1162 pupils that enter secondary education in a havo-vwo class and 608 pupils that enter in a mavo-havo or mavo-havo-vwo class. In 1993, the sample consists of 1107 pupils that enter secondary education in a havo or havo-vwo class and 521 that enter in a mavo-havo or mavo-havo-vwo class.

[44] Since we use only within-country data we have to do partial analyses in order to get a view on the total system. In a cross-country analysis in which a comprehensive system is compared to an early tracking system both countries have, roughly speaking, the same distribution of abilities of the pupils which are thus randomly divided over a tracked and a comprehensive system. However, including all pupils (which implies all advice groups and all class types) simultaneously in our analysis would substantially increase the endogeneity problem. Specific advice groups (for example lbo, vwo) may be more likely to be tracked while others are more likely to be subject to a comprehensive system. Hence, there will be self-selection of specific advice types with particular unobservables into an early tracking system which substantially increases the risk of omitted variable bias.

Apart from this, the analysis would become fuzzy which would make it very complicated to identify separate effects, which is one of the advantages of our analysis compared to a macro-analysis.

[45] We include havo-vwo in the high track because there are very few schools which offer categorial havo classes in the Netherlands. We observe no pupils in a categorial havo in 1989 and only a few in 1993.

[46] In this analysis we choose to include both the mavo-havo and mavo-havo-vwo classes as comprehensive classes. In this way we use exactly the same definition as in the analysis on the mavo advice pupils. Pupils with a mavo advice benefit from entering these classes, and this analysis investigates the effects for pupils with a havo advice that enter these classes.

The results indicate that there is no significant effect of early tracking on both participation in and completion of higher education. This suggests that pupils with a havo advice do not experience negative effects of being in a comprehensive class together with lower ability pupils.

Table 7.1 The impact of early tracking on participation in and completion of higher education (havo advice group)

	Cohort 1989	Cohort 1993	Pooled sample 1989-1993
Early tracking on participation	0.010	0.036	0.024
	(0.027)	(0.029)	(0.020)
Early tracking on completion	– 0.006	0.055	0.022
	(0.025)	(0.033)	(0.021)
SES	yes	yes	yes
Test score	yes	yes	yes
N	1770	1628	3279

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

We cannot use the instrumental variable analysis to address the potential endogeneity issue here, because there are no (severe) restrictions in the supply of high track classes. However, it is plausible to assume that potential self-selection would positively affect educational outcomes for the high track, as pupils with better unobservable characteristics are likely to self-select into the high track and may be expected to enroll more often in higher education. Hence, the estimated coefficients can reasonably be interpreted as the upper bounds of the effects of early tracking. This makes that potential endogeneity is less of a problem here as we find potentially upwards biased effects that are already statistically insignificant.

Secondly, we investigate the effects of tracking for the group of pupils with a mavo-havo advice, who are in between the havo and mavo pupils.

This group has never been tracked at an education level that coincides with perceived ability: if they are in a low stream they are tracked at a lower level and in the high stream they are tracked at a higher level. We can use this group to perform two kinds of partial analyses.

In the first analysis, we estimate the effect of being assigned to a low track (categorial mavo) compared to a comprehensive class (mavo-havo or mavo-havo-vwo). In the second analysis, we investigate the effect of being assigned to a high track (havo, havo-vwo) compared to being in a comprehensive class (mavo-havo or mavo-havo-vwo). Table 7.2 presents the OLS estimation results of both analyses for all cohort samples. The first two rows present the estimated effects of early tracking on participation in and completion of higher education for the first analysis, while the latter two rows present the estimated effects for the second analysis. In each of the models, all available control variables are included.

