Economics of Education Review xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Does the timing of tracking affect higher education completion?

Roel van Elk[a,*], Marc van der Steeg[a], Dinand Webbink[b]

[a] CPB Netherlands Bureau for Economic Policy Analysis, The Netherlands
[b] CPB Netherlands Bureau for Economic Policy Analysis, Erasmus School of Economics, Rotterdam, IZA, Tinbergen Institute, The Netherlands

ARTICLE INFO

Article history:
Received 9 October 2009
Received in revised form 15 April 2011
Accepted 19 April 2011

JEL classification:
I2
I28

Keywords:
Early tracking
Higher education graduation

ABSTRACT

This paper investigates the effect of the timing of tracking on completion of higher education by exploiting unique variation from the Dutch education system. At the age of 12 Dutch students can enrol in tracked schools or in comprehensive schools. The comprehensive schools postpone enrolment into tracked classes by one or two years. OLS- and IV-estimates, using regional variation in the supply of schools as instruments, show that early tracking has a detrimental effect on completion of higher education for students at the margin of the Dutch high and low tracks. The negative effects of early tracking are larger for students with relatively high ability or students with a higher socioeconomic background. In addition, we find no negative effects of comprehensive classes on higher ability students. These results suggest that increasing participation in comprehensive classes would increase graduation from higher education.

1. Introduction

One of the most remarkable differences between school systems is the age at which students are tracked into different levels of education. Countries such as Germany, Austria, Hungary and the Slovak Republic track students as early as age 10. Other countries, such as Norway, Sweden, Canada, the US, the UK and Japan, have a comprehensive school system until the end of secondary education. Differences in the age of tracking might have far-reaching consequences for the equity and efficiency of educational outcomes. However, little is known about the effects of this key institutional feature of education systems.

A researcher interested in the effect of the age of tracking would like to know what would have happened with a student if he/she had not been tracked at, for example, the age of 12 but rather at 14. Investigating this question requires variation in the age of tracking. However, in most countries or states tracking takes place at one age or not at all. Hence, variation in the age of tracking is often not available within a single education system. Therefore, researchers often rely on variation between countries or states, and estimate regression models using a large set of covariates (for instance, Schuetz, Ursprung, & Woessman, 2005). The multitude of potential confounders due to differences between countries or states poses serious problems for the identification of the effect of the age of tracking in these models. A recent innovative study by Hanushek and Woessman (2006) addresses this problem by using a difference-in-differences approach exploiting international test scores taken at different ages. However, this study has also been criticised for neglecting incentive effects of tracking (Eisenkopf, 2007). Studies that investigate the effects of tracking within countries compare the outcomes of tracked students with non-tracked students (Duflo, Dupas, & Kremer, 2008; Figlio & Page, 2002) or exploit the gradual introduction of a comprehensive system (Pekkarinen, Uusitalo, & Kerr, 2009). To our knowledge there are no previous studies that exploit differences in the timing of tracking within the same education system.

The Dutch system offers the opportunity to investigate the effect of the age of tracking within a single school system. At the age of 12 Dutch students can enrol in tracked classes (categorial classes) or in comprehensive classes. The

* Corresponding author. Tel.: +31 703386129.
E-mail address: r.a.van.elk@cpb.nl (R. van Elk).

0272-7757/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2011.04.014

comprehensive classes consist of a combination of school types and take one or two years. Basically, these comprehensive classes delay enrolment into tracked classes until age 13 or 14. This paper compares higher education completion rates of students that are tracked at the age of 12 with the outcomes of students that are tracked at the age of 13 or 14 within the Dutch education system. The advantage of this approach is that students in the tracked and comprehensive parts of the system encounter exactly the same educational environment after the period of treatment. Hence there are no concerns for confounding factors due to differences between countries. Relative to existing studies that compare outcomes of tracked versus non-tracked pupils, our paper contributes by comparing two alternative tracking ages. We compare a group of pupils that is tracked early with a group of pupils that is tracked later. This implies that there are two differences in treatment between both groups that may affect future educational outcomes. First, there is the difference of being in a tracked versus a comprehensive class for one or two years. Tracked pupils might, for example, benefit from more homogenous classrooms, while pupils in a comprehensive class might benefit from the interaction with their peers. Second, there is a difference in the age at which pupils enter the tracked system in which they are allocated to different education types. This may also affect outcomes if the ability of teachers and parents to select the most suitable education types for pupils depends on the pupil's age. Tracking pupils at an early age may imply more uncertainty with respect to the pupil's true capabilities and hence a higher risk of sending them to an inappropriate school type. Hence, the probability of misallocation of pupils to tracks may decrease with tracking age. This aspect is not taken into account in studies that compare tracked versus non-tracked pupils. Suppose a researcher compared the effects of being tracked versus non-tracked at age 12 and found a positive result for tracking. Tracking at age 14 might be better still if a positive effect of realising better matches between students and education types at a later tracking age dominates the negative effect of being in a comprehensive class for one or two years. This illustrates that comparing tracking ages, taking into account both aspects, is of critical importance for policymakers aiming to improve the school system.

The main challenge for our analysis is the potential self-selection of students into categorial or comprehensive classes. If categorial or comprehensive classes attract more able or more motivated students, unobserved differences might bias the estimates. To address this potential problem we control for a large set of covariates including various measures of socioeconomic background and test scores on arithmetic, language and information processing. Moreover, we adopt an instrumental variable approach in which we exploit regional variation in the supply of schools. Previous papers in the economic literature used regional differences in the supply of schools to instrument for educational attainment (see for instance Card, 1993; Currie & Moretti, 2003; Park & Kang, 2008). In this paper we use regional differences in the supply of tracked or comprehensive classes to instrument for the age of tracking.

