CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 300

Does the teacher beat the test?

The additional value of teacher assessment in predicting student ability

Eva Feron
Trudie Schils
Bas ter Weel

Does the teacher beat the test? The additional value of teacher assessment in predicting student ability*

Eva Feron
Department of Economics, Maastricht University
e.feron@maastrichtuniversity.nl

Trudie Schils
Department of Economics, Maastricht University
t.schils@maastrichtuniversity.nl

Bas ter Weel
CPB Netherlands Bureau for Economic Policy Analysis &
Department of Economics, Maastricht University
b.ter.weel@cpb.nl

January 2015

Abstract

This research investigates to what extent subjective teacher assessment of children's ability adds to the use of test scores in the explanation of children's outcomes in the transition from elementary to secondary school in terms of initial track allocation, track switching in the first three years of secondary education and subsequent test scores. We apply micro-data from the Netherlands about cognitive test scores and teacher assessment in elementary schools and about track placement, track switching and test scores in secondary schools. Our estimates suggest that subjective teacher assessment is about twice as important as the elementary school cognitive test scores for initial track placement in secondary school. In addition, teacher assessment is more predictive of track allocation in 9th grade compared to cognitive test scores. Next, children who switch tracks are more likely to be placed in tracks based on test scores. Also, test scores in 9th grade are predicted by subjective teacher assessment, not by test scores in 6th grade. Finally, a back-of-the-envelope calculation shows that switching could be reduced by at least ten percent if children would have been allocated according to the teacher's assessment.

* The programs and data used in this paper are available from the authors upon request. We would like to thank Lex Borghans, Flavio Cunha, Bart Golsteyn, Erzo Luttmer, Marc van der Steeg, Dinand Webbink and Karen van der Wiel for comments on an earlier draft. We are also grateful for useful feedback from seminar participants at the CPB, Maastricht University and the University of Philadelphia , conference participants at de *Nederlandse Economen Dag* in Amsterdam, de *Onderwijsresearch Dagen* in Brussels, the ESPE conference in Braga, the IWAEE conference in Catanzaro and the EALE in Ljubljana and teachers in Communities of Learning where we discussed this topic and received feedback from the perspective of the field.

1. Introduction

The weight of cognitive test scores in allocating children to different levels of education has increased over time (e.g., Kautz et al., 2014). The reason why cognitive test scores are widely used as assessment measures to sort children is that they are assumed to give objective measures of ability levels. Assessments made by teachers represent more subjective measures of ability and have been criticized for being biased; for example towards gender and against minorities (e.g., Dee, 2005; Burgess and Greaves, 2013; Fairlie et al., 2014). However, teacher assessments could also be valuable complements to cognitive test scores. Teachers work with children on a daily basis, which allows them to also assess other determinants of ability such as motivation and classroom behavior (e.g., Segal, 2008).

This research investigates to what extent teacher assessment contains additional information that is useful in determining elementary school children's ability level on top of the information provided by cognitive test scores. We make use of a database that contains information about cognitive test scores and teacher assessment in elementary school and about initial track placement and subsequent careers in secondary school of Dutch pupils. The Netherlands has an educational system that involves early tracking (i.e. at the age of 12). Our empirical analysis benefits from this system because we observe both high-stakes cognitive test scores and teacher assessments at the end of 6th grade and the transition from primary to secondary school during which children are allocated to different (hierarchical) education tracks.

The research strategy involves three steps. The first step is to document whether or not there are non-random differences between cognitive test scores and teacher assessment at the individual level. In the second step we investigate whether or not cognitive test scores or teacher assessment is more predictive of track placement in 7th grade and track allocation in 9th grade. In a complementary analysis we document the determinants of math and language test scores in 9th grade by correlating them with both ability signals. In the third step we pay attention to track switchers (between 7th and 9th grade) to investigate if these children have been allocated according to teacher assessment or cognitive test scores.[1] The application of the assessment measures by secondary schools to allocate children to different tracks (or the lack thereof) yields information about the usefulness of objective and subjective assessment measures in allocating children.

[1] The benefits of an early tracking system are highest when children are allocated to tracks most suitable given their ability level and that they stay within this track throughout their secondary school careers. These seem to be the most important reasons for early tracking in for example Germany (e.g., Woessmann, 2004) and the Netherlands (e.g., Diris, 2012), two countries that track children the earliest among all OECD countries.

Our database for the empirical analysis includes 4,500 children. The database is constructed by making use of administrative data from school tracking systems and survey data of children in 6th and 9th grade in the period 2009-2012 (i.e., these children were in 6th grade in 2009 and were followed until 9th grade in 2012). We have access to a collection of data consisting of both objective and subjective assessment measures in 6th grade, track placement from 7th to 9th grade, test scores on an identical test across all different tracks in 9th grade and information about switching tracks in secondary school. The survey data include measures about demographic factors as well as socio-economic status. The objective assessment measure is a high-stakes standardized test score (so-called *Cito Eindtoets*) which children take in 6th grade. This objective test score provides a well-defined measure of levels of achievement and is annually taken by almost all 6th grade children in the Netherlands.[2] The subjective assessment measure is the teacher's assessment of the child's ability level. This subjective assessment is made by teachers in 6th grade after they have observed the test score from the objective assessment.