Table 7.2 The impact of early tracking on participation in and completion of higher education (mavo-havo advice group)

	Cohort 1989	Cohort 1993	Pooled sample 1989-1993
Early tracking on participation	– 0.131***	– 0.125**	– 0.125***
(low track vs comprehensive)	(0.048)	(0.048)	(0.035)
	(N=945)	(N=1293)	(N=2172)
Early tracking on completion	– 0.085**	– 0.080	– 0.085***
(low track vs comprehensive)	(0.042)	(0.050)	(0.034)
	(N=945)	(N=1293)	(N=2172)
Early tracking on participation	0.034	0.018	0.027
(high track vs comprehensive)	(0.033)	(0.040)	(0.026)
	(N=1068)	(N=1335)	(N=2332)
Early tracking on completion	0.023	– 0.053	– 0.017
(high track vs comprehensive)	(0.040)	(0.042)	(0.030)
	(N=1068))	(N=1335)	(N=2332)
SES	yes	yes	yes
Test score	yes	yes	yes

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Robust standard errors in parentheses.

Entering secondary education in a low track decreases the probability to enroll in (and complete) higher education, compared to entering in a comprehensive class. The estimated coefficients are statistically significant, except for the impact of tracking on completion in the 1993 cohort. These negative effects are in line with the effects we found for the mavo-advice group, although point estimates are larger in absolute value now. The mavo-havo advice pupils experience a more negative impact of being in a low track, which may be explained by the fact that they are placed in a track below their perceived ability.

In addition, it turns out that there is no significant difference whether a pupil enters the high track or a comprehensive class. This result is in line with the effect we found for the havo-advice pupils.

In conclusion, these analyses suggest that both pupils with a havo-advice and pupils with a mavo-havo advice experience no significant difference whether they are in a high track or in a comprehensive class. The pupils with a mavo-havo advice experience a negative effect of entering a low track instead of a comprehensive class.

8 Conclusions and discussion

This paper investigates the effect of early tracking with respect to participation in higher education. In our analysis, we use data from the Secondary Education Pupil Cohorts of 1989, 1993 and 1999 and exploit differences in the timing of tracking between schools in the Netherlands. To deal with potential endogeneity problems we restrict our estimation sample to a particular school advice group which is homogeneous with respect to ability, use a large set of covariates and adopt an instrumental variables approach.

Our main analysis focuses on pupils with a mavo advice, as this group seems most interesting considering a potential negative effect of early tracking. If early tracking would have a detrimental effect on participation in higher education, the mavo advice pupils who are tracked early are likely to be the group which is most negatively affected.

In the 1989 cohort, we find (mainly statistically significant) negative effects of early tracking on both participation and completion of higher education. The OLS estimates are supported by the IV estimates. Similar analyses on the 1993 and the pooled 1989-1993 sample all yield comparable significant negative effects of early tracking. These findings turn out to be robust in various specifications. The OLS estimates show that early tracking decreases the probability of both enrollment and completion of higher education by around 4%- points for pupils with a mavo advice. In the 1989 and 1993 cohorts, average enrollment and completion of higher education for the tracked pupils is around 35 and 22%, respectively. Hence, pupils with a mavo advice that are in a categorial mavo can increase their probabilities to enroll and complete higher education to around 39 and 26%, respectively, by entering secondary education in a combined mavo-havo(-vwo) class. The IV analyses yield even larger effects in absolute value. Hence, the pupils with a mavo advice perform better in a comprehensive class than in a low track.

In the 1999 cohort, we also find a negative impact of early tracking on enrollment in higher education. However, we have to be careful with these results as the 1999 cohort is less suitable for studying the effects of early tracking in general because of the shorter time span over which pupils are observed and some indications for increased selectivity.

Although we find that mavo-advice pupils can gain from being in a comprehensive class, these results are not sufficient to judge the efficiency of an early tracking regime as a whole. After all, higher ability pupils may be negatively affected by the lower ability ones in a comprehensive class. To obtain a more complete view on the impact of early tracking, we have analyzed effects for pupils with higher advices. Based on OLS estimations we conclude that pupils with a havo advice experience no significant difference whether they are in a high track or in a comprehensive class. In addition, two partial analyses on the pupils with a mavo-havo advice

suggest that they are better of in a comprehensive class compared to a low track, while there is no significant difference for them between entering a high track or a comprehensive class.

Table 8.1 summarizes the estimation results for all advice groups. While pupils with a mavo and mavo-havo advice experience a negative effect of being assigned to a low track compared to a comprehensive class, pupils with a mavo-havo and havo advice do not experience a positive effect of being tracked in a high track compared to a comprehensive class. These results suggest an inefficiency of the early tracking regime with respect to participation in and completion of higher education.