The main outcome variable in the analysis is the completion of higher education. This outcome plays a prominent role in the Dutch policy discussion on participation in higher education. A recent report suggests that the early tracking regime in the Netherlands constrains the growth of higher education participation significantly (OECD, 2007). We focus the analysis on a sample of students that is most likely to be affected by early tracking: those who were advised to enrol in the lower track of secondary education. In addition, we investigate the effects of early tracking for higher ability pupils that were advised to enrol in the higher track of secondary education.

Our main finding is that tracking at the age of 12 compared to tracking at 13 or 14 has a negative effect on the probability of higher education completion. Both ordinary least-squares (OLS) and instrumental variables (IV) estimations show negative effects of early tracking. We also find that the comprehensive school has no negative effect on pupils with a higher ability. This suggests that postponing tracking to older children would increase the number of people completing higher education.

The structure of this paper is as follows. Section 2 discusses the previous literature. Sections 3 and 4 describe the Dutch context and data. The empirical strategy is discussed in Section 5. Section 6 presents the main estimation results. In Section 7 several robustness analyses are presented. Section 8 presents an additional analysis for higher ability students. Finally, Section 9 concludes and discusses potential policy implications.

2. Previous studies

The empirical literature has produced conflicting results on the effects of tracking. Several studies use cross-country data and regress international test scores on institutional differences between countries. Ariga and Brunello (2007) study the effect of the number of years spent in a tracked system on the performance of young adults in a standardised cognitive test. Using constraints on educational participation, such as financial constraint or family reasons, as instrumental variables they find a positive effect of tracking on performance. Hanushek and Woessman (2006) use a difference-in-differences approach to identify the effect of an early tracking regime on standardised test scores. They match international primary school tests to secondary school tests and compare differences in test scores across countries. They find that early tracking increases inequality, while it does not have a clear impact on average achievement. Recently, this study has been criticised. Jakubowski (2007) argues that the differences in design of the international tests may affect the results. While PISA measures pupils aged 15 (independent of their grade), PIRLS/TIMMS measures achievement in specific grades (independent of age). He concludes that early tracking does not increase inequality when comparing pupils of the same age and grade. Eisenkopf (2007) argues that the findings by Hanushek and Woessman (2006) are biased because they neglect the incentive effects of tracking. Brunello and Checchi (2007) focus on the interaction between family background and tracking. They estimate regression models with many covariates including country-by-cohort dummies and interactions with family background. Their estimates suggest that tracking reinforces the role of family background with respect to educational attainment and labour market outcomes. Schuetz et al. (2005) 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. They also find that early tracking increases the effect of family background.

A second group of studies uses variation within countries. Duflo et al. (2008) provide evidence from a randomised experiment in Kenya. In this experiment 121 primary schools, who all had a single grade 1 class, received funding to hire an extra teacher and split this class into two sections. In 60 randomly chosen schools the pupils were randomly assigned to one of the two sections. In the other 61 schools the pupils were assigned based on a ranking of their previous educational performance. After 18 months pupils in the schools that selected based on performance scored on average 0.14 standard deviation higher than pupils in schools that selected randomly. In addition they find that pupils at all levels of the distribution benefited from tracking. The authors suggest that tracking was beneficial because it helped teachers to focus their teaching to a level appropriate to most students in class. Evidence for developed countries is provided by several studies that exploit educational reforms from countries that switch from a tracked to a comprehensive system. Galindo-Rueda and Vignoles (2005) investigate the effect of the change from selective schools towards a comprehensive system in England. However, Manning and Pischke (2006) note that the gradual introduction of the new system started in poorer areas which might bias the estimates. A recent study by Pekkarinen et al. (2009) exploits the replacement of the Finnish two-track school system by a uniform nine-year comprehensive school. Identification comes from the gradual introduction of the reform during a six-year period. They find that the reform had a small positive effect on verbal test scores but no effect on the mean performance in arithmetic or logical reasoning tests. In addition, the reform improved test scores of students from lowly educated families. Figlio and Page (2002), using US data, investigate the effect of tracking on the improvement in maths test scores between the 8th and 10th grade for different ability groups. 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. 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 suggest that low-ability students may actually gain from tracking. Betts and Shkolnik (2000) examine the effect of formal ability grouping in the US, making use of a nationally representative data set which asks principals whether their schools use tracking in their math classes.[1] The authors find no overall effect on math achievement growth and find little or no differential effects for high-achieving, average or low-achieving students.

Our study provides the following contributions to the literature.

First, we examine the effects of the timing of tracking. In the Netherlands, each pupil ends up in a tracked system at age 14 and we investigate whether tracking at age 12 improves on tracking at 13 or 14. Previous studies, as cited above, investigate whether tracking improves educational outcomes or not.

Second, we exploit the coexistence of a tracked and a comprehensive system within the same school system. Previous studies exploit the gradual introduction of a reform from a tracked system to a comprehensive system within countries, which means that these systems have not been in existence for a long time.

Third, we investigate long-term outcomes of tracking. Most of the previous work focuses on test scores at a relatively young age. This might compromise a proper evaluation of the impact of tracking on performance if tracking is not given sufficient time to work out its effect. We look at later outcomes by comparing higher education completion rates across early versus later tracked students.

3. Variation in the timing of tracking in Dutch education

In order to analyse 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 the Dutch educational system as a whole is considered to be an early tracking regime, 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, 12-year-old pupils are directly tracked into categorial classes of a certain education level. Other schools offer one- or two-year comprehensive classes, in which pupils are kept together before they are tracked in a particular school type. Hence, pupils starting secondary education in a comprehensive class postpone their choice of school type by at least one year. This difference causes variation in the timing of tracking, which we can exploit to analyse the effects of early tracking by comparing the educational outcomes of pupils who start secondary education in a categorial class (the 'tracked' pupils) to those who start in a combined class (the 'non-tracked' pupils). Strictly speaking, we investigate the effect of being tracked one or two years earlier.

