The Impact of Early Tracking on Later-life Outcomes: An Instrumental Variable Approach

Lex Borghans[*], Ron Diris[†], Wendy Smits[‡], and Jannes de Vries[§]

October 26, 2012

Abstract

Educational tracking is used in many nations as a means of dealing with heterogeneity in ability. This paper exploits within-country variation in the age at which tracing takes place, to estimate its effect on later-life outcomes. Students in the Netherlands can be tracked at any age between 12 and 15, depending on school policy. We employ the relative supply of early tracking schools in each municipality as an instrument for early tracking. Outcome variables on educational attainment and labor market success are based on administrative data. Early tracking leads to overall negative impacts on educational attainment and earnings. This negative effect reduces if we move upward in the distribution, but only becomes positive for students of very high ability.

Keywords: economics of education, early tracking, labor market success

JEL Classification: I21, J24

[*]Department of Economics and Research Centre for Education and the Labour Market (ROA), Maastricht University, lex.borghans@maastrichtuniversity.nl.

[†]Department of Economics, Catholic University Leuven, ron.diris@kuleuven.be

[‡]Centraal Bureau voor de Statistiek, w.smits@cbs.nl

[§]Centraal Bureau voor de Statistiek, j.devries@cbs.nl

1 Introduction

The age at which students are selected into different school tracks constitutes an important source of variation in educational systems worldwide. The choice of early versus late tracking is a choice between a more directed school curriculum and a more homogeneous peer group versus the possible positive interactions from having a peer group of heterogeneous ability. Moreover, early tracking might increase the risk of misallocating students to an inappropriate school type. Since skill development is a dynamic process and can exhibit strong complementarity over time,[1] even a difference in tracking age of only one year can have strong implications for subsequent skill development and a large range of later-life outcomes.

This paper estimates the effect of early tracking on several measures of educational attainment and labor market outcomes. We use an instrumental variable model where we employ the relative supply of early tracking schools in the municipality as an instrument for early tracking. We use Dutch data from a secondary school study, matched with administrative data on tertiary education and the labor market. We find that early tracking has a strong negative impact on educational attainment, and this is translated into lower earnings for some subgroups as well. The negative effect of tracking on these outcomes reduces when we move upward in the distribution. However, it only becomes positive for the very highest ability level. The effect of early tracking on earnings can be attributed to a decrease in employment and hours worked, rather than to a decrease in the hourly wage. We employ several robustness checks that provide evidence in favor of the validity of our instrument.

Developed nations track students as early as age 10, while some have a comprehensive system throughout secondary school. The Dutch educational system is in between these extremes, although it is generally seen as an early tracking system. Children in the Netherlands enter secondary school at age 12, but actual track assignment can still be postponed for one or two years. This induces variation in the tracking age within the Netherlands that can be exploited. Where other studies rely on between-country differences[2] or changes in a country's tracking system over

[1] See, e.g. Cunha and Heckman (2007); Cunha et al. (2010).

[2] See e.g. Hanushek and Woessmann (2006); Jakubowski (2010).

time,[3] we look at differences between students in the same country at the same point in time. Moreover, we specifically look at long-term outcomes, in the form of educational attainment, earnings, employment and use of welfare.

This paper is organized as follows. Section 2 gives a brief overview of relevant characteristics of the Dutch educational system. Section 3 presents a literature overview, in which we discuss previous research in the field. Section 4 discusses the data we use in the empirical analysis. The methodological approach of our analysis is explained in section 5. Section 6 presents the main results, after which robustness issues are discussed in section 7. Section 8 concludes.

2 Dutch educational system

2.1 Track assignment

The Dutch Educational System is characterized by relatively early tracking, namely at the age of 12. Students are selected into three main tracks: *vmbo* (vocational), *havo* (higher general) and *vwo* (pre-university). They can still drop to a lower track in higher grades, in case of bad performance, but moving to higher tracks is only possible when the current track has been completed.[4] The current *vmbo* track used to consist of the separate *lbo* and *mavo* tracks. These tracks were joined in 1999, although there is still selection within *vmbo* between theoretical and practical education. For the remainder of this paper, we refer to these tracks as T1 for the (lower) *vmbo* track (or T1a and T1b for *lbo* and *mavo*, respectively), T2 for the (intermediate) *havo* track and T3 for the (higher) *vwo* track.

The key determinant of track selection is the Cito test, which is employed by more than 80% of Dutch primary schools (Inspectie van het Onderwijs, 1997). The Cito score leads to an advice, given by the primary school, for any of the tracks. Students and their parents can decide to deviate from that advice.[5] In the majority of cases the advice corresponding to the Cito score is adhered. There is additional room for students to be selected into a track that does not correspond to their

[3] See e.g. Galindo-Rueda and Vignoles (2005); Pekkarinen et al. (2009)

[4] Attending *havo* is only possible after completing the highest of four *vmbo* levels.

[5] If secondary schools allow for this, which is not always the case.

Cito score, because selection is often postponed until the second or third year of secondary education. This occurs through the existence of comprehensive 'bridge-grades', where students of two or more tracks are still kept together. This is most common for Track 2 and Track 3 students, while students in Track 1 are generally selected early.[6]

The use of bridge-grades induces variation across the country in the actual age at which students are being tracked. It is this specific variation that we will use in estimating the effect of tracking age on measures of educational attainment and labor market outcomes.

2.2 Post-secondary education

The Dutch educational system has three levels of post-secondary education. The lowest level is *mbo*, which has a vocational orientation. *Mbo* has four main levels, which last one to four years. Higher education can be divided into two categories. *Hbo* provides higher professional education, while *wo* consist of university education. Students with a high school diploma from the T2 track or higher can enter *hbo*, while *wo* is only available for students who completed T3.[7] Hence, students who are selected to a specific T1 or T2 track lose the direct opportunity to attend *hbo* or *wo*, respectively. Completing the highest level of *mbo* makes one eligible for entering *hbo*. Hence, it is still possible for T1 students to complete higher education, although the route is less direct.

3 Literature

Previous literature on the effects of tracking generally focuses on its effects on inequality, often in relation to efficiency. We can distinguish studies that use across-country differences from those that use within-country variation in tracking over time. Cross-country studies generally make use of difference-in-difference estimation, by looking at differences in test score between countries, both before and after tracking has taken place. Hanushek and Woessmann (2006) use differences

[6] In our sample, roughly 90% of T2 and T3 students is in a comprehensive grade for at least one year, and 35% for 2 years. About 75% of students in Track 1a and 1b is in a specific track at age 12 already, which contains almost all T1a students.

[7] Completing *hbo* also gives access to university education.

in tracking ages across countries and find that early tracking increases inequality, which does not seem to be compensated by any gains in efficiencies. These studies assume that differences would have remained stable when tracking would have occurred at the same age. Moreover, Jakubowski (2010) points out that there are differences in relative age levels between the studies Hanushek and Woessmann (2006) use, although he largely confirms their findings that tracking reduces mean performance and comes especially at the expense of students of lower ability. Ariga and Brunello (2007) contrast these findings. They use an IV-approach, employing reasons to drop out from school as an instrument, and find that the time spent in a track is positively related to performance on achievement tests.

A number of other studies has looked at variation in the design of a tracking system over time. Pekkarinen et al. (2009) use a change in the Finnish education system from multi-track to one track to assess the effect of tracking as a whole. The study uses variation in the exact timing of the policy change across municipalities to identify this impact. Pekkarinen et al. (2009) find no impact of tracking on several measures of educational attainment, test scores or income. Subgroup analysis shows that the policy change had beneficial effects for reading subscores and for those with low parental education. A similar policy change in England has been evaluated by Galindo-Rueda and Vignoles (2005). They find positive effects of tracking on high-ability students, and an absence of negative effects for the rest of the distribution. They use the political constituency of the area as an instrument, on the ground that Conservative areas are more likely to postpone conversion to the comprehensive system (which was done gradually across the country). However, Manning and Pischke (2006) question the validity of their IV approach, finding a similar impact with this method when they use test scores obtained before tracking, which naturally should be unaffected.