Our four most important findings can be summarized as follows. First, for 19 percent of our sample we observe a substantial difference between the objective and subjective assessment measure of ability in 6th grade. In three quarters of these cases the subjective teacher assessment measure is higher than the test score. We find that there are systematic differences between the objective and subjective assessment measure. Our most important findings relate to gender and social-economic status. Girls are more likely to receive a teacher assessment that is higher than their test score compared to boys and children from families with lower socio-economic status are less likely to receive a teacher assessment that is higher than the test score compared to those children from higher level families.[3]

Second, we estimate several models to analyze the determinants of track placement in 7th grade and track allocation in 9th grade. Our estimates suggest that subjective teacher assessment is able to explain about two times more of the gap between the lowest and the highest track placement compared to the objective test score in both 7th and 9th grade. These results suggest that secondary schools put more emphasis on the subjective assessment measure relative to the objective assessment measure when allocating children to tracks. It also suggests that teacher assessment is

---

[2] Approximately 85 percent of all children who are in 6th grade participate in the *Cito Eindtoets* every year.

[3] These numbers are consistent with previous studies using data about the Dutch education system. For example, Timmermans et al. (2012) investigate the relationship between assessment measures and secondary school careers. They use a different dataset, which contains information about children from all regions in the Netherlands but no information on track placement and switching. Their descriptive statistics about assessment measures are similar to ours, which suggests that our analysis is likely to be representative for the Netherlands. Golsteyn and Schils (2014) show that boys and girls score differently on the 6th grade tests according to differences in personality that might also be related to higher teacher assessments.

not only predictive of initial track placement but also of the longer term career in secondary school compared to the cognitive test score. In addition, secondary schools seem to allocate children in accordance with the highest assessment signal of ability. Finally, our estimates are conservative or lower-bound estimates as we only consider eight different educational tracks in secondary education for these analyses and are very strict in labeling signals as different. [4]

Third, we observe that the teacher assessment has non-random deviations from the test score. The question is whether these non-random deviations are efficient with respect to later outcomes. Therefore, we look at the determinants of switching between different tracks. We distinguish 11 types of secondary school tracks between which children are able to switch. It turns out that approximately 24 percent of our sample makes such a major switch between tracks between 7th and 9th grade. Our analysis suggests that those children who are allocated to tracks that are not in accordance with either the objective or the subjective assessment measure have the highest probability of switching tracks. Children who are allocated according to the teacher's assessment are the least likely to switch tracks. Furthermore, we find that switching between tracks is negatively related to test scores on a math/language test in 9th grade.

Finally, we investigate to what extent test scores in 9th grade correlate with the ability signals documented in 6th grade and to what extent switching correlates with lower 9th grade test scores. Our database contains information about an identical cognitive test (including math and language) that all children had to make in 9th grade. It is a low-stakes test. The estimates suggest that the teacher's assessment positively correlates with the 9th grade test scores, whereas the cognitive test score in 6th grade does not explain test scores in 9th grade (when controlling for the teacher's assessment). Switching tracks between 7th and 9th grade seems to have a negative effect on this test score in 9th grade, pointing towards costs of switching.

This paper contributes to the literature about the consequences of using objective and subjective assessment measures for tracking and successive performance. Dee (2005), Lindahl (2007), Lavy (2008), Gibbons and Chevalier (2008), Cornwell et al (2012) and Burgess and Greaves (2013) all use objective and subjective assessment measures to study discrimination and uncertainty. It is shown that systematic differences exist between these two types of instruments in the assessment of children's performance, such as between boys and girls, or between blacks and whites. Bernardi et al. (2014) show that additional information from cognitive and non-cognitive tests is able to help

[4] We group the tracks in pre-vocational education which leads to eight educational tracks as the number of children allocated to each of these tracks is relatively low. This decreases the variation in the data compared to analyzing 11 educational tracks.

children make a more efficient track allocation choice. Our contribution to this literature is that we analyze differences in both assessment measures for track allocation and switching, where we are able to observe children's later outcomes in 9th grade in the form of track allocation, track switches and their scores on a math and language test.

Other studies have shown that test scores in secondary school are predictive of labor-market outcomes (e.g., Murnane et al., 1995 and Currie and Thomas, 1999). We obtain a set of estimates suggesting that switching seems to lead to lower test scores in 9th grade. This seems to support arguments that switching tracks harms children's accumulation of human capital, which is documented in Van Elk et al. (2009) and Diris (2012) for the Netherlands.

Our work also contributes to the literature on early school tracking. The long-run effects of early tracking for human capital development and educational opportunities have been summarized by Hanushek and Woessmann (2006) and Brunello and Checchi (2007). According to the OECD, the early tracking regime in the Netherlands causes a severe constraint for the growth of higher education participation. It states that "In the end, postponement of the present early tracking regime seems inevitable; although this is a major change in the way Dutch society thinks of itself" (OECD, 2007, p.38). Consistent with this advice, other studies using Dutch data suggest that relatively low-ability children could improve their educational outcomes by about 30 percentage points when tracking is postponed by one year. At the same time, most children do not seem to be hurt by the presence of low-ability children in the first year of secondary school. Only children who are considered to have the highest ability seem to be hurt by the presence of lower ability peers (e.g., Van Elk et al., 2009 and Diris, 2012). We show that a substantial fraction of our population is not allocated to the right track, which seems akin to restraints on optimal human capital development.