The structure of Dutch education after leaving primary education at the age of 12 is shown in Fig. 1. In 1989 secondary education consisted of four tracks: 'lbo' (pre-vocational secondary education), 'mavo' (lower general secondary education), 'havo' (higher general secondary education) and 'vwo' (pre-university education). After secondary education students can enrol in 'mbo' (upper secondary vocational education) or higher education, depending on the type of school completed. Mbo is oriented towards vocational training and is offered at four levels which prepare students for specific professions in

[1] Rees, Brewer, and Argys (2000) have criticised the study because of concerns on the reliability of principal-reported measures of tracking.

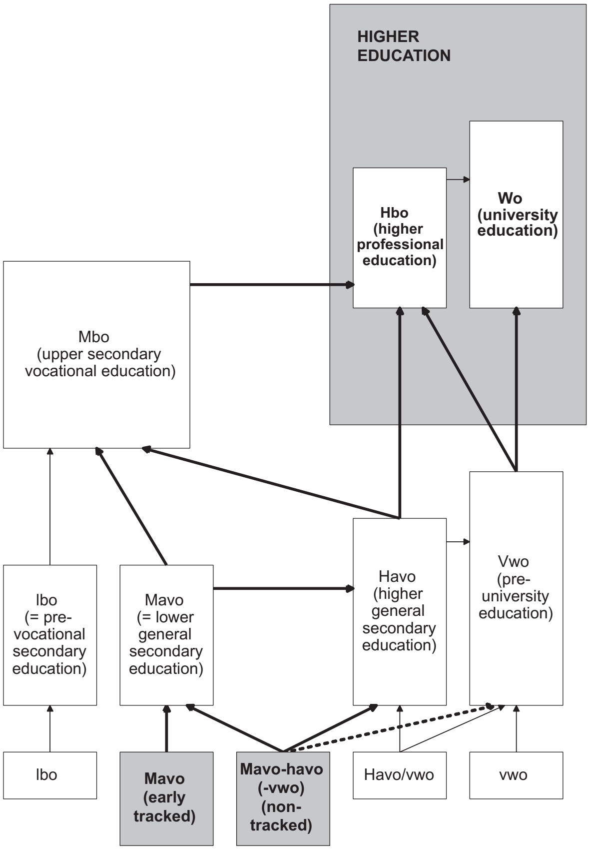

Fig. 1. Flow chart of Dutch education system: routes towards higher education for tracked versus non-tracked pupils in first year of secondary education.

the labour market. 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 four (the highest level) degree of 'mbo'. Hence, pupils who have completed havo or vwo are qualified for access to higher education, while pupils with a mavo degree cannot directly enrol. We label all pupils who start secondary 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 start 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'. After all, both are tracked at the age of 12 in either the low track or the high track. The pupils who enter some combined class that consists of at least mavo–havo delay tracking until age 13 or 14. We make no distinction whether this is a mavo–havo, mavo–havo–vwo or another combined class that includes both mavo and havo. As long as the choice of both the low track and the high track is open it is labelled as a comprehensive class. Those pupils can be considered to be subject to a 'comprehensive system'.

In the literature different definitions of tracking are used. For example, the US has, strictly speaking, 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 particular courses are taught at different levels of complexity. Often, this distinction is ignored in the existing literature.

We define tracking as the allocation of pupils into high or low track classes. In the Netherlands first grade categorial classes are in general offered by categorial schools and first grade comprehensive classes by comprehensive schools. Parents are equally free to choose for categorial or comprehensive schools and there are typically no differences in the way courses are taught between the two school types.

Only a few schools offer both comprehensive and categorial classes. The course programmes that are offered at these schools are in principle the same as those offered at the single comprehensive or categorial schools.

In the Netherlands pupils at the end of primary education are advised about the type of secondary school most appropriate for them. The advice on education level 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. In our analysis, we primarily focus on pupils that are advised to follow the mavo track when they leave primary education and compare higher education completion rates 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. Hence, we select a homogeneous group with respect to ability and compare the subsample of this group that moves into a low track to the subsample that moves into a comprehensive class. From the OECD-perspective (see Section 1 and OECD, 2007) this group seems most interesting as the former subsample does not qualify for direct access to higher education. If early tracking has a detrimental effect on the completion of higher education, the mavo advice pupils who are tracked early are likely to be those who are most damaged due to their subsequent inability to access higher education directly.

Our main analysis addresses the question whether pupils that are recommended by the teachers to go to the low track benefit from being in a comprehensive class compared to being in the low track. In addition we will analyse the effects of tracking for higher ability pupils, by focusing on pupils that are advised to follow the havo track and compare outcomes of those who enter a comprehensive class to those who are tracked early into the high track.

4. Data

In our empirical analysis, we use data from the Secondary Education Pupil Cohort 1989. These are longitudinal data collected by Statistics Netherlands and the University of Groningen (Driessen & van der Werf, 1992; Statistics Netherlands, 1991). The cohort consists of a representative sample of around 20,000 pupils that enrolled in the first grade of secondary education in 1989. These pupils were followed during their school careers until they left the education system. Hence, for each pupil in each calendar year their corresponding school type and grade is known. The public files of the Secondary Education Pupil Cohorts follow the pupils until the school year 2003–2004. We received access to the private files of Statistics Netherlands with the most recent available data until the school year 2007–2008. This provided us with the most up-to-date information on completion of higher education. A comparison of our sample with the population shows that the shares of pupils that completed higher education (around 30%) in the cohort 1989 are quite similar to the nationwide population figures.[2]

Our data include information on the highest level of education completed from which we derive a dummy variable for completion of higher education. The data also provide information on a pupil's school type and grade for each calendar year. We observe whether a pupil is enrolled in secondary education in a categorial mavo class or in a comprehensive mavo–havo or mavo–havo–vwo class. The former group is tracked early while the latter delays tracking by one or two years. From this information we construct our main tracking variable. Our data contain a large set of covariates. First, the data provide a large set of socioeconomic background variables. The parents of the pupils were asked to fill in questionnaires which include questions on ethnicity (5 categories), education level (7 levels), profession level (7 categories) and family composition (8 categories). Second, the data contain information on the urbanisation of residence (5 categories), gender and age of the pupils. In addition, we condition on scores from 'entrance tests' in arithmetic, language and information processing, which pupils undertook at the start of secondary education. These tests are comparable to the main standardised test at the end of Dutch primary education (the CITO test) and the results serve as an indicator for pupil ability. Scores on each of the tests are between 0 and 20 (maximum score).