Table 8.1 Summary of results: effects of early tracking on participation in and completion of higher education

Advice Group	Effect of low track vs comprehensive	Effect of high track vs comprehensive
Mavo	—	
Mavo-havo	—	0
Havo		0

A potential additional advantage of a comprehensive class is that it has a favorable effect on duration until entrance in higher education. These pupils need on average 0.3 to 0.5 less years to enter higher education.[47] This finding is intuitive as a comprehensive class provides an institutionally shorter way to higher education. Early tracked pupils more often take the longer route via mbo towards higher education, whereas non-tracked students predominantly take the shorter route via havo.

Our analysis is not informative on the underlying mechanism that translates early tracking to outcomes. Potential mechanisms are peer effects or teacher quality effects. In addition, extending school choice may reduce uncertainty with respect to the pupils' true capabilities which lowers the risk of sending them to the wrong school type. While it may be difficult for pupils in a low track to change to another education level that provides access to higher education, pupils in a comprehensive class are not restricted yet in their potential routes to higher education. The role of these effects in early tracking remains an interesting topic for further research.

Moreover, this study only investigates effects on participation in higher education and does not address the highest and lowest advice groups and classes as these are less relevant in this respect. This also implies that potential effects of early tracking on performance in secondary or

[47] This finding is confirmed in some OLS regressions in which we regress the time between the start of secondary education and enrollment in higher education on a dummy variable for tracking and different sets of covariates. As there is some concern on issues like sample selection a more elaborate analysis is needed to study duration until entrance in more detail.

higher education, which may serve as indicators for excellence, are beyond the scope of this research.

Our study supports the OECD conclusion on the Dutch education system that the early tracking regime has a negative effect on participation in higher education. Pupils with a mavo advice are more likely to enroll in and complete higher education when they start secondary education in a comprehensive class. In the light of this conclusion, the decrease in the number of early tracked pupils we observe over time in the cohort samples under consideration, has been an encouraging development yet. Enrollment in and completion of higher education can be further increased by stimulating participation in combined first-grade classes that keep pupils with a mavo or havo advice together for an additional one or two years.
Any more specific policy implications cannot be derived from our results. With respect to vmbo schools for example, we might expect that vmbo-tl students gain from being in a combined class together with havo pupils. However, we do not know the effect of such a change for the lower advice pupils in vmbo. In addition, more information on underlying mechanisms is needed to be able to target specific policies at the key factors that drive the impact of early tracking.

References

Angrist, J., 2003, Treatment effect heterogeneity in theory and practice, *The Economic Journal*, vol. 114, pp. C52-C83.

Antenbrink, P., K. Burger, M. Cornet, M. Rensman, D. Webbink, 2005, Nederlands onderwijs en onderzoek in internationaal perspectief, The Hague: CPB Netherlands Bureau for Economic Policy Analysis.

Ariga, K., G. Brunello, 2007, Does secondary school tracking affect performance? Evidence from IALS, IZA Discussion Paper 2643.

Argys, L., D. Rees, D. Brewer, 1996, Detracking America's schools: equity at zero cost?, *Journal of Policy Analysis and Management*, vol. 15, no. 4, pp. 623-645.

Bound, J., D. Jager, R. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association*, vol. 90, pp. 443-450.

Brunello, G., D. Checchi, 2007, Does tracking affect equality of opportunity? New International Evidence, *Economic Policy*, vol. 52 , pp. 781-861.

Brunello, G., M. Giannini, 2004, Stratified or comprehensive? The economic efficiency of school design, *Scottish Journal of Political Economy*, vol. 51, no. 2, pp.173-193.

Driessen, G., M. van der Werf, 1992, Het functioneren van het voortgezet onderwijs: beschrijving steekproef en psychometrische kwaliteit van de instrumenten, Groningen/ Nijmegen: RION/ITS.