We proceed as follows. First, we present background features of the Dutch education system and explore the research strategy. Section 3 documents the data description and statistics of our core variables. Sections 4-6 present the results on the differences between objective and subjective assessment measures, track switching and 9th grade test scores. Section 7 briefly addresses the policy perspective of our analysis with a focus on reducing switching. Section 8 concludes. We present additional results and detailed data descriptions in the Appendix to this paper.

2. Background and strategy

We observe five main outcomes for each child: the test score which serves as an objective assessment of ability at the end of elementary school (6th grade), the elementary school teacher's assessment which serves as a subjective assessment measure of ability (6th grade), track allocation in the first and third year of secondary education (7th and 9th grade), the results from a cognitive test in 9th grade, and track switching in the first three years of secondary education (7th-9th grade). We now present information on these measures and information about the Dutch education system.

2.1. Dutch education system

Countries differ in the age at which they first track children into different types of schools. In the majority of OECD countries, tracking takes place between the ages of 14 to 16. Some countries, including the Netherlands, undertake the first tracking at the age of 12 when children progress from elementary to secondary school (i.e., from 6th to 7th grade).[5] We take advantage of this system by studying the allocation in secondary school, the transition from elementary to secondary school and performance in 9th grade.

In the Netherlands primary education consists of eight years of which the first two are spent in kindergarten. As of the third year of elementary school (i.e., 1st grade), children formally learn how to read and write. Most children start kindergarten at the age of 4, enter 1st grade at the age of 6 and finish elementary school at the age of 12. As of secondary school children are allocated to tracks. The track allocation decision is made by the secondary school. It is based on test scores and the elementary school teacher assessment. Some schools set a threshold test score level below which children are not allowed to enter a certain level of secondary education.

The Dutch secondary education system is hierarchically structured by ability and consists of three main tracks that differ in duration and qualification (see the left-hand column of Figure 1). The four-year track (VMBO or T1) qualifies children for vocational education, the five-year track (HAVO or T2) qualifies children for higher vocational education and the six-year track (VWO or T3) qualifies children for university education. The next column in Figure 1 shows four sub-tracks at the lowest level of secondary education (T1a to T1d). The difference between these four sub-tracks is the importance of a practical versus theoretical focus in the curriculum. Time spent on more

[5] Research has been conducted to investigate the effects of early tracking on differences in human capital formation and educational opportunities (e.g., Hanushek and Woessmann, 2006 and Brunello and Checchi, 2007 for comprehensive review studies). The overall conclusion of these investigations has been that the earlier the tracking, the greater the impact of socio-economic factors, such as parental education, on educational outcomes and opportunities. For the Netherlands, Van Elk et al. (2009) and Diris (2012) find that early tracking reduces participation in and completion of education for relatively low-ability children placed in lower tracks at the age of 12, while there does not seem to be a positive effect on those placed in the higher tracks. One of the main reasons of implementing an educational system with early tracking is to provide children with a learning environment in which they benefit from classrooms with homogenous populations. Such populations would benefit from peer effects. In addition, it allows for a focused curriculum and appropriately paced instructions.

theoretically oriented courses increases with the tracks from T1a to T1d.[6]

The third column of Figure 1 shows all possible tracks, some of which are combinations of the three major tracks. Both the objective and subjective assessment measure are tailored toward allocating children into one of these 11 track combinations. The 6th grade test distinguishes brackets which are consistent with these 11 track combinations, shown in the fourth column of Figure 1. Teacher assessment is measured on the same 11-point scale.

2.2. Background

Approximately 85 percent of all Dutch children complete the high-stakes objective assessment test in 6th grade.[7] The children in our data have completed this *Cito Eindtoets* in 2009, which is the most common objective assessment test in elementary school in the Netherlands. The test was taken during three days at the beginning of February 2009. The test is standardized meaning that the test procedure is the same for the whole country. During the assessment children have to answer questions in the areas of math, reading, study skills and science. The degree of difficulty of the test is approximately the same between different years. The test is calibrated every year to ensure that children's average scores on the test are comparable across different years. The performance is measured on a scale between 501 and 550.

The aim of the cognitive test score is to provide an independent and appropriate perspective on children's expected performance and their best track placement in secondary education. That is why the percentage of children with similar scores is reported for every different track in secondary education. The test is not meant to provide a signal of grades or performance on an intelligence test, which can lead to negative connotations. The test institute offers guidelines for children's track allocation by reporting brackets of scores and accompanying track assessments. We followed these guidelines for constructing our objective ability measure variable. We use the brackets as the outcome measure of the objective assessment measure.

High scores on the standardized test are an important way in which elementary schools try to signal the quality and value-added of their educational efforts. Elementary schools seem to use their average scores on this test to attract new children. In addition, the Dutch Education Inspectorate

[6] The six-year track (T3) can also be divided into two sub-tracks of which one includes the option to take courses in Greek and Latin. Our data do not allow us to distinguish between these two sub-tracks, which means that we combine them in T3. In practice, test scores and teacher assessment do not distinguish the two tracks.