4.1. Sample statistics

The sample of pupils that leave primary education with a mavo-advice includes 4912 pupils, of which 3123 (63.6%) enter secondary education in a categorial mavo class and 1190 (24.2%) in a combined mavo–havo or mavo–havo–vwo class. The sample statistics of the variables we use in our analysis are shown in Table 1. We exclude pupils who are missing these variables, which leaves us with an estimation sample of 3936 students.[3] Column (1) reports the sample means for the group of pupils that start secondary education in a tracked class, column (2) reports them for the group of pupils that start secondary education in a comprehensive class. Column (3) reports the p-value of the difference, calculated using a two-tailed t-test or a chi-squared test.

A comparison of the covariates of the two groups of students reveals some differences in socioeconomic background. Parents from students in comprehensive classes are slightly higher educated and their professional level is also slightly higher. Parents of students in tracked classes have Dutch nationality more often. In addition, students in tracked classes are more likely to live in families with two parents and in regions with a lower degree of urbanisation. Despite these differences, the personal characteristics of the students in both groups are remarkably comparable. The age and gender composition is equal. Even more importantly, the scores on all three ability tests taken in the first year of secondary education are also equal. A regression of tracking on all covariates yields no significant effects. As such, the sample statistics suggest that the observable characteristics of the groups are quite similar.

The bottom panel of Table 1 gives a first impression of the potential effect of early tracking on educational achievements. We observe that both enrolment and completion of higher education are significantly lower for early tracked pupils. The difference in completion of higher education between the two groups is approximately 5 percentage points.

Our data also allow us to observe the educational career of students in the different tracks. This may indicate to what extent the lower participation in higher education among the tracked pupils can be explained by differences in enrolment opportunities after the first years in secondary education. Pupils that started secondary education in a comprehensive class still have an opportunity to enrol into a higher track (i.e. havo or vwo) in subsequent years, which grants admission to higher education. Pupils that start in a categorial class do not have these possibilities. Only in special cases where pupils turn out to have been placed in the wrong track, are pupils allowed to change tracks, in which case they have to repeat one grade. Table 2 shows what happened with the students in each group during the first nine years after enrolment in secondary education. We observe that some of the non-tracked pupils move on directly to higher education types, which results in more non-tracked pupils ending up in education levels that provide direct access to higher education. 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 educational careers translate into differences in higher education enrolment in favour of non-tracked students.

5. Empirical strategy

We use two approaches for estimating the effect of early tracking on completion of higher education. In our first approach we use linear probability models that include various controls and estimate the following equation:

$$Y_i = \alpha X_i + \beta T_i + \varepsilon_i, \qquad (5.1)$$

[2] 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% (Minne, Rensman, Vroomen, & Webbink, 2007).

[3] There does not appear to be any systematic difference between our sample of pupils with non-missing values on all covariates and the excluded pupils on observable characteristics. Inclusion of the additional observations hardly affects the estimation results presented in Section 6.

Table 1
Sample statistics for sample with mavo-advice, early tracked versus later tracked.[a]

Variables	Early tracked: mavo (1)	Non-tracked: mavo–havo (–vwo) (2)	*P*-value of difference (3)
Ability test score arithmetic[b]	10.6	10.6	0.864
Test score language[b]	11.6	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	5	
Lower employee	11	12	
Intermediate employee	22	21	
Higher profession	12	16	
Other	17	19	
Ethnicity			0.000
The Netherlands	89	81	
Other (4 categories)	11	10	
Family composition			0.000
Father and mother	87	77	
Other (seven categories)	13	23	
Urbanisation 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
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.

Table 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.

	<Mavo	Mavo	Comprehensive: mh or mhv	Havo	Havo/vwo	Vwo	Mbo	Higher education	Left
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 6									
Non-tracked	2	3	0	16	0	4	49	3	22
Tracked	2	2	0	11	0	1	64	0	20
Year 9									
Non-tracked	0	0	0	0	0	0	16	27	57
Tracked	0	0	0	0	0	0	16	22	62

Table 3
Supply-ratio of categorial schools.

Municipality type	No. of categorial schools	No. of comprehensive schools	Supply-ratio
Countryside A1	0	0	
Countryside A2	0	0	
Countryside A3	4	1	0.80
Countryside A4	12	0	1.00
Urbanised countryside B1	11	1	0.92
Urbanised 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

where Y_i is a dummy variable for completion of higher education, X_i denotes a vector of background characteristics, T_i is a dummy variable which indicates whether a pupil is tracked early or not and ε_i is the error term. T_i takes the value 1 if a pupil starts secondary education in a categorial mavo and takes value 0 if the pupil starts in a comprehensive class. The parameter of interest is β. Estimation of Eq. (5.1) 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. In addition, we are able to control for differences in ability by including pupils' test scores.

Nevertheless, it is conceivable that unobservables exist that are correlated both with early tracking and the outcome variable. The differences in socioeconomic background characteristics we observe between tracked and non-tracked pupils (see Table 1) may give rise to concerns about unobserved heterogeneity which threatens the unconfoundedness assumption underlying standard OLS regressions. Motivated parents, for example, may rather place their offspring in a comprehensive class, which gives better opportunities to move into a more advanced type of secondary education (havo or vwo) later on.

Our second approach addresses this potential endogeneity problem. We use an instrumental variables (IV) approach that exploits regional variation in the supply of schools. This approach is similar to previous papers in the economic literature that used regional differences in the supply of schools to instrument for educational attainment (Card, 1993; Currie & Moretti, 2003; Park & Kang, 2008).