Driessen, G., 2006, Het advies voortgezet onderwijs: is de overadvisering over?, *Mens & Maatschappij*, vol. 81, no. 1, pp. 5-23.

Epple, D., E. Newlon, R. Romano, 2002, Ability tracking, school competition, and the distribution of educational benefits, *Journal of Public Economics*, vol. 83, no.1, pp. 1-48.

Figlio, D., M. Page, 2002, School choice and the distributional effects of ability tracking: does separation increase inequality?, *Journal of Urban Economics*, vol. 83, pp. 497-514.

Hanushek, E., L. Woessman, 2006, Does education tracking affect performance and inequality? Differences in differences evidence across countries, *The Economic Journal*, vol. 116, C63-C76.

Hoxby, C.M., 2000, Peer effects in the classroom: learning from gender and race variation, NBER Working Paper 7867, Cambridge, MA: National Bureau of Economic Research.

Imbens, G., J. Angrist, 1994, Identification and estimation of local average treatment effects, *Econometrica*, vol. 62, no.2, pp.467-475.

Jakubowski, M., 2007, Effects of tracking on achievements growth, exploring difference-in-difference approach to PIRLS, TIMMS and PISA data, unpublished manuscript.

Leuven, E., M. Lindahl, H. Oosterbeek, D. Webbink, 2007, The effect of extra funding for disadvantaged pupils on achievement, *Review of economics and statistics*, vol. 89, no. 4, pp. 721-736.

Minne, B., M. Rensman, B. Vroomen, D. Webbink, 2007, Excellence for productivity?, The Hague: CPB Netherlands Bureau for Economic Policy Analysis.

OECD, 2006, Where immigrant students succeed. A comparative review of performance and engagement in PISA 2003, Paris: OECD.

OECD, 2007, Thematic Review of Tertiary Education – The Netherlands, Country Note.

Oreopoulos, P., 2006, Estimating average and local average treatment effects of education when compulsory schooling laws really matter, *American Economic Review,* vol. 96, no. 1, pp. 152-175.

Sacerdote, B., 2001, Peer effects with random assignment: Results for Dartmouth roommates, *Quarterly Journal of Economics*, vol. 116, no. 2, pp. 681-704.

Schuetz, G., H. Ursprung, L. Woessman, 2005, Education policy and equality of opportunity, CESIFO Working Paper 1518.

Staiger, D., J. Stock, 1997, Instrumental variables regression with weak instruments, *Econometrica*, vol. 65, no. 3, pp.557-586.

Van de Werfhorst, H., J. Mijs, 2007, Onderwijsdifferentiatieen ongelijkheid. Nederland in vergelijkend perspectief, Amsterdam: University of Amsterdam.

Appendix A

This appendix provides an overview of the education careers of both the tracked and non-tracked pupils in the estimation samples of the cohorts 1993 and 1999.

Table A.1 **Routes through education in first nine years after entering secondary education, tracked (=mavo) versus non-tracked (=mavo/havo or mavo/havo/vwo); mavo-advice estimation sample, cohort 1993**

	< Mavo	Mavo	Comprehen-sive: (l)mh or (l)mhv	Havo	Havo/vwo	Vwo	Mbo	Higher education	Left[a]
Year 1									
Non-tracked	0	0	100	0	0	0	0	0	0
Tracked	0	100	0	0	0	0	0	0	0
Year 2									
Non-tracked	11	39	30	13	6	1	0	0	0
Tracked	5	92	0	1	1	0	0	0	1
Year 3									
Non-tracked	14	59	3	19	1	3	0	0	1
Tracked	10	87	0	2	0	0	0	0	1
Year 4									
Non-tracked	15	61	0	19	0	3	0	0	2
Tracked	12	84	0	2	0	0	0	0	2
Year 5									
Non-tracked	5	13	0	26	0	3	41	0	13
Tracked	4	12	0	13	0	0	55	0	16
Year 6									
Non-tracked	0	0	0	15	0	2	50	6	26
Tracked	0	0	0	10	0	0	61	1	28
Year 7									
Non-tracked	0	0	0	3	0	3	48	15	34
Tracked	0	0	0	2	0	0	55	6	37
Year 8									
Non-tracked	0	0	0	0	0	1	32	24	44
Tracked	0	0	0	0	0	1	35	17	48
Year 9									
Non-tracked	0	0	0	0	0	0	14	32	54
Tracked	0	0	0	0	0	0	13	25	62

[a] The percentages shown in this column seem to overstate real shares of people that left education. This is because the cohort may have lost track of some students that moved to other schools (in other cities) or that have temporarily left education and returned after some time. This also explains why the reported percentages active in higher education are somewhat lower than the ones we found after having matched the students from the VOCL cohorts with a central register of all higher education participants over time.