[7] All children have to take an objective assessment test in 6th grade. Schools are free to choose which test their children take. In our database all children have taken the *Cito Eindtoets* in 2009.

uses these results, controlled for individual characteristics, as one of the inputs for their overall evaluation of the school's quality and value-added. Children also have an incentive to obtain a high test score because it is an important signal of their ability. In that sense, the test is a high-stakes assessment.

In addition to the objective test score, teachers make a personal assessment of each child's level of ability. The assessment is based on the teacher's experience and interaction with the child, observable demographic and socio-economic factors and the child's performance throughout all grades in elementary school. Teachers also know the test score from the objective assessment measure when they present their assessment of the child's ability. This subjective assessment is provided in the spring of 2009 before children apply to a secondary school. The teacher's assessment is provided in similar brackets as the objective assessment and fits with possible track allocations in secondary schools.

Elementary school teachers do not have a strong incentive for strategic behavior in such a way that their assessment overstates the child's ability. The teacher's compensation scheme does not depend on the assessments made. Furthermore, the elementary school's population usually goes to the same secondary schools every year. This means that over time the information asymmetry between the primary and the secondary schools reduces and secondary schools learn how to interpret the assessments from elementary schools. Furthermore, each year children are assigned to a class in which they are taught by one elementary school teacher. This teacher is involved with teaching all subjects in elementary school. Differences in test scores across different parts of the test are therefore unlikely to be driven by different teacher characteristics.

Secondary schools allocate children to tracks. They obtain the information about the objective and subjective assessment measures. Secondary schools have an incentive to allocate children to the track that matches their ability level. Inputs for the Education Inspectorate evaluation of secondary schools' performance include the percentage of children who graduate every year as well as the percentage of children who switch tracks. Allocating children to tracks that are too high (too low) leads to switching downward (upward) and would induce negative (positive) evaluations on this part of the performance assessment. Nevertheless, secondary schools also benefit from having more children in the higher tracks as this is beneficial for signaling the quality of the schools' education, which potentially helps to attract more children.

2.3. Strategy

Our analysis first focuses on the way in which both the objective and subjective assessment measures in 6th grade help to explain track placement in 7th grade. Both assessment measures aim to measure ability, but face the problem that the true underlying and unobserved level of ability is unknown. We define the objective assessment measure of child i as his test score (TS_i). This test score depends on i's true and unobserved ability (A_i), a vector of observed characteristics (X_i) and the elementary school he attends (P_i). These ingredients lead to the following measurement of child i's test score:

$$TS_i = f(A_i, X_i, \beta_{P_i}). \tag{1}$$

In an ideal world $TS_i = A_i$. In practice this is not true because TS_i is measured with noise and observed characteristics X_i are likely to influence the measurement of A_i by TS_i. The reason for adding school fixed effects (or dummies in our cross-sectional specifications) is that characteristics of elementary schools can be related to test results of children in 6th grade, which can influence TS_i. Some of these school characteristics we cannot observe. Hence, we add school fixed effects to the model.

We define the subjective assessment measure of child i as the teacher's assessment of his ability (TA_i). This measure includes the same ingredients and the observed test score:

$$TA_i = g(A_i, X_i, TS_i, \beta_{P_i}). \tag{2}$$

The information about the child's test score influences the teacher's assessment of the child's ability. We abstract from differences in school fixed effects that may select teachers with certain attitudes because we do not observe individual teachers in the data. Because children are assigned to one teacher in the final year of primary education and most elementary schools have only one 6th grade class, potential teacher effects are captured by school fixed effects. This is also the reason for indexing all variables with child i and for ignoring teacher j effects.

Finally, we observe child i's initial track placement (TP_i) in 7th grade (at secondary school) and thereafter allocation in 9th grade. The decision about initial placement is made by the secondary school and includes the objective and subjective assessment measures. Adding secondary school fixed effects to the model would create additional endogeneity as not all schools offer the same track levels and secondary school fixed effects are related to TS_i and TA_i.

$$TP_i = h(A_i, X_i, TS_i, TA_i, \beta_{P_i}). \tag{3}$$

In the first part of the empirical analysis we analyze if there are any systematic differences between the test score, the teacher assessment and track placement for various socio-economic background characteristics of children. Furthermore, we are primarily interested in explaining the track placement of child i. We estimate equation (3) with an ordered probit model to find the determinants of track placement in both 7th and 9th grade. Second, we do not observe A_i but two signals TS_i and TA_i. Third, in the empirical analysis we incorporate the possibility that secondary schools take into account both TS_i and TA_i, although TA_i includes information about TS_i. In this way we have the two signals competing with each other. A statistically significant coefficient of TS_i would in all likelihood suggest that secondary schools put weight on both signals of ability. The equation we first estimate is:

$$TP_i = C + a_1 X_i + a_2 TS_i + a_3 TA_i + \beta_{P_i} + \varepsilon_i, \quad (4)$$

where ε_i is the error term.