We use the relative supply of categorial schools in particular municipality types as an instrument for early tracking and estimate Eq. (5.1) by two-stage least squares (2sls). The first stage, in which early tracking is regressed on the supply-ratio and all covariates, is

$$T_i = \gamma X_i + \delta S_i + u_i, \qquad (5.2)$$

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. In our data 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 schools offering first grade categorial mavo classes (categorial schools) and the total number of schools offering first grade comprehensive classes (comprehensive 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. Four schools offer both comprehensive and categorial classes. They have been counted both as a categorial and a comprehensive school.

Table 3 provides an overview of the number of categorial and comprehensive schools in our estimation sample for each of the municipality types. The first column reports the twelve different municipality types. For each of these types, the numbers of categorial and comprehensive schools in our sample is reported in the second and third column, respectively. The last column shows the corresponding supply-ratios. The relative supply-ratios are substantially larger in the countryside, rural and small cities compared to larger cities and specific commuter municipalities. Hence, for pupils living in these first types the choice-set includes relatively fewer comprehensive schools.

The supply-ratio is a legitimate instrument if the exclusion restriction holds. Hence, the effect of early tracking is identified on the assumption that the supply-ratio only affects the outcome variable of interest through early tracking. To further reduce potential endogeneity bias from unobserved regional characteristics that are potentially correlated to both the supply-ratio and outcomes, we include an urbanisation indicator in the analysis to control for possible effects of the municipality of residence on the outcome variable. This urbanisation indicator is an additional measure for the type of municipality of residence. In contrast to the classification in 12 categories, this indicator is only based on the number of residents.

A concern with using the supply of schools as an instrument is that differences in supply may reflect differences in the demand for schools. In our case differences in supply-ratios may reflect differences in demand for categorial or comprehensive classes. We pursue this issue by investigating the robustness of the results to inclusion of variables that might affect the demand for certain school types (see Section 7.2). An additional issue regarding the exclusion restriction concerns potential education quality differences related to the supply of school types. In case of correlation between education quality and our instrument, the identifying assumption underlying instrumental variable analyses would be violated. We discuss this point in Section 7.2 and empirically address it by investigating the relationship between the supply-ratio and an observable measure of school quality.

Another concern with our instrument is that we measure the supply of categorial and comprehensive schools at an aggregated level. We only have information on the pupil's municipality type, and not specifically of the pupil's own municipality. If there is a lot of variation within municipality types, then our instrument may not reflect the actual options that are truly facing the pupils, resulting in a weak first stage. Furthermore, information at the individual level, that would also include some distance measure (for example the distance between the nearest categorial and the nearest comprehensive class), would be even more preferable, and would probably generate an even stronger first stage. Unfortunately, these individual data are not available. Nevertheless, as the next section will show, our first stage using aggregated information on the supply of schools remains sufficiently strong.

6. Main estimation results

This section shows the main estimation results of the two approaches used in this paper. The top panel of Table 4 shows the estimation results of a linear probability model that regresses completion of higher education on a dummy variable for early tracking for three specifications. Column (1) includes no further controls; column (2) controls for age, age squared, gender, ethnicity, educational and professional level of parents, family composition and urbanisation of residence; and column (3) also controls for test scores at the start of secondary education. The middle and bottom panel of Table 4 show the estimation results of the first and second stage of the instrumental variable approach, respectively. Each estimate is based on a separate regression. Robust standard errors corrected for clustering at the school level are in parentheses.

The top panel of Table 4 shows that early tracking is associated with a reduction in the probability of completing higher education of approximately 5 percentage points. All three OLS-regressions yield a statistically significant negative effect of early tracking on completion of higher education. The estimated coefficients slightly decrease (in absolute value) when additional control variables are included in the model.

A concern with the OLS-estimates is that they might be biased by unobserved factors. Therefore, we also use an instrumental variable approach. The middle panel of Table 4 shows the estimation results of the first stage Eq. (5.2). As expected, we find that an increase in the supply of comprehensive schools increases the probability of enrolment in comprehensive schools. A well-known concern with the IV-approach is the problem of weak instruments (Bound, Jager, & Baker, 1995). Staiger and Stock (1997) proposed using a cut-off value of 10 for the *F*-value of the excluded instrument. In the previous section we noted that our instrument is based on aggregated information and that an individual measure of the supply of schools would be preferable. Nevertheless, the *F*-values of our excluded instrument are well above the cut-off level, suggesting that the issue of weak instruments is not really a concern for our analysis. Including some smooth polynomial of the supply-ratios or particular step-functions of the supply-ratios yields similar results for the first stage. The bottom panel of Table 4 shows the second stage results. The models in column (2) and (3), which include controls for socioeconomic background and ability, yield negative point estimates. In the full model the estimated effect is −0.13, which is statistically significant at the 10% level. The results from these two IV-regressions are larger than the OLS-estimates. This difference in the size of the estimates might be related to the difference in the treated populations used for the estimation. The OLS estimates give the average treatment effect (ATE) for the whole population. The IV-estimates give the local average treatment effect (LATE). This is the effect of the treatment on the subpopulation of compliers (Imbens & Angrist, 1994): in our case the subpopulation that is more likely to choose the comprehensive class if the supply of comprehensive classes increases. 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 (Oreopoulos, 2006).

Next, we investigate whether the effects of early tracking are different for specific groups, such as girls and boys, high and low ability students and students with different socioeconomic backgrounds. Table 5 presents both OLS and IV estimation results of the full model (including all covariates) for various subsamples of pupils. Columns (1) and (2) show the results for the samples of girls and boys, respectively. Columns (3) and (4) show the results for pupils with relatively high and low ability. For this analysis we divided the sample in two based on the total test score ('low ability' is defined as having a score below the median total test score of 34). Finally, columns (5) and (6) show the estimation results for the pupils having parents with a relatively high and a relatively low education level. All pupils with parents having a highest education level above lower secondary education are in the top group; all other pupils are in the bottom group.