There has been special interest in whether tracking reinforces the effect of family background, thereby causing increased segregation and hampering equality of opportunity. Schnepf (2003) finds that family background, gender and degree of urbanization have a significant impact on the track someone ends up in, independent of ability. Waldinger (2007) finds that tracking reinforces the effects of family background on student performance, but he identifies that this is solely due

to primary school selection and not exacerbated by secondary school tracking. This underlines the identification problem in estimating the impacts of (early) tracking. Possible reasons of negative effects of tracking, especially on students of lower ability, can occur for a number of reasons, including peer effects, teacher sorting and misallocation of students. As of yet, there is no clear consensus on which of these issues is the strongest contributor to any possible negative effects, since disentangling all these effects is troublesome.

In the literature, tracking is often used interchangeably with the more informal practice of ability-grouping, which refers to the splitting up of students according to ability, for certain school subjects. The overall findings do not differ strongly across definitions, although they appear to be slightly more favorable for ability-grouping. Brunello and Checchi (2007) find that tracking has negative impacts on the intergenerational mobility of differences in educational attainment and early labor market outcomes, but positive impacts for literacy and participation in on-the-job training. Moreover, studies have identified positive effects for high-ability students (Epple and Romano, 2002) and an absence of statistically significant effects on inequality.[8]

This paper employs an Instrumental Variable approach to estimate the effect of tracking age, using regional variation in the supply of schools that track early. As such, it has similarities with the study of Van Elk et al. (2011). They use the relative supply of early tracking schools by degree of urbanization, which is divided into nine categories, as an instrument for early tracking. They estimate the effect of early tracking on completion of higher education, for the subgroup with a secondary school advice for lower general secondary education. We use the relative supply of early tracking on the level of the municipality as an instrument. An advantage of this study over other papers that examine the effect of the tracking age, which we share with Van Elk et al. (2011), is that we are looking at variation in the tracking age *within* a country, at the same point in time. Thereby, we avoid relying on differences in tracking across countries, which can be highly endogenous. Moreover, we examine a long-established tracking system, in contrast to studies that look at a recent reduction or expansion in the number of tracks. Furthermore, we estimate effects on long-run outcomes, where research on tracking generally relies on test scores.

[8] See e.g. Slavin (1990), Figlio and Page (2002).

4 Data

We employ data from the national database of Statistics Netherlands (*CBS*). The dataset contains a post-primary education study (*VOCL*) combined with administrative data, including demographic information and labor-market data. The individuals that participated in *VOCL* are retrieved in the administrative data. The available data include test scores at the start of secondary education and two years later, information on the educational career of each individual (secondary education track, years of schooling, highest degree etc.) and labor market status from 1999 until 2007 (employment status, earnings etc.). Hence, we are able to examine what the effect of tracking age is on educational attainment and earnings after multiple years. Data are available for the *VOCL* cohorts of 1977, 1982, 1989, 1993 and 1999. The cohorts are named after the year in which the included individuals entered secondary education. This is when the *VOCL* study started, and students are 12 years old. We will only use the 1989 and 1993 cohorts, since our data on the supply of early tracking schools is based on national data from 1997. Going back as far as 1982 would lead to a too inaccurate instrument. On the other hand, the students in the 1999 cohort are too young to have entered the labor market. Hence, we have labor market information ranging from those who just started to work up until people who are 30 years old (the 1989 sample in 2007). The cohorts also contain test scores from year three of secondary school. However, the exact tests differed across tracks which is why we do not employ them as outcome variables. Summary statistics are provided in table 1.

Our main indicator for educational attainment is completion of higher education. This constitutes both completion of higher professional education (*hbo*) or university education (*wo*). We also employ completion of university education as a separate outcome variable, as well as dropout, which refers to only completing primary education. We only use the 1989 cohort for our estimation of labor market outcomes. Individuals in the 1993 cohort are aged 26 and younger. Some observations are still in education, or at the very start of working life. Including them will undervalue the earnings potential of the most well-educated individuals and thereby not give a proper picture of the true impacts of early tracking. The 1989 cohort is aged 27 to 30 during the years we include in

our earnings estimation.[9] We use the raw reported monthly earnings as an outcome, as well as corrected earnings. The latter correct the monthly wage for the part-time factor, exclude wages below the minimum wage and topcode wages at 10,000 euro's per month. Other labor market outcomes we use are: a dummy 'employment' variable which takes the value of 1 if the individual earns at least a minimum (monthly) wage, and a dummy that indicates whether the individual was ever on welfare from 1999 to 2007.[10] We use both cohorts when we estimate impact on welfare benefits, because the difference in the gradient between those with different levels of education should not be an issue here.

We have information on the exact school type each individual attends for all years of secondary education and thereafter. We can identify when students enter a non-comprehensive track for the first time. This constitutes their tracking age. A student is in a non-comprehensive track if he or she is separated from the other track under examination. When we are looking at tracking between the low (T1) and medium (T2) track, this implies that a T1 student is separated from T2 students and vice versa. The T2 student can still be in a comprehensive class together with T3 students. The same reasoning applies for tracking between medium and high track. We also observe the type of education students are in after they finish secondary education. Thereby, we can follow the educational levels they complete until the year 2007, when the 1989 cohort is 30 years old and the 1993 cohort is 26 years old.

Summary statistics can be seen in table 1.Government reports show that the share of students aged 25-34 that have a higher educational degree is 34 percent in 2007 (Ministerie van Onderwijs Cultuur en Wetenschap, 2011). Hence, the overall sample is very representative of the total Dutch population in this age group, although 1989 is slightly less and 1993 slightly more affluent. Figure 2 shows where each advice group ends up in terms of educational attainment. It shows that higher education (the sum of the two right most bars) is completed by students of all subgroups, but that this strongly increases by advice group. The prevalence of higher education completion is more than twice as high in the subgroup with T2 advice than for those with T1b advice. This reflects,

[9] We use the earnings data from 2004 to 2007, since the vast majority of individuals in the 1989 cohort have finished education by then.

[10] This includes unemployment benefits and disability benefits.

besides the higher level of T2 students, also the easier access to higher education for T2 students.

5 Methodology

5.1 Defining early tracking

We estimate the effect of being tracked early or late on different levels of the Dutch educational system. Since the Dutch system has four school tracks, there are three levels at which students can be tracked or held together: between track 1a and 1b; track 1b and track 2; and track 2 and 3. For ease of notation we refer to tracking between tracks 1b and 2 as Track 1/Track 2 tracking, or T1/T2 tracking. Correspondingly, the other levels of tracking are labeled as Track 1a/Track 1b tracking and Track 2/Track 3 tracking, or T1a/T1b and T2/T3 tracking.

We will not estimate the effect of T1a/T1b tracking since the variation in tracking age is too low at that level to ensure a powerful first stage. The vast majority of students in these two tracks are already separated from each other at age 12, and only few thereafter. Tracking at age 12 is very rare in the case of T2/T3 tracking. Only 9% of students who end up in T2 or T3 are already in a specific track by age 12. Hence, early tracking is defined at this level as at age 13 or earlier. The definition of early tracking is less straightforward for T1/T2 tracking. There is a fairly even division of tracking at ages 12, 13 and 14. Choosing a division at age 12 or thereafter provides a stronger first stage for low-ability subgroups, but choosing age 13 or thereafter produces strong F-values for a larger part of the distribution.[11] Since we are looking at effects for all subgroups, we employ the latter. Hence, we define early tracking at the T1/T2 level as before age 14, or the third year of secondary education.[12]

[11] This is mainly because the T1/T2 students of relatively higher ability are often tracked in T1/T2/T3 comprehensive classes, which generally do not track at age 12.

[12] Van Elk et al. (2011) choose age 12 or thereafter as a division point. This produces stronger instrumentation for the T1b subgroup, which is the subpopulation that comprises their sample.

5.2 Endogeneity

Whether you attend a school that tracks early is, to a large extent, a free choice of the student and his parents. Hence, it is an endogenous variable which could be strongly correlated with several other characteristics that affect outcome variables such as educational attainment and earnings.