In the second part of the empirical analysis we estimate the determinants of track switching in the first three years of secondary school (i.e., in the period spanning 7th, 8th and 9th grade). To do so, we first estimate a set of probit models in which we show what type of children tend to switch tracks. Second, we estimate probit models in which we estimate the probability of switching tracks (SW_i) for child i:

$$SW_i = C + a_1 X_i + a_2 TS_i + a_3 TA_i + a_4 TP_i + \gamma_{S_i} + \varepsilon_i, \quad (5)$$

where ε_i is the error term. We estimate different versions of the model in which the dependent variable is switch, switch up or switch down. We only include secondary school fixed effects because switching takes place in secondary school. We show below that elementary school fixed effects have no impact on track placement, which makes us confident that including only secondary school fixed effects is sufficient to estimate the determinants of switching.

Finally, we estimate models to investigate to what extent test scores on an identical (low-stakes) test in 9th grade are correlated with the ability signals from the teacher and the test score in 6th grade. We also estimate to what extent switching is correlated with test scores in 9th grade.

The strength of the data at our disposal is that we are able to observe performance in both primary and secondary school. In addition, we have detailed information about elementary school teacher assessments and initial track placement in secondary school. This is a unique feature in the literature. Nevertheless, the analysis is constrained by the fact that we are not able to identify a source of exogenous variation in our data. Ideally, one would want to conduct an experiment in which a random portion of the sample was placed according to the test scores' signal, another part according to the teacher's signal and a final slice of the population as it is currently done (i.e., decided by the secondary school). The alternative is to find instruments to deal with the "self-fulfilling prophecy" that creates endogeneity. The "self-fulfilling prophecy" is the idea that when a child is placed on a higher or lower track than he should be according to his true ability, the child is more likely to switch back to the track that matches his true ability. This is to be kept in mind when interpreting the results in the switching section. Furthermore, all tests, but also the teacher assessment, contain measurement error and analyses concerning ability suffer from omitted variable bias. However, such instruments are not readily available. Our analyses focus on outcomes between 6th and 9th grade. In order to find exogenous variation we would need a set of instruments related to (one of) our assessment measures and at the same time unrelated to any unobservable variables that might influence our outcome variable. Since children's true ability is unobserved, this is problematic. We are aware of the endogeneity concerns with respect to omitted variable bias and also with the fact that potential measurement error is an important disclaimer when interpreting the estimated coefficients, but try to deal with this in the best way possible by using primary and secondary school fixed effects and a rich set of covariates, including track placement in 7th grade.

3. Data

Before we present our results, we first document the most salient features of our data to reveal information about the allocation of children and to present a number of key descriptive statistics. More information as well as additional regression analyses are presented in the four sections of the Appendix at the end of the paper.

3.1. Descriptive statistics

For the analyses we use a unique dataset on the educational development of children in a given region (Limburg) of the Netherlands. These data are collected in a cooperative project between (elementary and secondary) schools, school boards, municipalities and Maastricht University to analyze school performance in order to foster educational improvement. The unique feature of this program is the participation of almost all schools in the region in this project, implying almost full coverage of children (about 95 percent of the elementary schools participate in the program and about 90 percent of the secondary schools in the region). In 2009, information about all 6th grade children was collected and these children were again reviewed when they were in 9th grade in 2012. In both years the data collection includes administrative data from the school information systems, surveys among children and their parents and test results. The data covers children from all tracks with the exception of those who are in special needs education.

Table 1 documents the distribution of test scores TS_i in 7th and 9th grade, teacher advice TA_i, and (initial) track placement TP_i, across the 11 different tracks (see Figure 1). The numbers in each row add up to 100 percent. In addition, the table documents the number of switchers from the initial track to which they are allocated in 7th grade. The numbers represent the fraction of children who switch away from each of these tracks. The distribution of test scores and teacher advice seems to be broadly consistent but there are also important differences. Track placement and teacher assessment at higher levels of secondary education is different from the test scores, with almost a quarter of the sample being placed in T3 in 7th grade. Teacher assessment seems to be more favorable for the higher tracks compared to the test score. We explore these differences further in Section 4.

Furthermore, teachers seem to be reluctant to advise tracks in the middle tracks of vocational education (i.e., tracks in T1). As can be observed from Table 1, some T1 sub-tracks contain only few observations. These are combination tracks to which only a few children are allocated.[8] We join the combination tracks and the regular tracks for T1 when analyzing differences between TS_i and TA_i. This means that for the analysis of track allocation we combine tracks T1a and T1a/T1b, T1b and T1b/T1c and T1c and T1c/T1d into three categories. This results in a more conservative estimate of the determinants of allocation in Section 4. As the difference between TS_i and TA_i does not have a direct impact on the number of children who switch tracks, we use 11 tracks for the analyses of switching tracks. We define differences in TS_i , TA_i and TP_i when there is a difference of at least two tracks. Furthermore, we define a switch between tracks when children switch at least two tracks. For example, if children switch from track T2 to track T1d/T2 or track T2/T3 this is not defined as a switch. When children switch from track T2 to track T1d or track T3 this is defined as a switch.