The estimation results in Table 5 show no clear difference between boys and girls. While the OLS point estimate is somewhat larger (in absolute value) for girls, the IV estimate is larger for boys. However, for the other subsamples we observe differences in the effect of early tracking. The effects of early tracking seem to be more negative for students with a higher ability and students with higher educated parents. The size of the estimates is larger for both the OLS and IV approach. Additional analyses, in which pupils are divided into two groups based on the professional level of the parents show a similar pattern (not shown in Table 5). These findings suggest that the detrimental effects of early tracking are larger for pupils who are more likely to enrol in and complete higher education.

Table 4
The impact of early tracking on completion of higher education (OLS and IV estimates).

	(1)	(2)	(3)
OLS			
Early tracking	−0.054***	−0.051***	−0.045**
	(0.019)	(0.018)	(0.018)
IV			
First stage			
Relative supply ratio	0.776***	0.866***	0.878***
	(0.224)	(0.254)	(0.255)
F-value excluded instrument	12.0	11.6	11.9
Second stage			
Early tracking	0.026	−0.107	−0.129*
	(0.070)	(0.067)	(0.068)
Socio-economic status variables (SES)	No	Yes	Yes
Test scores	No	No	Yes
Observations	3936	3936	3936

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

Table 5
The impact of early tracking on completion of higher education for subsamples.

	Girls	Boys	High ability	Low ability	High education level parents	Low education level parents
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	−0.049**	−0.036	−0.076***	−0.010	−0.067***	−0.008
	(0.024)	(0.024)	(0.024)	(0.022)	(0.025)	(0.021)
IV	−0.115	−0.131	−0.159*	−0.098	−0.163*	−0.102
	(0.094)	(0.090)	(0.086)	(0.097)	(0.095)	(0.093)
SES	Yes	Yes	Yes	Yes	Yes	Yes
Test scores	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2175	1761	2103	1833	2295	1641

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

Summarising, this section shows that early tracking reduces the probability of graduation from higher education for pupils at the margin of the high and low tracks (the pupils with a mavo advice). Both the findings from the OLS- and IV-regressions suggest that early tracking has a negative effect on the completion of higher education. The detrimental effects are larger for students with a relatively high ability and a higher socioeconomic background.

7. Robustness analyses

This section presents estimation results from several alternative model specifications and analyses designed to probe the robustness of our main results. We concentrate on potential bias of the OLS estimations due to unobserved heterogeneity (Section 7.1) and concerns regarding the validity of our instrument in the IV approach (Section 7.2).

7.1. Robustness of OLS analysis

We present robustness analyses in order to explore the magnitude of potential selection bias, which helps us to establish bounds on the true causal effect of early tracking. First we address selection concerns by focusing the analysis on a range of test scores. Excluding pupils with the highest and lowest ability from the analysis may result in a more homogeneous subsample, thereby reducing the risk of unobserved heterogeneity and selection bias in the estimations.

Table 6 presents the OLS estimation results of the full model (including all covariates) in which the estimation sample is narrowed based on test scores. The first model presents the results for the pupils with a total test score between 29 and 39 (the second and third quartile of the total test score distribution), the second model for the

Table 6
The impact of early tracking on completion of higher education (narrowing subsamples).

	(1)	(2)
	OLS Test scores between 29 and 39	OLS Test scores between 32 and 37
Early tracking	−0.046*	−0.070**
	(0.024)	(0.033)
SES	Yes	Yes
Test score	Yes	Yes
Observations	2027	1178

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

Table 7
Sample statistics for sample with mavo-advice, early tracked versus later tracked, subsample of pupils living in regions A4, C2, C3, C5.

Variables	Early tracked: mavo (1)	Non-tracked: mavo–havo (–vwo) (2)	*P*-value of difference (3)
Ability total test score	33.83	32.92	0.071
Parental education level			
Highest education level parents			0.001
No primary education	2	3	
Primary education	14	22	
Secondary education low	30	25	
Secondary education high	39	34	
Higher education first phase	14	13	
Higher education second phase	2	3	
Higher education third phase	0	0	
Observations	296	1228	

pupils with test scores between 32 and 37 (quantiles 37.5 and 62.5). Narrowing the subsamples with respect to test scores does not decrease the estimated effect (in absolute value) for early tracking. Narrowing to a subsample including only the middle 25 percent of the total test score distribution even increases the estimated coefficient (in absolute value) to −0.070.

The major concern regarding potential selection bias is that pupils with 'better' unobservable characteristics might self-select into the comprehensive classes, in which case the main OLS estimations would be biased upwards (in absolute value). As such, we are particularly interested in establishing a lower bound of the effect of early tracking. Our second analysis aims to provide a lower bound by focusing on a set of municipality types for which observable characteristics are 'better' for pupils in tracked classes. We use parental education level as the main criterium in the selection of the municipality types, as tracked and non-tracked pupils differ significantly on this observable variable (see Table 1).

Our data show that this condition, better observable characteristics for tracked pupils, is met for pupils living in municipality types A4, C2, C3 and C5. Table 7 presents the covariates on parental education level and total test scores of the pupils. Early tracked pupils have significantly higher educated parents and higher test scores. Other combinations of 4 or more municipality types yield relatively better characteristics for non-tracked pupils.

The OLS estimation results of the full model for this subsample are presented in Table 8.

The estimates show that early tracking decreases the probability of completion of higher education with 3.6 percentage points in the full model. In this analysis tracked pupils have a significantly higher parental education level and higher test scores. Assuming that these 'better' observable characteristics also imply 'better' unobserved characteristics for tracked pupils, the OLS estimation results may be expected to be biased downwards (in absolute value). Hence, the estimation result may be interpreted as the lower bound on the true effect of early tracking.