Table 2 shows the means of the groups that are tracked early and those who are not, for tracking on both the T1/T2 and T2/T3 level. The fourth and eighth column of table 2 give the p-value of a two-sided Wald test on whether these means are equal. The table clearly shows that those who are tracked early have less favorable observables in case of T1/T2 tracking, and more favorable in the case of T2/T3 tracking. This is reflected by significant differences in Entrance Test scores, parental education, social class and ethnicity. The difference are more pronounced for T1/T2 tracking. Table 2 shows that high-ability students and their parents have a preference for being in a separate T3 track, while the relatively highly-able students at the T1/T2 level prefer to be in a mixed class to be able to make it to the T2 track. The majority of early tracked students in the T1/T2 sample is tracked to a T1 track. This is partly because the T1 track is much larger and partly because a comprehensive track is less attractive for future T1 students that do not have any expectations of moving to a higher track.[13] This division already contributes to a difference in ability between students in comprehensive and tracked schools at this level.[14] The table also lists the outcome variables we employ. They show that the early tracked students are generally worse off for both levels of tracking. For T2/T3 tracking, these students are also of higher ability, suggesting that the actual effect is even stronger. The differences for T1/T2 tracking are often severe, but given that non-tracked students are of better ability, the true effects can be substantially lower.

5.3 The IV model

Hence, one needs a source of exogenous variation to identify the effect of early tracking. We employ Instrumental Variable estimation to correct for endogeneity. We construct separate instru-

[13] In many cases, they will not even be allowed to enter by secondary schools.

[14] Ability refers to less favorable observables here, as shown in table 2, which suggests relevant unobservables are less favorable as well.

ments for T1/T2 tracking and T2/T3 tracking. When there is at least one student in the school that is tracked separate from the other track under examination, this school is classified as a tracking school at that specific grade level. Our instrument measures the fraction of *students* in a municipality that are in a school where that specific track is offered separately before age 14. This is the municipality where the school resides rather than the student. These values are not based on the sample at hand, but determined on a national level. Graphs 3 and 4 show that there are multiple municipalities where the instrument takes either value 0 (no tracking anywhere) or 1 (all schools track), but there are many values in between that range as well. Our explicit assumption is that the relative supply of early tracking schools on the municipal level is not influenced by demand side variables. Figure 5 shows the relation between mean Entrance Test scores on the municipal level and our instrument, for tracking on both levels. The figure shows that there is no relation between a municipality's average student level and the degree of early tracking in that municipality.[15]

We employ a long list of control variables. We compare results with and without controls to get an indication of the validity of our instrument. Our control variables are: month of birth, gender, ethnic origin,[16] social class, highest education level of the parents,[17] whether the child lived in one of the big 4 cities in the Netherlands,[18] the score on an intelligence test taken in year 1 of secondary school and the scores on all three test domains of the Entrance Test, also taken at the beginning of secondary education.[19] Our model then becomes:

$$D_i = \beta_0 + \beta_1 w + \beta_2 X' + \epsilon_i$$

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X' + \epsilon_i \tag{1}$$

D indicates whether the child is tracked early or not. The variable *w* is our instrument, indicating the relative supply of early tracking schools at the municipal level. X' is a vector of controls

[15] This is confirmed by formal estimation of this correlation, which proves to be insignificant.

[16] Defined as a dummy variable that takes the value of 1 if the child or his parents were not born in the Netherlands.

[17] We employ two dummy variables. One takes the value 1 if the highest educational level of the parents is in the top two categories, the other when it is in the lowest two categories (six categories are defined).

[18] Amsterdam, Rotterdam, The Hague and Utrecht

[19] These domains are language, mathematics and information processing. The scores are entered separately for each domain.

(described above) and ϵ is our error term. We employ robust standard errors, which are clustered on a school level. Y is our outcome variable. These outcomes are measures of educational attainment or labor market outcomes.

In addition to overall effects, we estimate effects by subgroup. We categorize these subgroups by the track advice they receive at the end of primary school. This advice is strongly based on the Cito score, but can deviate from that in some cases. This occurs when the teacher believes this score does not reflect the students true ability. Using these groups allows us to see the impacts of early tracking on different types of students, which are essentially ranked according to ability. Schools and teachers can give specific track advices or mixed advices. Figure 1 shows where the students of these different advice groups end up at their point of actual track selection. Adherence to the T1 and T3 advices is very strong. This is lower for the smaller T2 track, where there is mainly upstream towards T3. This is also the case for the mixed T2/T3 advice, which most often leads to assignment towards T3.

Table 3 shows the F-values of a Wald test on the instrument, measuring the strength of the instrument. We report this separately for educational attainment and labor market outcomes. The rows list the values for the overall samples and the subgroups according to secondary school advice. We only use the 1989 cohort in our estimation of labor market effects, which naturally makes these F-values smaller. Our instrumentation is strongest for T2/T3 tracking, especially for educational attainment. The informal critical value[20] of 10 is not reached in T1/T2 tracking for some subgroups with respect to educational attainment. The overall sample does not pass the critical value for labor market outcomes and only one subgroup does. We report the effect for these estimations nonetheless, but they are very imprecise and carry less weight.

6 Results

We estimate the effect of early tracking on several outcome variables, which we organize in two categories: educational attainment and labor market effects. In the first category, we look at com-

[20]First proposed by Staiger and Stock (1997).

pletion of higher education, completion of university and school dropout (not finishing secondary education). Labor market outcomes consist of two different measures of earnings, employment and use of welfare. One measure of earnings consists of those earnings reported to the tax authorities. In a second measure, we exclude earnings below the minimum wage, correct for part-time factor and topcode at 10,000 euro's per month.

Results using OLS are portrayed in table 4, for comparative purposes. We expect them to be positively biased in the case of T2/T3 tracking and negatively biased for T1/T2 tracking, based on the comparison of observables in table 2. The OLS estimates for T2/T3 tracking are largely insignificant. We can observe a clear pattern of negative effects for the low ability groups and positive effects for the higher ability groups. The effects for T1/T2 tracking are significant and negative, especially for the low advice groups. The effect of adding controls is in some cases substantial, mainly for the higher advice groups. This confirms our belief that early tracking is endogenous,[21] which is partly picked by our control variables.

6.1 Educational attainment

Table 5 shows the IV estimates of early tracking on multiple measures of educational attainment, both for the sample as a whole as for subgroups according to ability. Early tracking reduces the chance of completing a higher education degree, both on the T1/T2 and on the T2/T3 level. The reduction is approximately 16 percentage points for T1/T2 tracking and 8 percentage points for the T2/T3 group. This negative effect originates primarily from students who got a secondary school advice for the T1 track or a mixed advice for T1/T2 for the former. For T2/T3 tracking, the effects are most prominent for the T1 (which is relatively small), T1/T2 and T2 advice groups. The fact that the impact is larger for the T1/T2 sample can be driven by the fact that early tracking closes off some paths towards higher education for those tracked to a T1 track early. The vast majority of early tracked students in the T1/T2 sample end up in a T1 track. For the subgroup with a T2/T3 advice, the effect is insignificant but still negative. The only positive effect occurs

[21] The fact that this is also present in our subgroup estimates indicates that there is also endogeneity within these groups. Hence, the better students with T3 advice are more likely to be tracked early than students of relatively lower ability with the same advice.

for the highest advice group, but is far from significant.[22] The effect of early tracking gradually decreases for higher ability levels, which is in line with previous literature (e.g. Hanushek and Woessmann (2006)). The coefficients in the next two columns suggest that this reduction in the completion of higher education can largely be attributed to higher completion rates of higher professional education, rather than completion of university. Effects are negative here as well, but never reach significance. The sign chance occurs one category earlier than for higher education. The last two columns present suggestive evidence that part of the negative impact on educational attainment occurs through an increase in dropout. Although the coefficients are not significant, they are consistently positive and close to significance for the overall sample and the T2 advice group for T2/T3 tracking and the T1/T2 advice group for T1/T2 tracking.

If we compare the columns with and without controls, we observe a very strong consistency in coefficients, for all outcome variables and both levels of tracking. All differences are neglectable and stay within half of a standard error even in the most extreme cases. This is suggestive evidence that our instrument is indeed exogenous, since it does not correlate strongly with those determinants of the outcome variable that are observable. There can still be a correlation between the instrument and unobservables, but given that our range of control variables is vast and offers multiple proxies for relevant unobservables, the fact that adding these controls has such a small effect over such a wide range of coefficients is a strong indication of instrument validity.

The estimates of Van Elk et al. (2011) are not easily comparable to our results. They employ a different age of what constitutes early tracking[23] and focus on one specific subgroup for which our instrumentation is weak.[24] Their IV-estimates suggest a decrease of 13 percentage points in the completion of higher education for the T1b advice group. These estimates are even higher in the robustness analyses, with a maximum effect of 21 percentage points. This is roughly equal to our overall estimate for T1/T2 tracking, as well as our subgroup effect for the T1b advice group.