Table 2 reports descriptive statistics. Our data include slightly more girls than boys. In 6th grade children are on average 12 years old. Most of the children (and their parents) were born in the region (Limburg) but there are also children (and their parents) who were born in another part of the Netherlands or abroad. The education level of the father and mother is measured as the highest completed level of education. The average education level of parents suggests that they have completed vocational education. Parental education levels are measured in four different categories: (1) lower education, (2) vocational education, (3) higher education and (4) university education. The labor-market position of the father and mother is also measured in four categories: (1) employed, (2) unemployed, (3) sick/unable to work and (4) not in the labor force. Almost all fathers are employed, while more mothers report to be unemployed. The number of workdays per

[8] It is possible that this is related to the availability of this type of school track in the area.

week of the father and mother is a measure of the amount of time allocated to paid work. Most fathers work fulltime, while more mothers report to work part-time. In The Netherlands, part time employment is an important form of employment for women with young children (e.g., Bosch et al., 2010). Working time is defined in three categories: (1) part-time 1-2 days per week, (2) part-time 3-4 days per week and (3) fulltime. Almost 80 percent of the children in our sample live with both parents in 9th grade.[9] Here teacher assessment, track placement and test score are all measured on a scale from 1-8. The original test score in 6th grade is measured on a scale from 501-550. Based on test score ranges provided by the institution that supplies the test in 6th grade, we rescaled the test score to a scale from 1-8. Average test score and average test score (short) correspond to a T1d/T2 track. Average teacher assessment and average track placement correspond to a T2 track (see Figure 1). Almost 24 percent of children switch tracks between 7th and 9th grade.

Figure 2 shows the distributions for the test score (Panel A), teacher assessment (B), and track placement (C) in the eight brackets we use for the first section of the empirical analyses. These eight categories are underlying our definitions of TS_i , TA_i and TP_i. The vertical axis in each of the three panels shows the percentage of children in each category. These percentages add up to 100. As can also be observed from the numbers displayed in Tables 1 and 2, the differences in means and other moments of the distribution between the three measures are relatively small.

Table 3 shows the bivariate correlation coefficients and their statistical significance of the main variables in our empirical analyses.[10] We observe that the correlations among teacher assessment, test scores and track placement are high and positive, which comes without surprise given the patterns displayed in Figure 2. What is also interesting to observe is that track placement shows a higher positive correlation with teacher assessment than with test scores. Furthermore, we observe that there is a negative correlation between switching tracks and the test score, teacher assessment and track placement. The majority of the switches happen in the pre-vocational track as can be observed from Table 1.

Next, Figure 3 shows that approximately 24 percent of our sample of children switches tracks between 7th and 9th grade. Most of the switches (55 percent) happen from 8th to 9th grade. Approximately 71 percent of all children who switch between tracks switch down and only about 29 percent of all children who switch between tracks switch up. In Figure 3 we observe that most children who switch tracks switch two tracks up or down measured on the track scale from 1-11 as

[9] The corresponding survey question is: “Do you live at home with both parents?”

[10] Appendix A presents a correlation table of all variables of interest and a detailed description of the covariates.

described in Figure 1.[11] Furthermore, correlations between teacher assessment, test score, track placement and switching are negative. This is due to the nature of our data. Since we observe track switches until 9th grade, we capture almost all of the switches in the T1 track but we capture less of the switches in the T2 and T3 tracks as these children still have the opportunity to switch tracks after 9th grade. This finding is confirmed by results from the Inspectorate of Education for all children in The Netherlands (Education Inspectorate, 2007).

Finally, we use information about a math and language test that children in our sample have taken in 9th grade. This test was a low-stakes test and part of the research project conducted at the schools in our sample. It main purpose was to have a school-independent test score for students in 9th grade. The difficulty level of the test differs according to the students' track level. Since we want to compare the effect of switching tracks across tracks we only use the questions on math or language that are identical for all students. This leaves us with 11 questions on math and 8 questions on language. Since we do not have enough observations for all tracks, we only look at students in track T1c, T1d, T2 and T3 in relation to their score on the math or language test. The average percentage of questions answered correctly increases from 34 in track T1c to 49 in track T3. The standard deviation is around 20 percent in all tracks. The distribution of the test score in 9th grade is presented in the last row of Table 1. The row adds up to 100 percent and shows the percentage of students per initial track placement in 7th grade.

3.2. Possible selection

Data has been collected from 155 elementary schools (95 percent of all schools in the given region) and 30 secondary schools (90 percent of all schools in the given region). This results in a database of $n = 4{,}500$ for the first section, $n = 4{,}019$ for the second section and $n = 1{,}812$ for the third section of the empirical analyses of this paper. In the first section of the empirical analyses we discuss non-random differences between the test score, the teacher assessment and track placement in 7th and 9th grade. In the second section of the empirical analyses we discuss track switching because mistakes in initial track placement or suboptimal allocation can be made undone in the first part of secondary education. In the third section we correlate both ability signals and switching with 9th grade test scores. Data has been collected for 9,092 children. For the analyses in the first section of the paper we only use those children for whom we observed the teacher assessment and the test score in our data. For the analyses in the second section we also need to know children's track

[11] Our data includes 1,266 switchers. Of these switchers 43 children switch down/up multiple tracks and 65 children switch both up and down between 7th and 9th grade. We analyze overall switches which means that we refer to 1,158 switchers (i.e., $1{,}266 - 43 - 65 = 1{,}158$). Figure 3 includes all 1,266 switchers.

placement in 7th, 8th and 9th grade, which reduces the sample size to 4,019. Finally, 9th grade test scores are available for 1,812 children.