7.2. Validity of the instrument

To address potential endogeneity bias in the OLS estimation, we use an instrumental variable analysis using the relative supply of tracked schools as instrument for early tracking. Identification depends crucially on the assumption that the supply-ratio only affects the outcome through early tracking. If the only impact of the supply-ratio on higher education completion is through early tracking, then the supply-ratio should be statistically insignificant in a regression on higher education completion that also includes a dummy for tracking. As a first test on the validity of the instrument, we include the supply-ratio in the OLS model in column (3) of Table 4. This yields an insignificant effect of the supply-ratio.

A concern with using the supply of schools as instrument is that it may be driven by demand for specific school types. We address it by investigating the robustness of our estimation results to inclusion of variables that may reflect the demand for schools in various municipality types.

The first variable we use is a measure of the change in supply-ratios between 1989 and 1993. Changes over time may reflect differences in demand for specific school types in 1989. We construct a measure of the change in ratios by dividing the supply-ratio in 1993 by the supply-ratio in 1989 for each municipality type. A value larger than 1 implies an increase in the relative supply of categorial schools between 1989 and 1993.

Table 8
The impact of early tracking on completion of higher education (subset of regions).

	(1)	(2)	(3)
OLS			
Early tracking	−0.027	−0.040	−0.036
	(0.028)	(0.028)	(0.029)
Socio-economic status variables (SES)	No	Yes	Yes
Test scores	No	No	Yes
Observations	1524	1524	1524

Robust standard errors in parentheses.

Table 9
The impact of early tracking on completion of higher education (OLS and IV estimates).

	(1)	(2)	(3)
OLS			
Early tracking	−0.046**	−0.045**	−0.046**
	(0.018)	(0.018)	(0.018)
IV			
First stage			
Relative supply ratio	0.788**	0.812***	0.782***
	(0.317)	(0.251)	(0.273)
F-value excluded instrument	6.2	10.5	8.2
Second stage			
Early tracking	−0.191*	−0.153*	−0.213*
	(0.106)	(0.091)	(0.120)
Change in supply-ratio over time	Yes	No	No
Fraction of highly educated 1	No	Yes	No
Fraction of highly educated 2	No	No	Yes
Socio-economic status variables (SES)	Yes	Yes	Yes
Test scores	Yes	Yes	Yes
Observations	3936	3936	3936

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

A second indicator is the fraction of highly educated parents in a pupil's municipality type. If higher educated parents place more value on educational attainment and rather send their children to a comprehensive class with better opportunities to end up in a higher track later on, the fraction of highly educated parents may reflect the demand for comprehensive schools. For robustness, we use two different definitions for highly educated parents: the fraction of pupils with parental education levels 4, 5 and 6 and the fraction of pupils with parental education levels 5 and 6.

Table 9 shows the estimation results of both the OLS and IV models including these variables. The first model is the full model in column 3 of Table 4, including a variable indicating the change in the supply ratio between 1989 and 1993. The second and third models include the two variables for the fraction of highly educated parents in the full model.

The OLS estimates are hardly affected by the inclusion of these variables.

The IV results are comparable to those in the main estimations. Estimated coefficients are even larger now (in absolute value), with point estimates between −0.15 and −0.21. Models in which we include combinations of the three variables do not affect the results. We conclude that we do not find evidence that the demand for education might bias our findings.

An additional concern regarding the exclusion restriction lies in potential education quality differences across schools or regions. Differences in the supply of school types might be related to the quality of education, and hence to the outcome. We address the issue of potential confounding regional differences in general by including an urbanisation indicator in our main regressions. This indicator is statistically insignificant in all of our models. Specifically with respect to school quality, structural variation originating from differences in resources does not seem plausible as the Dutch system provides equal funding to schools. Nevertheless, schools may not be of equal quality because of differences in for example teacher effort or peer quality. In order to empirically address this issue we use average total test score at the school level as an observable measure of school quality and investigate its correlation with our instrument. In a regression of school average total test score on the supply-ratio and all other covariates, we find a statistically insignificant positive effect of the supply-ratio. This suggests that, conditional on all covariates, our instrument is not significantly related to school quality. Moreover, the positive correlation between a larger relative supply of tracked schools and school quality implies that, if anything, unobserved quality differences would bias our estimates downwards (in absolute value). Hence, we do not find evidence that our finding of a negative effect of early tracking is biased by unobserved school quality.

8. Effects of early tracking for higher ability pupils

The previous results suggest that pupils with a mavo advice can gain substantially 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 the pupils with a havo advice, the level just above a mavo advice. 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.

We concentrate on the pupils who leave primary education with a havo advice. We use this group to analyse the effect of being in a high track compared to being in a comprehensive class. The non-tracked pupils are again 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 edu-

Table 10
The impact of early tracking on completion of higher education (havo advice group).

	(1)	(2)	(3)
Early tracking	0.040	0.004	− 0.008
	(0.029)	(0.027)	(0.026)
Socio-economic status variables (SES)	No	Yes	Yes
Test scores	No	No	Yes
Observations	1770	1770	1770

Robust standard errors in parentheses.

cation in a havo or havo–vwo class.[4] 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. We estimate similar models as in Table 4. Unfortunately, we cannot estimate IV-regressions because there are no severe restrictions in the supply of high track classes. The possible bias due to self-selection into the high track classes will be discussed below.

Table 10 presents the OLS estimates of early tracking on completion of higher education. The estimation sample consists of 1162 pupils that entered secondary education in a havo–vwo class and 608 pupils that entered in a mavo–havo or mavo–havo–vwo class.

The results indicate that there is no significant effect of early tracking on 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. As mentioned before, we cannot use an IV approach to address the potential endogeneity issue. 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 to complete higher education more often. Hence, we expect that the estimated coefficients can be interpreted as the upper bounds of the effects of early tracking. As such, endogeneity does not seem to be a problem because all of our estimates, that might be upward biased, are statistically insignificant.

Summarising, these analyses suggest that pupils with a havo-advice experience no significant difference whether they start secondary education in a high track or in a comprehensive class. Hence, pupils that are advised to follow the high track, do not gain from enrolment in a high track compared to enrolment in a comprehensive class. Together with the previous result that pupils advised to enrol in the low track, do gain from being in a comprehensive class, this suggests that higher education completion can be improved by tracking all pupils later.