[22] This can be partly due to the fact that the vast majority of students in this group attains high education anyway. Moreover, the T3 advice group is relatively small.

[23] We employ before age 14 as early tracking, while Van Elk et al. (2011) use age 13.

[24] Because the vast majority of students in this particular subgroup is tracked early, the instrumentation is weak both when we choose age 12 as a cutoff or age 13. This is exacerbated by the discrepancy between the frequency of early tracking in the sample, and the nationwide average where our instrumentation is based on. Van Elk et al. (2011) construct the instrument with the sample at hand, which increases their first stage power.

However, the instrumentation is weak for the latter. We can estimate the effect for this subgroup in a less robust way, constructing the same instrument using only the sample data. This results in an effect of early tracking for those with T1b advice of -0.066,[25] or a lowering of higher education completion of 6.6 percentage points. This estimate is, although still negative, clearly lower than the one in Van Elk et al. (2011).[26]

6.2 Labor market outcomes

Table 6 presents results for several outcomes relating to labor market performance. As table 3 already showed, the instrumentation is too weak for estimation of tracking on the T1/T2 level on labor market outcomes. We report these results for completion, but they need to be interpreted with care. For T2/T3 tracking, the first stage is powerful enough for all estimations besides the subgroup estimation for the students with T3 advice.

When we look at the overall effects and the effects for subgroup T2 (the only subgroup with a powerful enough first stage here) for tracking on the T1/T2 level, the effects are positive but insignificant. This contrasts with the strong negative results on educational attainment, and might suggest that higher degrees have not translated into higher earnings. However, we cannot draw any conclusions here, given the low first stage power.

We identify strongly significant and negative impacts for T2/T3 tracking. These are present for reported earnings and on employment, but not for corrected earnings. This indicates that the effect of early tracking is mainly due to higher employment and hours worked, since it disappears when we exclude low earnings and correct for part-time factor.[27] The effect originates mainly from the subgroup of students with a T2 advice. The earnings of the T2/T3 advice group are negatively affected by early tracking as well, while the impacts on educational attainment are low and insignificant for this subgroup. This could be indicative evidence of the effect of peer quality on later earnings. The results again show that the only positive coefficients are present for the very

[25] Significant at the 1% level.

[26] The mean difference between both groups is 10 percentage points for the subgroup with T1b advice. Observables are slightly in favor of comprehensive students which suggests that the ATE will be below 10 percentage points. Still, the LATE could exceed this.

[27] The corrected earnings also topcode very high wages, but this proofs not to make a difference.

highest subgroup, but these effects do not reach significance. We also estimate the effect of early tracking on earnings in the latest reported year (2007). These estimates (not shown) are highly similar as those for the mean earnings over 2004 to 2007. They are slightly more negative for both levels of tracking (with the exception of the highest subgroup in each case, where the effects become more positive), which is what we would expect when age increases. We also see here that adding controls has a very small impact on coefficients, for all outcome variables and both levels of tracking.

We also examine whether someone was ever on welfare from 1999 to 2007. The effect for T1/T2 tracking is negative, but insignificant. Table 2 showed that the mean difference was significant and higher for the comprehensive group. Given the fact that comprehensive students are of higher ability, this suggests that welfare receipt is even higher when tracked late, which is in contrast to results for educational attainment.[28] The overall effect size is insignificant for T2/T3 tracking but this hides a significantly positive effect for the T2 advice group[29] and a significantly negative effect for the T3 advice group. This confirms earlier patterns of how the effect of early tracking develops over the distribution, although this is the only variable that is significantly affected for the highest advice group.

Our estimation of labor market outcomes is weakened by the fact that we do not use the 1993 cohort in the previous analyses. We also estimated results when both cohorts are included. We only use earnings for the most recent year, 2007, in this estimation. The 1993 observations are 26 years old then. We obtain estimates (not shown) that are lower in magnitude than when we only use cohort 1989. This is not surprising, since the students from comprehensive classes have a higher probability of still being in education. Still, the effect for the T1/T2 advice group in T2/T3 tracking is significantly negative. The effect for the highest advice group is positive and close to significance. Given the fact that the early tracked students are *more* likely to still be in education here,[30] this effect could become even larger once all students have entered the labor market. All other subgroups are not significantly affected, but this could change when individuals mature in

[28]The fact that IV estimates are insignificant is mainly driven by relatively high standard errors. The effect sizes are larger than the mean difference in table 2, as one would expect given the direction of the bias.

[29]Although the significance disappears when controls are added.

[30]Based on the insignificant but positive estimates from table 5.

the labor market.

6.3 Magnitudes

Compared to the OLS estimates from table 4, the IV-estimates are more negative for T2/T3 tracking. This is exactly what we would expect when high-ability students are more likely to go to a specific track early on. On the other hand, the IV estimates are more negative for T1/T2 tracking as well, where the observable characteristics of the comprehensive students are clearly better. However, OLS estimates refer to an average treatment effect, while the IV coefficients are the local average treatment effect (Imbens and Angrist, 1994). This is the effect of early tracking on those students who would respond to a change in the supply of schools that track early. It is expected that the students who have the ambition to enter higher tracks or complete higher education, are the ones that would respond strongly to an increase in comprehensive classes. These classes increase the opportunity of attending high education. Even considering that our effects are LATE, the impact of early tracking on the T1/T2 level on high education are large. A large part of this effect could be driven by the institutional design. For students in the T1b track, which constitutes the large majority of T1/T2 students, it is less straightforward to enter higher education.[31] The path to higher education is more direct when they first go to a comprehensive track. We have seen that the advices given at the end of primary school are conservative; a lot of students that go to comprehensive classes end up in higher tracks.

We also saw that there is large upstream for students with a T2 advice from primary school. This can explain why the results for T2/T3 tracking are especially large for this subgroup (along with the T1/T2 subgroup). Early tracking has strong negative effects for this group on completion of higher education, earnings and employment, while the treatment effects for completion of university, dropout and welfare receipt are not far from being significant. It is likely that a large part of these effects can be attributed to misallocation from early selection. More than 30% of students with a T2 advice still go to Track 3. Naturally, such upstream is less likely when selected early.

[31] They can still achieve higher education, if they follow the T2 track after the T1b track, or when they complete the highest level of lower vocational tertiary education. This ensures that still 22 (1989 cohort) to 27 (1993 cohort) percent of T1b students ultimately completes higher education.

We do not find significantly positive estimates of early tracking for any subgroup. However, our results are indicative that early tracking at the T2/T3 level increases outcomes for the T3 advice group. This subgroup is small (n=724) and the first stage is not strong. The effects for these students are consistently positive and close to significance in some cases. Moreover, they become larger when we use more recent earnings estimates and estimates are significant for welfare receipt, which has a more powerful first stage. Since these are highly educated individuals, the labor market gains from early tracking can strongly increase after age 30 for this subgroup.

7 Robustness

We have seen that adding control variables has a very small effect on our estimates, which is evidence in favor of the validity of our instruments. In addition, we check whether our model finds any estimates for variables that should not be affected. We use all three domains of the Entrance Test at the beginning of secondary education and whether one of the parents completed one of the highest two educational attainment levels.[32] We find that the effects of early tracking on these variables is far from significant, as it should be. The t-statistics are all lower than 1, both with and without controls and for T1/T2 and T2/T3 tracking.

Our approach assumes that the supply of schools in the municipality is the choice set you face. If mobility of students is high, this might not be a valid assumption. It could be that individuals of a specific ability level are more likely to commute to a school outside of their municipality of residence. This can induce a correlation between the instrument and unobserved ability. To address sensitivity, we use the supply of early tracking schools by municipality of residence, rather than municipality where the school resides, as an alternative instrument. We employ the latter as an instrument in our main analysis, because there are multiple municipalities without any school that offers the specific track. Hence, using municipality of residence weakens our first stage. If such behavior occurs in practice, this should lead to differences in estimates from using these different instruments. Estimates with the alternative instrument are portrayed in column 2 of table

[32] Naturally, these variables cannot be used as controls in the respective regressions.