We address possible selection issues with regard to the sample we use for our analyses. We investigate whether individual characteristics are able to predict whether children end up in our sample for analyses. We are mainly concerned about schools only reporting data for their well performing children and holding back information on other children. After controlling for school fixed effects we find that individual characteristics are not significantly related to selection into our sample. We conclude that selection is not an issue (see Appendix B for a more elaborate analysis).

4. Track placement

This section presents our first set of estimation results. We first investigate to what extent there are non-random differences between the objective and subjective assessment measures. Second, we present a set of estimates about the determinants of track placement in 7th and 9th grade.

4.1. Differences between objective and subjective assessment measures

To compare differences between our objective and subjective assessment measures we have created three categories: $TA_i < TS_i$, $TA_i = TS_i$ and $TA_i > TS_i$. Figure 4 shows that 81 percent of the children in our sample are faced with objective and subjective assessment scores that are equal. If there is no systematic difference between the teacher assessment and the test score, we would expect both assessments to be equal on average. Any deviations should be approximately symmetric. From Figure 4 we observe however that it is more likely that the subjective assessment measure is higher when there is a difference between the two measures. In 5.1 percent of the cases the subjective assessment is lower compared to the objective assessment, in 13.9 percent it is higher. Since the teacher assessment makes use of the information revealed by the test score, the teacher has access to a child's educational history in elementary school and has knowledge about a child's background characteristics and earlier test scores, she has an information advantage.[12]

When we look into the differences between the test score, teacher assessment and track placement in 7th grade, we find that most of the differences we observe are related to gender and social-economic status. Table C1 and C2 in the Appendix show descriptive statistics and estimated marginal effects from ordered probit models. From these analyses we observe the following main patterns. First, we observe that girls seem more likely to receive a teacher assessment that is higher compared

[12] If we use 11 different tracks instead of 8, 32.9 percent of all children is faced with different assessment measures. Of these children 83 percent receives a higher subjective assessment measure of ability.

to the test score, while they seem less likely to receive a teacher assessment that is lower compared to the test score. In addition, girls are more likely to receive a track placement equal to the subjective assessment measure and more likely to have a test score that is lower compared to their track placement. This suggests that girls not only get a more favorable assessment from the elementary school teacher, but also with regard to track placement in secondary school. Second, from the labor market position of the mother we observe that children who have mothers who are unemployed are less likely to receive a subjective assessment that is higher than the objective assessment. These children are also less likely to have a test score that is lower than their initial track placement and more likely to receive a track placement that is equal to their test score. Furthermore, children of mothers who are sick/unable to work are less likely to receive a subjective assessment that is higher than the objective assessment. These children are also less likely to receive a test score that is lower than their track placement and more likely to receive a track placement that is equal to or lower than their objective assessment. This indicates that secondary schools seem to allocate these students unfavorably. In our data over 70 percent of the mothers who are unemployed and over 85 percent of the mothers who are sick\unable to work completed only lower education or vocational education. Burgess and Greaves (2013) obtain similar results with respect to ethnic minorities, which they attribute to negative stereotyping of particular groups in society.

4.2. Determinants of track placement

Figure 5 shows how children are allocated to tracks according to the two assessment measures. The figure is divided into three panels. The first panel documents track placement of children who are faced with $TA_i < TS_i$. The second panel displays track placement of those with $TA_i = TS_i$ and the final panel shows placement of those with $TA_i > TS_i$. Track placement can be equal to either TA_i or TS_i in the upper and lower panel or $TA_i = TS_i$ in the middle panel of Figure 5. In addition, placement in the category labeled "else" represents those children who are placed in a track that was recommended by neither the subjective assessment nor the objective assessment measure. In almost all of these cases the subjective and objective assessment measures differ by more than three levels and track placement is in between these two measures. In some cases track placement is higher or lower than both assessment measures indicate. In the latter case our data suggest that it is more likely that track placement is higher than both assessment measures would merit.

The bars in Panel A of Figure 5 suggest that when $TA_i < TS_i$ children are more likely to be placed according to TA_i than TS_i (40.2 percent vs. 17.9 percent). At first sight, this suggests that secondary schools seem to act in a relatively conservative way by following the lower of the two signals. They seem to attach more value to the teacher's assessment of the child's ability relative to the test score.

At the same time, the share of children placed in tracks that do not directly correspond with one of the assessment measures is relatively large (41.9 percent). The numbers in Panel C suggest that secondary schools are more likely to follow the teacher's assessment even when it is higher than the test score. More than two thirds of the population with $TA_i > TS_i$ is placed according to TA_i. Also the share of children placed in tracks that do not directly correspond with the assessment measures is relatively low compared to the case in which $TA_i < TS_i$. Combined with the information from Panel A, a picture emerges that secondary schools attach more value to the teacher's assessment of children's ability relative to test scores. Note that the teacher's assessment of a child's ability is on average higher than the test score would suggest. Finally, the statistics in Panel B of Figure 5 reveal that when both assessment measures give the same signal about children's ability almost all children are placed in the corresponding track. Nevertheless, 7.3 percent of the children are allocated to a different track. Upon closer inspection most of these children are allocated to higher tracks relative to what the teacher's assessment and the test score recommend. Overall, it seems to be the case that secondary schools have a preference to allocate children according to the teacher's assessment measure and/or according to the assessment that signals the highest ability.