9. Conclusions and discussion

This paper investigates the effect of early tracking on the completion of higher education. In our analysis, we use data from the Secondary Education Pupil Cohorts of 1989, 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. 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 because the pupils in the mavo track have no direct access to higher education.

We find negative effects of early tracking on completion of higher education. The OLS estimates are supported by the IV estimates. The OLS estimates show that early tracking decreases the probability of completion of higher education by approximately 5 percentage points for pupils with a mavo advice. In the 1989 cohort, average completion of higher education for the tracked pupils is around 21%. Hence, pupils with a mavo advice that are in a categorial mavo can increase their probability of completing higher education by around 25% by entering a comprehensive class. The IV analyses yield even larger effects in absolute value. Hence, pupils with a mavo advice perform better in a comprehensive class than in a low track.

This finding for mavo-advice pupils is 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 also analysed effects of early tracking for pupils with a havo advice. 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. Hence, while pupils with a mavo-advice gain from being in a comprehensive class (compared to being in a low track), pupils with a havo-advice experience no significant difference whether they are in a high track or in a comprehensive class. These results suggest an inefficiency of the early tracking regime with respect to completion of higher education.

Our analysis is not informative on the underlying mechanisms that translate early tracking to outcomes. Potential mechanisms might be, for example, peer effects in the classroom or teacher quality effects. Low ability pupils may benefit from the interaction with their high-achieving peers in comprehensive classes. If peer effects are non-linear, this gives rise to efficiency gains. In the empirical literature, however, there is no consensus yet on the size

[4] We include havo-vwo in the high track because there are hardly any schools which offer categorial havo classes in the Netherlands. Actually, there turns out to be no school that offers such a categorial havo class in our estimation sample.

and functional form of peer group effects and the role of peers in the discussion on early tracking remains disputed.[5] If better teachers prefer to teach relatively high ability classes, teacher sorting may result in a higher quality of education in these classes.[6]

In addition, our empirical results may reflect differences between tracking ages in the ability of educators to allocate students to appropriate school types. The role of these mechanisms remains a valuable topic for further research which may contribute to a more profound understanding of the effects of early tracking.

Our study supports the OECD conclusion regarding 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 complete higher education when they start secondary education in a comprehensive class. As such, we expect that the recent decrease in the number of early tracked pupils in the Netherlands will increase future graduation from higher education. Completion of higher education can be further increased by stimulating participation of pupils with a mavo or havo advice in combined first-grade classes that postpone the point in the educational career at which these children enter different tracks.

References

Ammermüller, A., & Pischke, J.-S. (2009). Peer effects in European primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics*, *27*(3), 315–348.

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

Betts, J., & Shkolnik, J. (2000). The effects of ability grouping on student math achievement and resource allocation in secondary schools. *Economics of Education Review*, *19*(1), 1–15.

Bound, J., Jager, D., & Baker, R. (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*, *90*, 443–450.

Brunello, G., & Checchi, D. (2007). Does tracking affect equality of opportunity? New International Evidence. *Economic Policy*, *52*, 781–861.

Card, D. (1993). “Using Geographic Variation in College Proximity to Estimate the Return to Schooling,” Working Papers 696, Princeton University.

Currie, Janet, & Moretti, Enrico. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics*, *118*(4), 1495–1532. November.

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

Duflo, E., Dupas, P., & Kremer, M. (2008). Peer effects and the impact of tracking: Evidence from a Randomized Evaluation in Kenya, NBER Working Paper 14475.

Eisenkopf, G. (2007). Tracking and incentives. A comment on Hanushek and Woessman, Thurgau Institute of Economics, Research Paper Series no. 22.

Eisenkopf, G. (2010). Peer effects, motivation and learning. *Economics of Education Review*, *29*, 364–374.

Figlio, D, & Page, M. (2002). School choice and the distributional effects of ability tracking: Does separation increase inequality? *Journal of Urban Economics*, *83*, 497–514.

Galindo-Rueda, F., & Vignoles, A. (2005). The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability, CEE Discussion. Paper No. 52.

Hanushek, E. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*, 466–479.

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

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

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

Manning, A., & Pischke, J. (2006). Comprehensive versus selective schooling in England and Wales: What do we know? IZA Discussion Paper 2072.

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

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*, *96*(1), 152–175.

Park, C., & Kang, C. (2008). Does education induce healthy lifestyle? *Journal of Health Economics*, *27*(6), 1516–1531.

Pekkarinen, T., Uusitalo, R., & Kerr, S. (2009). School tracking and development of cognitive skills, IZA Discussion Paper No. 4058.

Rees, D., Brewer, D., & Argys, L. (2000). How should we measure the effect of ability grouping on student performance? *Economics of Education Review*, *19*(1), 16–20.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools and academic achievement. *Econometrica*, *73*(2), 417–458.

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

Staiger, D., & Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.

Statistics Netherlands. (1991). *Schoolloopbaan en achtergrond van leerlingen: Cohort 1989*. Voorburg/Heerlen: CBS.

Sund, K. (2009). Estimating peer effects in Swedish high school using school, teacher and student fixed effects. *Economics of Education Review*, *28*, 329–336.

[5] Sund (2009) finds empirical evidence for the existence of nonlinear peer effects. Other recent work on peer effects includes Ammermüller and Pischke (2009) and Eisenkopf (2010).

[6] Betts and Shkolnik (2000) find some evidence that schools would be more likely to allocate highly educated teachers to higher ability classes. Rivkin, Hanushek, and Kain (2005) and, more recently, Hanushek (2011) point out the central importance of teacher quality for student achievement.

Please cite this article in press as: van Elk, R., et al. Does the timing of tracking affect higher education completion? *Economics of Education Review* (2011), doi:10.1016/j.econedurev.2011.04.014