7. Since instrumentation is weaker in this case, standard errors are higher. The estimates for higher education completion for T1/T2 tracking are somewhat higher under the alternative specification. The latter difference is much smaller when we run the original model on this reduced sample. Hence, the change in sample size drives a large part of this difference. All other coefficients for T1/T2 tracking, and all those for T2/T3 tracking, are very robust. This indicates that mobility between municipalities does not drive our results.

One possible concern is that the supply of early tracking schools can be driven by the demand for tracking. Schools in municipalities with students of a higher than average level might be induced to offer early tracking if these better able students and their parents have a preference for early tracking. We include two variables in our model that are indicators of such differences in the demand for early tracking schools. These are the mean Entrance Test score and mean parental education level, both defined per municipality. Table 7 shows that the effect of adding these variables on coefficients is virtually non-existent.

Van Elk et al. (2011) address robustness by adding the change in the supply of early tracked schools over time as an extra control. This change might reflect changes in the demand for these schools. However, our instrument is based on national figures from 1997 and has the same value for both cohorts. We can construct this variable based on the supply from the sample at hand. This reduces the number of observations and first stage power considerably, because not all municipalities occur in both cohorts. Therefore, the coefficients from column 4 in table 7 are not the same as the original estimates. Comparing to this benchmark, the effect of adding this ratio is very minimal.

Our instrument proves to very robust to any robustness analysis employed. This is somewhat in contrast with other studies that use Instrumental Variables when estimating the effects of tracking. Manning and Pischke (2006) show that the use of political constituency as an instrument in Galindo-Rueda and Vignoles (2005) is invalid, since the Conservative areas are relatively more wealthy and have better performing students. Moreover, the approach used in Galindo-Rueda and Vignoles (2005) also gives significant estimates on predetermined variables. This is not the case for our instrument, which does not lead to significant estimates for variables that should be unaffected.

Moreover, it is not correlated with any of the control variables we employ, while the estimates from Van Elk et al. (2011) change significantly when controls are added. This suggests that areas with different density or political constituency will differ more systematically than municipalities with a different relative supply of early tracking schools.

8 Conclusion

We have identified that early tracking leads to negative impacts on educational attainment and labor market outcomes. These impacts are more pronounced for those with lower ability. Treatment effects of early tracking decrease in magnitude as we move up in the distribution, but they are only positive for the very highest subgroup. Both tracking at the T1/T2 level as well as on the T2/T3 level lead to negative treatment effects. The effects are stronger for educational attainment for the former, but our instrumentation at the T1/T2 level is not powerful enough to make a similar claim with respect to labor market effects.

We use an instrumental variable approach where we exploit difference in tracking age across municipalities to identify the effect of early tracking. The relative supply of early tracking schools proves to be a strong predictor of early tracking and multiple analyses strongly suggest that it is not correlated with any other determinants of our outcome variables. The effect of adding a large range of controls is virtually non-existent, as is the adding of the mean level of parental education and mean Entrance Test score on the municipal level. Our model also gives very insignificant results when we employ pre-determined outcome variables. The strong robustness of our instrumentation is in contrast with other studies that use IV-approaches in relation to tracking.

The fact that early tracking negatively affects educational attainment does not immediately imply is has a detrimental overall effect. By construction, early tracking puts some subgroups of students already in a track where the path towards completion of higher education is more complicated. However, the reason to put them in such tracks is because they are not deemed fit for a path that leads to higher education. Although the payoffs of attending higher education are generally large and positive, they do not have to be so for every single individual, and maximizing the

fraction of higher education degrees should not be a goal in itself. Still, our estimates suggest that the impacts on educational attainment are translated into lower earnings as well, at least for T2/T3 tracking. This works mainly through a lowering of employment and hours worked. Whether higher earnings compensate sufficiently for the extra time spent in school depends on the utility function of the individual. The effects we find can occur through changes in peer quality, school quality, differences in instruction, or a different allocation of students towards tracks when selection is postponed.

Our estimates for earnings at the T1/T2 level lack the statistical power to draw strong conclusions with respect to the labor market effects for this specific level of tracking. Moreover, we do not have earnings data beyond age 30. Future studies could focus more on estimating effects of the tracking age on earnings, for a larger sample of individuals and a larger age range, to get a better indication of the lifetime consequences of early tracking. Additionally, our study cannot distinguish whether the estimated impacts of early tracking should be attributed to peer effects, curriculum effects or teacher sorting. Nonetheless, the results indicate that providing more comprehensive classes will increase the share of students that obtains a higher education degree, which translates into higher earnings as well.

References

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

Brunello, G. and D. Checchi (2007). Does school tracking affect equality of opportunity? new international evidence. *Economic Policy* (52), 781–861.

Cunha, F. and J. J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Epple, D., E. N. and R. Romano (2002). Ability tracking, school competition, and the distribution of educational benefits. *Journal of Public Economics* (83), 1–48.

Figlio, D. N. and M. E. Page (2002). School choice and the distributional effects of ability tracking: Does separation increase inequality? *Journal of Urban Economics* (51), 497–514.

Galindo-Rueda, F. and A. Vignoles (2005). The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability. CEP discussion paper, Centre for Economic Performance, London School of Economics.

Hanushek, E. A. and L. Woessmann (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *The Economic Journal 116*, 63–76.

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

Inspectie van het Onderwijs (1997). Aansluiting voortgezet onderwijs op het basisonderwijs. Technical report, Utrecht: Inspectie van het Onderwijs.

Jakubowski, M. (2010). *Institutional Tracking and Achievement Growth: Exploring Difference-in-Differences Approach to PIRLS, TIMSS, and PISA Data*.

Manning, A. and J.-S. Pischke (2006). Comprehensive versus selective schooling in england and wales: What do we know? CEP discussion paper, Centre for Economic Performance, London School of Economics.

Ministerie van Onderwijs Cultuur en Wetenschap (2011). Kerncijfers 2006-2010. Technical report.

Pekkarinen, T., R. Uusitalo, and S. Kerr (2009). Schooltracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics 93*, 965–973.

Schnepf, S. V. (2003). Inequalities in secondary school attendance in Germany. S3RI Applications Working Paper A03/16.

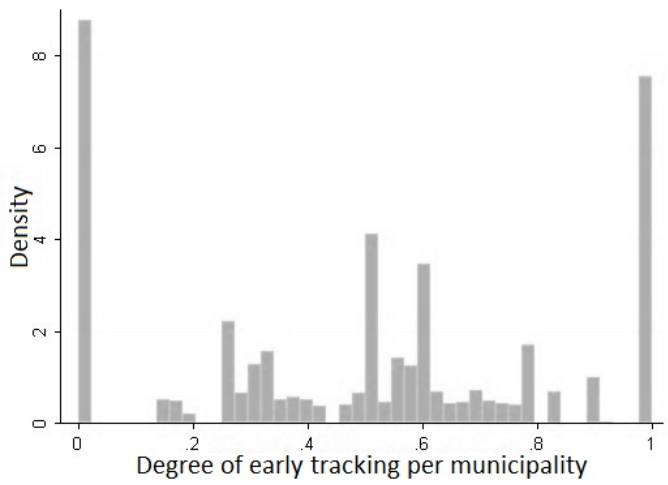
Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research 60*(3).

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

Van Elk, R., M. Van der Steeg, and D. Webbink (2011). Does the timing of tracking affect higher education completion? *Economics of Education Review 30*, 1009–1021.

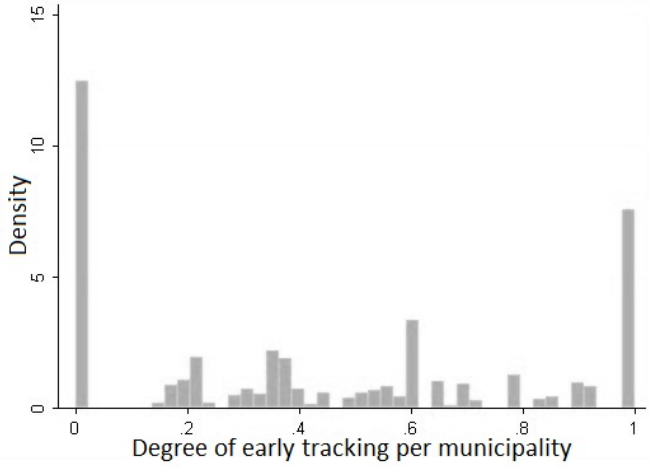
Waldinger, F. (2007). Does ability tracking exacerbate the role of family background for students' test scores? Unpublished Manuscript, London: London School of Economics.