Table A.2 **Routes through education in first eight years after entering secondary education, tracked (=vmbo) versus non-tracked (=mavo/havo or mavo/havo/vwo); mavo-advice estimation sample, cohort 1999**

	Vmbo	Comprehen-sive: (l)mh or (l)mhv	Havo	Havo/vwo	Vwo	Mbo	Higher education	Left[a]
Year 1								
Non-tracked	0	100	0	0	0	0	0	0
Tracked	100	0	0	0	0	0	0	0
Year 2								
Non-tracked	47	37	10	5	1	0	0	5
Tracked	96	1	2	0	0	0	0	0
Year 3								
Non-tracked	73	1	20	2	2	0	0	3
Tracked	94	0	3	0	0	0	0	3
Year 4								
Non-tracked	75	0	18	0	2	0	0	5
Tracked	92	0	3	0	0	0	0	5
Year 5								
Non-tracked	15	0	24	0	1	47	0	13
Tracked	13	0	9	0	0	57	0	21
Year 6								
Non-tracked	2	0	12	0	2	65	9	10
Tracked	1	0	8	0	0	75	2	14
Year 7								
Non-tracked	0	0	3	0	1	61	19	16
Tracked	0	0	2	0	0	73	6	19
Year 8								
Non-tracked	0	0	1	0	0	48	25	26
Tracked	0	0	1	0	0	56	14	29

[a] The percentages shown in this column seem to overstate real shares of people that left education. This is because the cohort may have lost track of some students that moved to other schools (in other cities) or that have temporarily left education and returned after some time. This also explains why the reported percentages active in higher education are somewhat lower than the ones we found after having matched the students from the VOCL cohorts with a central register of all higher education participants over time.

Table A.3 **Routes towards higher education: percentages of last completed education level before entering higher education (cohort 1993)**

Education level	Early tracked: mavo	Non-tracked: mavo-havo(-vwo)
Mbo	62	41
Havo	25	43
Vwo	3	7
Other	10	9

Table A.4 **Routes towards higher education: percentages of last completed education level before entering higher education (cohort 1999)**

Education level	Early tracked: vmbo	Non-tracked: vmbo-havo(-vwo)
Mbo	66	45
Havo	31	50
Vwo	2	5
Other	1	0

Appendix B

This appendix provides a detailed description of the data for the estimation samples in the cohorts 1993 and 1999. Tables B.1 and B.2 present the descriptive statistics for the cohorts 1993 and 1999, respectively.

Table B.1 Descriptive statistics for estimation sample, mavo-advice group, cohort 1993, early tracked (mavo) versus later tracked (mavo-havo-(vwo))[a]

Variables	Early tracked: mavo	Non-tracked: mavo-havo(-vwo)	P-value of difference
1. Ability			
Test score arithmetic[b]	11.3	11.7	0.001
Test score language[b]	12.0	12.3	0.009
Test score information processing[b]	12.2	12.4	0.024
2. Personal and SES variables			
Female	56	54	0.175
Age (years)	12.56	12.56	0.920
Highest education level parents			0.091
No primary education	0	0	
Primary education	9	11	
Secondary education low	20	17	
Secondary education high	53	51	
Higher education first phase	15	17	
Higher education second phase	3	3	
Higher education third phase	0	1	
Profession level parents			0.344
Worker	32	29	
Self-employed without personnel	10	9	
Self-employed with personnel	12	12	
Lower employee	13	13	
Intermediate employee	20	24	
Higher profession	13	14	
Other	0	0	
Father's country of birth			0.000
The Netherlands	92	86	
Marocco	2	1	
Antilles, Surinam, Aruba	1	2	
Turkey	2	3	
Other	4	8	
Urbanization city of residence			0.000
Very high	11	11	
High	17	25	
Median	23	19	
Modest	28	21	
Low	22	24	
3. Education outcomes			
Ever participated in higher education	37.3	44.0	0.000
Completed higher education	23.3	28.3	0.003
Time to enrollment in higher education in years	7.72	7.29	0.000
Number of pupils	1864	1090	