Figure 3: Distribution of early tracking instrument: T1/T2



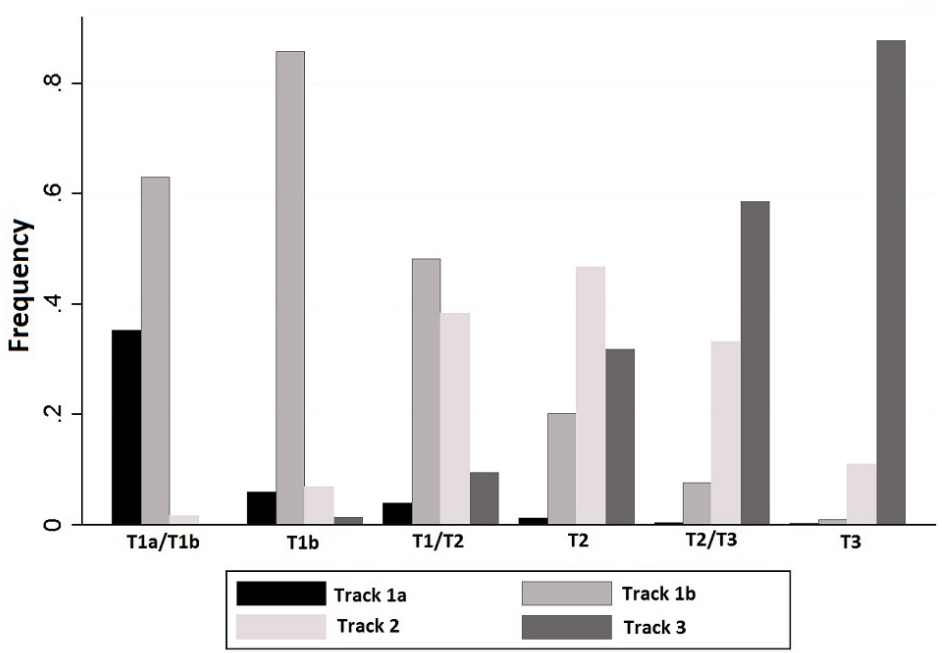
Notes: The graph shows the distribution of instrument values, for tracking at the T1/T2 level. The instrument represents the relative supply of schools that track early (before age 14) in the municipality where the school resides.

Figure 4: Distribution of early tracking instrument: T2/T3



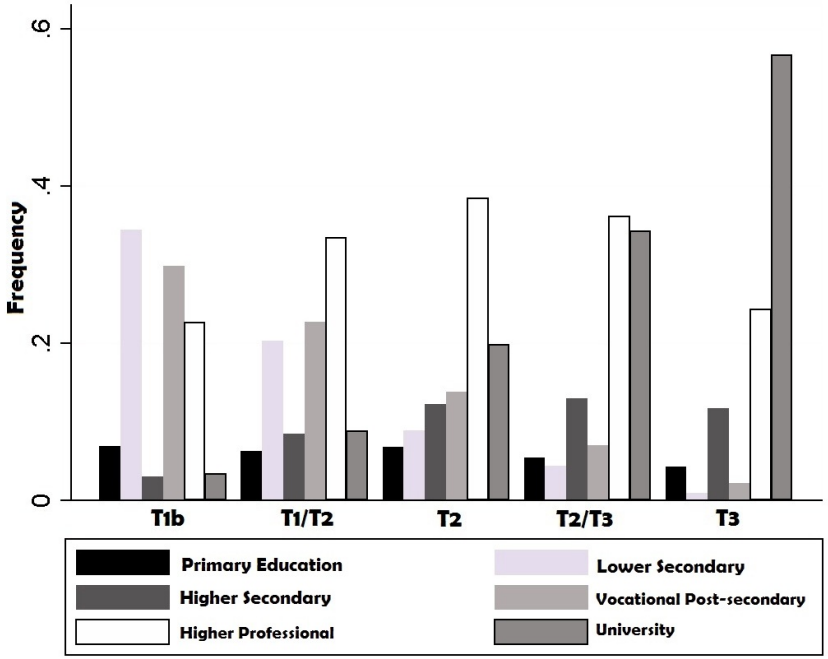
Notes: The graph shows the distribution of instrument values, for tracking at the T2/T3 level. The instrument represents the relative supply of schools that track early (before age 14) in the municipality where the school resides.

Figure 1: Ultimate track assignment per advice group



Notes: The graph shows the ultimate track assignment per track advice received for secondary education. Advices are given by the students' primary school at the end of primary education. The indicators on the horizontal axis indicate the track that was advised for the particular student. Advices can be direct or mixed.

Figure 2: Ultimate educational attainment per advice group



Notes: The graph shows the highest educational level completed per track advice received for secondary education. The indicators on the horizontal axis indicate the track that was advised for the particular student.. Advices are given by primary school at the end of primary education. Data on educational attainment are available until age 26 and 30 for the 1993 and 1989 cohorts, respectively. Advices can be direct or mixed.

Figure 5: Value of instrument by mean test score in municipality

Notes: The Graph represents the relation between the instrument and the mean Entrance Test score in the municipality, for both tracking at the T1/T2 level and T2/T3 level. The Entrance Test is taken at the beginning of secondary education.

Table 1: Summary statistics

	1977		1982		1989	1993
	Mean	Std. dev.	Mean	Std. dev.		
Track 1a (*lbo*)	0.333	0.471	0.359	0.480		
Track 1b (*mavo*)	0.381	0.486	0.323	0.468		
Track 1 (*vmbo*)	0.714	0.452	0.682	0.466		
Track 2 (*havo*)	0.135	0.341	0.152	0.359		
Track 3 (*vwo*)	0.151	0.358	0.165	0.371		
Entrance Test	34.23	11.33	35.20	11.42		
Years of Schooling	13.05	3.83	12.99	3.84		
Earnings 2007	2869.22	1065.80	2455.26	847.81		
Earnings 2006	2696.38	999.41	2266.34	807.70		
Earnings 2005	2426.07	811.36	2044.20	745.81		
Earnings 2004	2316.85	738.55	1915.78	686.53		
Earnings 2003	2209.92	683.26	1763.86	654.31		
Earnings 2002	2044.52	639.20	1544.61	614.55		
Earnings 2001	1868.88	628.76	1314.24	594.49		
Female	0.481	0.500	0.485	0.500		
Big 4 Cities	0.045	0.207	0.060	0.238		
Non-Dutch	0.109	0.311	0.090	0.286		
Lowest Social Class	0.088	0.284	0.088	0.284		
Highest Social Class	0.146	0.353	0.153	0.360		
Parent High Educ.	0.195	0.396	0.220	0.414		
Parent Low Educ.	0.153	0.360	0.086	0.281		
Intelligence Test	0.018	0.832	0.062	0.848		
Early Tracking T1/T2	0.379	0.485	0.351	0.477		
Early Tracking T2/T3	0.136	0.343	0.220	0.414		
Higher Education	0.295	0.456	0.382	0.486		
University	0.104	0.305	0.134	0.340		
Dropout	0.126	0.332	0.129	0.335		
Employment	0.870	0.336	0.847	0.360		
Welfare	0.090	0.287	0.068	0.252		

Notes: Table shows means and standard deviations of all relevant variables. Earnings are per month, pre-tax and in euro's. Reported earnings are corrected for part-time factor, topcoded at 10,000 euro's, and earnings below 1000 euro's are excluded. The variables Track 1a to Track 3 report the fraction of students that go to that track. Track 1 (*vmbo*) is the sum of Track 1a (*lbo*) and Track 1b (*mavo*). Years of schooling is coded to the highest degree obtained. Parent High Education is a dummy variable that takes value 1 when the highest education level of the parents is in one of the top two categories (six categories are defined). Parent Low Education takes value 1 when the highest education level of the parents is in one of the lowest two categories. Social class is based on the occupational status of the parents. 'Big 4 cities' indicates whether the child lived in one of the four big cities in the Netherlands. Higher education consists of higher professional education (*hbo*) and university education. Employment measures whether the individual earned at least a minimum wage in 2007. Welfare measures whether the individual ever received welfare benefits between 1999 and 2007.

Table 3: First stage results (F-values)

	Educational Attainment		Labor Market Outcomes	
T1/T2 Tracking				
All	21.73	26.33	5.03	6.68
T1a/T1b	9.01	8.64	2.66	2.61
T1b	7.53	8.94	0.42	0.89
T1/T2	13.73	14.26	0.72	0.87
T2	33.52	36.85	12.72	14.74
T2/T3	11.65	12.49	2.74	3.33
Controls		Yes		Yes
T2/T3 Tracking				
All	64.94	67.14	30.13	34.67
T1b	20.92	23.13	4.34	5.06
T1/T2	31.67	32.47	11.45	12.78
T2	70.36	73.62	39.47	43.84
T2/T3	42.18	42.06	33.26	37.48
T3	12.34	13.16	5.31	7.87
Controls		Yes		Yes

Notes: The table shows the F-values of a Wald test of the instrument in the first stage of the regression, either with or without controls. The test is conducted for tracking between the low and the medium track (T1/T2), and medium and high track (T2/T3). The left column indicates the subgroup for which the effect is estimated. The indicators refer to track advice they received at the end of primary school, which can be either specific or mixed. The degree of Early Tracking in the municipality is used as an instrument for Early Tracking. Control variables are: month of birth, gender, ethnic origin, social class, highest education level of the parents, whether the child lived in one of the big 4 cities in the Netherlands, the score on an intelligence test taken in year 1 of secondary school and the scores on all three test domains of the Entrance Test. F-values are lower for labor market outcomes, since estimation only includes the 1989 cohort. The treatment effects on use of welfare contains both cohorts and, correspondingly, F-values are equal to those for the educational attainment columns.

Table 2: Comparison of variables for different age of tracking

	T1/T2				T2/T3			
	Mean Comp.	Mean ET	Difference	P-value	Mean Comp.	Mean ET	Difference	P-value
Female	0.505	0.530	-0.025	0.003	0.520	0.521	-0.0015	0.881
Big 4 Cities	0.058	0.030	-0.028	0.000	0.051	0.058	0.0067	0.123
Non-Dutch	0.101	0.091	0.010	0.040	0.099	0.092	0.0067	0.240
Parent High Education	0.258	0.191	0.067	0.000	0.360	0.423	-0.063	0.000
Parent Low Education	0.074	0.098	-0.024	0.000	0.053	0.034	0.019	0.000
Lowest Social Class	0.105	0.099	0.0069	0.178	0.094	0.085	0.0097	0.078
Highest Social Class	0.178	0.140	0.037	0.000	0.239	0.282	-0.043	0.000
Intelligence Test Score	0.166	0.028	0.139	0.000	0.339	0.363	-0.024	0.136
Language Subscore	13.36	12.31	1.05	0.000	14.83	15.40	-0.567	0.000
Math Subscore	13.05	11.60	1.45	0.000	15.04	15.59	-0.554	0.000
Information Subscore	13.45	12.33	1.11	0.000	15.05	15.54	-0.489	0.000
Track 1	0.418	0.807	-0.389	0.000	-	-	-	-
Track 2	0.581	0.192	0.389	0.000	0.555	0.404	0.151	0.000
Track 3	-	-	-	-	0.445	0.596	-0.151	0.000
Mean Earnings (logs)	7.444	7.449	-0.0049	0.618	7.452	7.389	0.062	0.000
Earnings 2007 (logs)	7.671	7.627	0.044	0.000	7.738	7.708	0.030	0.015
Corrected Mean Earnings (logs)	7.719	7.703	0.016	0.002	7.762	7.747	0.015	0.032
Higher Education	0.457	0.299	0.157	0.000	0.692	0.706	-0.014	0.113
University	0.081	0.045	0.036	0.000	0.280	0.362	-0.083	0.000
Dropout	0.065	0.073	-0.0083	0.056	0.060	0.063	-0.0027	0.559
Employment	0.884	0.862	0.022	0.000	0.892	0.862	0.031	0.000
Welfare	0.086	0.069	0.016	0.000	0.088	0.040	0.048	0.000

Notes: The table shows the means of comprehensive (Comp.) and early tracked (ET) students, for both T1/T2 as T2/T3 tracking. The table lists p-values of a two-sided Wald test conducted on the difference between means. 'Big 4 cities' indicates whether the child lived in one of the big 4 cities in the Netherlands. Parent High Education is a dummy variable that takes value 1 when the highest education level of the parents is in one of the top two categories (six categories are defined). Parent Low Education takes value 1 when the highest education level of the parents is in one of the lowest two categories. Social class is based on the occupational status of the parents. All four test scores are taken at the start of the first year of secondary education. The Track variables indicate the fraction of students that end up in this specific track. Earnings are in logs. Corrected earnings are corrected for part-time factor, topcoded at 10,000 euro's, and earnings below 1000 euro's are excluded. Higher education consists of higher professional education (*hbo*) and university education. Employment measures whether the individual earned at least a minimum wage in 2007. Welfare measures whether the individual ever received welfare benefits between 1999 and 2007.

Table 4: OLS estimates of early tracking

	High Education		Mean Earnings	
T1/T2 Tracking				
All	-0.155***	-0.097***	-0.053***	-0.021
	(0.015)	(0.012)	(0.016)	(0.014)
T1a/T1b	-0.026	-0.013	0.019	0.0098
	(0.033)	(0.035)	(0.045)	(0.042)
T1b	-0.094***	-0.082***	-0.035	-0.016
	(0.020)	(0.017)	(0.026)	(0.026)
T1/T2	-0.042*	-0.028	-0.0065	-0.0086
	(0.023)	(0.022)	(0.036)	(0.035)
T2	-0.104***	-0.086***	0.045	0.044
	(0.026)	(0.024)	(0.031)	(0.028)
T2/T3	-0.030	-0.017	-0.025	-0.038
	(0.036)	(0.034)	(0.038)	(0.038)
Controls		Yes		Yes
T2/T3 Tracking				
All	0.010	-0.010	-0.011	-0.021
	(0.014)	(0.011)	(0.017)	(0.017)
T1b	-0.084*	-0.076*	-0.017	0.0026
	(0.046)	(0.040)	(0.065)	(0.069)
T1/T2	-0.041	-0.033	-0.081	-0.061
	(0.030)	(0.030)	(0.049)	(0.046)
T2	-0.029	-0.032	-0.026	0.021
	(0.021)	(0.019)	(0.030)	(0.029)
T2/T3	0.00061	-0.0076	-0.051**	-0.071***
	(0.017)	(0.017)	(0.024)	(0.023)
T3	0.033	-0.016	0.123***	0.118**
	(0.022)	(0.021)	(0.039)	(0.040)
Controls		Yes		Yes

Notes: Table shows the OLS estimates of early tracking on the specified outcome variables, either with or without controls. Estimation is conducted for tracking between the low and the medium track (T1/T2), and medium and high track (T2/T3). High education refers to the completion of either hbo (higher professional education) or university education. Earnings are the average of reported monthly pre-tax earnings from age 27-30 (2004 to 2007). Effects of tracking on earnings are estimated on the 1989 cohort only. See table 3 for a description on the indicators in the left column and the controls included.