[a] All numbers represent percentages, unless stated otherwise.
[b] Test scores are of entrance tests taken in the first year of secondary education. The maximum score for each test is 20.

Table B.2 Descriptive Statistics for estimation sample, mavo-advice group, cohort 1999, early tracked (vmbo) versus later tracked (vmbo-havo(-vwo))[a]

Variables	Early tracked: vmbo[b]	Non-tracked: vmbo-havo(-vwo)	P-value of difference
1. Ability measures			
Test score arithmetic[c]	11.0	11.7	0.000
Test score language[c]	11.5	12.0	0.000
Test score information processing[c]	11.6	11.9	0.010
2. Personal and SES variables			
Female	54	53	0.647
Age (years)	12.6	12.6	0.000
Highest education level mother			0.028
Primary education	10	7	
Secondary education low	24	20	
Secondary education high	37	39	
Higher education	8	10	
Other	21	24	
Highest education level father			0.022
Primary education	6	8	
Secondary education low	18	16	
Secondary education high	38	33	
Higher education	10	14	
Other	28	30	
Ethnicity			0.000
The Netherlands	88	80	
Marocco	1	3	
Antilles, Surinam, Aruba	2	4	
Turkey	2	3	
other	9	11	
Urbanization city of residence			0.000
Very high	11	12	
High	18	34	
Median	25	18	
Modest	24	14	
Low	23	23	
3. Education outcomes			
Ever participated in higher education	26.6	39.2	0.000
Time to enrollment in higher education in years	7.07	6.56	0.000
Number of pupils	1047	1106	

[a] All numbers represent percentages, unless stated otherwise.
[b] Vmbo includes former mavo (vmbo-tl) as well as vbo.
[c] Test scores are of entrance tests taken in first year of secondary education. The maximum score for each test is 20.

Appendix C

This appendix provides an overview of the number of categorial and combined schools in our estimation sample for each of the municipality types.

Table C.1 presents the supply-ratios of categorial schools in 1989. The first column reports the twelve different municipality types. For each of these types, the number of categorial and combined schools in our sample is reported in the second and third column, respectively. The last column shows the supply-ratio of tracked classes, which is defined as the number of categorial schools divided by the total number of schools in each municipality type. Equivalently, table C.2 shows the number of categorial schools, combined schools and the supply-ratios for the cohort 1993.

Table C.1 Supply-ratio of categorial schools in 1989

Municipality type	# Categorial schools	# Combined schools	Supply-ratio
Countryside A1	0	0	
Countryside A2	0	0	
Countryside A3	4	1	0.80
Countryside A4	12	0	1.00
Urbanized countryside B1	11	1	0.92
Urbanized countryside B2	29	5	0.85
Specific commuter municipality B3	13	9	0.59
Rural cities C1	7	1	0.88
Small cities C2	17	3	0.85
Medium-sized cities C3	5	4	0.56
Medium-sized cities C4	12	11	0.52
Big cities C5	25	9	0.74

Table C.2 Supply-ratio of categorial schools in 1993

Municipality type	# Categorial schools	# Combined schools	Supply-ratio
Countryside A1	0	0	
Countryside A2	0	0	
Countryside A3	3	1	0.75
Countryside A4	8	6	0.57
Urbanized countryside B1	9	1	0.90
Urbanized countryside B2	17	5	0.77
Specific commuter municipality B3	12	8	0.60
Rural cities C1	7	2	0.78
Small cities C2	11	14	0.44
Medium-sized cities C3	7	5	0.58
Medium-sized cities C4	7	12	0.37