Table 5: IV estimates of early tracking: educational attainment

	High Education		University		Dropout	
T1/T2 Tracking						
All	-0.163**	-0.172***	-0.0078	-0.012	0.037	0.032
	(0.066)	(0.050)	(0.026)	(0.021)	(0.026)	(0.023)
T1a/T1b	-0.121	-0.211	-0.038	-0.041	0.057	0.074
	(0.211)	(0.215)	(0.043)	(0.045)	(0.102)	(0.114)
T1b	-0.121	-0.192*	-0.018	-0.026	0.000035	0.020
	(0.136)	(0.115)	(0.042)	(0.039)	(0.067)	(0.055)
T1/T2	-0.147	-0.145*	0.021	0.017	0.071	0.068
	(0.091)	(0.090)	(0.046)	(0.044)	(0.045)	(0.044)
T2	-0.117	-0.091	0.00055	-0.0050	0.027	0.012
	(0.067)	(0.061)	(0.042)	(0.040)	(0.032)	(0.029)
T2/T3	-0.047	-0.092	0.044	0.041	0.030	0.043
	(0.103)	(0.099)	(0.065)	(0.056)	(0.051)	(0.050)
Controls		Yes		Yes		Yes
T2/T3 Tracking						
All	-0.082***	-0.082***	-0.0025	-0.018	0.022	0.017
	(0.030)	(0.025)	(0.041)	(0.033)	(0.015)	(0.013)
T1b	-0.221**	-0.173*	-0.157**	-0.143**	0.023	0.0016
	(0.106)	(0.097)	(0.067)	(0.058)	(0.055)	(0.045)
T1/T2	-0.139*	-0.134*	-0.0027	-0.017	0.024	0.030
	(0.078)	(0.079)	(0.043)	(0.041)	(0.029)	(0.029)
T2	-0.120***	-0.100***	-0.048	-0.039	0.041	0.026
	(0.043)	(0.037)	(0.038)	(0.036)	(0.026)	(0.023)
T2/T3	-0.030	-0.042	0.024	0.0014	0.0073	0.011
	(0.035)	(0.037)	(0.047)	(0.048)	(0.019)	(0.019)
T3	0.018	0.018	0.015	-0.010	0.013	0.0077
	(0.075)	(0.070)	(0.105)	(0.092)	(0.029)	(0.027)
Controls		Yes		Yes		Yes

Notes: Table shows the IV estimates of early tracking on several measures of educational attainment, either with or without controls. Estimation is conducted for tracking between the low and the medium track (T1/T2), and medium and high track (T2/T3). See table 2 for an explanation of outcome variables. See table 3 for a description on the indicators in the left column, the method of instrumentation and the controls included. The first stage is too weak for estimation of subgroups T1a/T1b and T1b for T1/T2 tracking (see table 3), but these effects are reported for completeness.

Table 6: IV estimates of early tracking: labor market effects

	Earnings		Corrected Earnings		Employment		Welfare	
T1/T2 Tracking								
All	-0.0085	0.026	0.081	0.081	0.0086	0.0068	-0.028	-0.036
	(0.126)	(0.090)	(0.078)	(0.060)	(0.069)	(0.059)	(0.028)	(0.025)
T1a/T1b	-0.405	-0.448	-0.197	-0.258	-0.182	-0.154	0.048	0.042
	(0.380)	(0.469)	(0.216)	(0.237)	(0.256)	(0.263)	(0.117)	(0.122)
T1b	0.972	0.630	0.801	0.513	0.735	0.423	-0.143	-0.137
	(1.57)	(0.823)	(1.13)	(0.551)	(1.28)	(0.604)	(0.097)	(0.087)
T1/T2	0.371	0.070	-0.133	-0.035	0.203	0.269	-0.013	-0.023
	(0.658)	(0.341)	(0.281)	(0.170)	(0.276)	(0.286)	(0.037)	(0.036)
T2	0.048	0.036	0.078	0.078	-0.051	-0.053	-0.0015	-0.013
	(0.089)	(0.082)	(0.057)	(0.055)	(0.059)	(0.059)	(0.026)	(0.025)
T2/T3	-0.218	-0.177	0.066	0.043	-0.193	-0.160	-0.022	-0.026
	(0.306)	(0.235)	(0.149)	(0.131)	(0.163)	(0.122)	(0.041)	(0.041)
Controls		Yes		Yes		Yes		Yes
T2/T3 Tracking								
All	-0.091**	-0.077**	-0.0087	-0.00025	-0.054**	-0.053***	-0.0010	-0.0062
	(0.040)	(0.036)	(0.028)	(0.026)	(0.021)	(0.019)	(0.010)	(0.0098)
T1b	-0.160	-0.101	0.021	0.031	0.035	0.054	-0.019	-0.039
	(0.241)	(0.203)	(0.097)	(0.092)	(0.117)	(0.126)	(0.045)	(0.044)
T1/T2	-0.208	-0.181	-0.043	-0.020	-0.104*	-0.095*	-0.0044	-0.013
	(0.158)	(0.136)	(0.067)	(0.060)	(0.060)	(0.055)	-0.022	(0.021)
T2	-0.106*	-0.095*	-0.0054	-0.00029	-0.085***	-0.083***	0.026*	0.017
	(0.059)	(0.055)	(0.034)	(0.033)	(0.029)	(0.028)	(0.014)	(0.013)
T2/T3	-0.101	-0.110*	-0.022	-0.030	-0.041	-0.048	0.0037	0.0035
	(0.067)	(0.061)	(0.046)	(0.042)	(0.031)	(0.030)	(0.017)	(0.018)
T3	0.310	0.298	0.104	0.111	0.094	0.090	-0.067*	-0.066*
	(0.187)	(0.178)	(0.099)	(0.100)	(0.081)	(0.074)	(0.041)	(0.040)
Controls		Yes		Yes		Yes		Yes

Notes: Table shows the IV estimates of early tracking on several measures of labor market success, either with or without controls. Estimation is conducted for tracking between the low and the medium track (T1/T2), and medium and high track (T2/T3). Effects are estimated on the 1989 cohort only. See table 2 for an explanation of outcome variables. See table 3 for a description on the indicators in the left column, the method of instrumentation and the controls included. The first stage is too weak for estimation of all but the T2 subgroup for T1/T2 tracking and for the T1b and T3 subgroups for T2/T3 tracking (see table 3). These estimates are reported for completeness. For welfare, this only applies to subgroups T1a/T1b and T1b for T1/T2 tracking.

Table 7: Robustness to alternative specifications

	Original Estimates	Residence Municipality	Add municipal ability	Comparison Supply Ratio	Add change Supply Ratio
T1/T2 Tracking					
Higher Education	-0.082**	-0.086***	-0.085***	-0.039	-0.040
	(0.025)	(0.029)	(0.024)	(0.092)	(0.087)
University	-0.018	0.0065	-0.020	-0.143	-0.140
	(0.033)	(0.0036)	(0.032)	(0.149)	(0.147)
Dropout	0.017	0.014	0.018	0.019	0.019
	(0.013)	(0.015)	(0.013)	(0.043)	(0.043)
Earnings	-0.077**	-0.101*	-0.077**	-0.417	-0.463
	(0.036)	(0.053)	(0.034)	(0.303)	(0.369)
Corrected	-0.00025	-0.017	-0.00054	-0.109	-0.131
Earnings	(0.026)	(0.033)	(0.026)	(0.100)	(0.116)
Employment	-0.053***	-0.063**	-0.053***	-0.169	-0.163
	(0.019)	(0.026)	(0.019)	(0.121)	(0.128)
Welfare	-0.0062	0.0067	-0.0050	-0.029	-0.029
	(0.0098)	(0.011)	(0.0096)	(0.052)	(0.052)
T2/T3 Tracking					
Higher Education	-0.172***	-0.276***	-0.178***	-0.095	-0.068
	(0.050)	(0.079)	(0.050)	(0.085)	(0.090)
University	-0.012	-0.027	-0.011	-0.042	-0.040
	(0.021)	(0.031)	(0.020)	(0.032)	(0.033)
Dropout	0.032	0.034	0.029	0.055	0.046
	(0.023)	(0.034)	(0.022)	(0.051)	(0.057)
Earnings	0.026	0.170	0.034	-0.240	-0.131
	(0.090)	(0.230)	(0.091)	(0.347)	(0.205)
Corrected	0.081	0.074	0.081	0.093	0.056
Earnings	(0.060)	(0.126)	(0.059)	(0.204)	(0.156)
Employment	0.0068	0.120	-0.0045	0.0029	0.039
	(0.059)	(0.162)	(0.059)	(0.190)	(0.157)
Welfare	-0.036	-0.010	-0.036	0.00058	-0.016
	(0.025)	(0.036)	(0.024)	(0.042)	(0.043)

Notes: Table shows the robustness of the coefficients from tables 5 and 6 to various alternative specifications. The columns report, in order: the original estimates; estimates when using the relative supply of Early Tracking schools by municipality of residence (in stead of municipality where the school resides); estimates when adding mean Entrance Test Score and mean Parental Education by municipality; estimates with the original model on a sample of municipalities that are prevalent in both cohorts; estimates on the same sample, adding the change in the supply-ratio from 1989 to 1993 cohort. All reported coefficients are based on specifications using all controls. See table 3 for a description on the original method of instrumentation and the controls included. See table 2 for an explanation of outcome variables.