We continue by estimating the determinants of track placement in 7th grade. Table 4 presents the estimation results of equation (4). The estimates are coefficients from ordered probit models where track placement (measured between 1 and 8) is the dependent variable. In the first two columns we either use the test score or the teacher assessment to estimate the determinants of track placement. As expected, we observe a strong positive relation between both the test score and track placement and teacher assessment and track placement. When we rescale the test score and the teacher assessment by the cut points in their respective regressions we observe that a one standard deviation increase in the test score (teacher assessment) bridges 35.1 percent (35.2 percent) of the gap between the lowest and the highest track placement, without adding any other control variables. In column (3) we add both the test score and the teacher assessment to the model. The test score and teacher assessment do not seem to be orthogonal. When the test score and teacher assessment are both added to the model we observe that a one standard deviation increase in the test score (teacher assessment) bridges 11.7 percent (25.2 percent) of the gap between the lowest and the highest track placement, without adding any other control variables. The difference in coefficients between the test score and teacher assessment is significant. An important observation is that teacher assessment seems to be a better predictor of track placement than the test score. In columns (4)-(6) we add control variables, elementary school fixed effects and a measure of children's GPA on math and language tests in previous years, respectively. Our estimates remain approximately the same in these different specifications. Overall, both teacher assessment and the test score seem likely to be important determinants of track placement in 7th grade. From our final specification in column (6) we conclude that teacher assessment appears to play a more important role in determining track placement compared to the test score. The estimated coefficient is about twice as large. This finding seems to be consistent with the subjective assessment measure having more information about the child's ability than the objective assessment measure.

In columns (7)-(12) we investigate the determinants of track placement in 9th grade. Track placement in 9th grade consists of six categories as over time the combination tracks disappear and children get allocated to their final track.[13] Columns (7)-(9) show the results for the sample of children who have not switched tracks between 7th and 9th grade. We obtain estimated coefficients that suggest that teacher assessment is still the best predictor of track placement for the children who have not switched tracks.

Columns (10)-(12) show the results for the entire sample of children we observe in 9th grade. We obtain a set of coefficients that suggests that there is no statistically significant difference between the predictive power of the test score and teacher assessment. These results indicate that the teacher is the best predictor of children's' ability both in 7th grade and later on in the children's secondary school career. However, the results also show that the predictive power of the teacher compared to the predictive power of the test score falls over time. A possible explanation for this is that children who were initially assessed too favorable by the teacher have switched to a lower track in the first three years of secondary education.

To take into account the covariation between the teacher assessment and the test score we also estimate the predictive power of the teacher assessment on track placement in 7th and 9th grade after correcting for the predictive power of the test score. This seems a natural thing to do because the teacher knows the test score of the child when the assessment is made. We find that the teacher's assessment is still highly predictive of track placement in 7th and 9th grade after controlling for the covariation between the teacher's assessment and the test score. The estimated coefficients of these analyses can be found in Appendix E.

5. Switching tracks

Children could be allocated suboptimally across different levels of secondary education. Suboptimal

[13] Secondary schools use combination tracks in the first year(s) of secondary education due to the uncertainty around ability. By using combination tracks schools can observe performance and allocate children to the most appropriate track after learning more about the ability. The combination tracks disappear into the neighboring main tracks. For example, children placed in track T2/T3 in the first year are expected to be re-placed in track T2 or T3 in year 2 or 3. Such a re-placement is not considered as switching tracks.

allocation encourages switching, which comes with a cost of suboptimal human capital investments and adjustment cost. In addition, children could have to stay for an additional period in the secondary education system because transitions between tracks are not always smooth. Note that switching tracks is the most drastic measure that secondary schools can take when children are not performing up to expectations. For example, the first option for children who are not able to keep up with the required level is to let them repeat the same grade. In the event that the school believes that grade retention will be insufficient, children have to switch tracks. Furthermore, it is possible that the costs of switching tracks are different when children switch up tracks compared to when they switch down tracks. It is likely that the costs of switching tracks for children who switch down are more related to demotivation and that the cost of switching up tracks are more related to previous underinvestment of human capital.

5.1. Documenting switchers

In between 7th and 9th grade approximately 24 percent of our sample switches tracks.[14] Switches are defined for all 11 tracks and are counted based on major switches, i.e., at least two steps. Appendix D documents the characteristics of switchers in detail and presents the coefficients of probit models.

Figure 6 shows track switching by differences in objective and subjective assessment measures. The figure is divided into three panels. The first panel documents the track switching of children who are faced with $TA_i < TS_i$. The second panel displays track switching of those with $TA_i = TS_i$ and the final panel shows switching of those with $TA_i > TS_i$. Similar to Figure 5, track placement is equal to either TA_i, TS_i or "else". The category labeled "else" represents those children who are placed in different tracks than the measures advised. Note that the number of switchers is determined on the basis of all 11 possible tracks.

The bars in Panel A suggest that when $TA_i < TS_i$ fewer children switch between tracks when they have been placed according to TA_i and more children switch when they are placed according to TS_i or in another track than either measurement pointed at. Comparison of TA_i and TS_i in Panel A suggests that those who are placed according to TS_i are more likely to switch up consistent with the argumentation that the teacher assessment is generally more generous about the children's ability than the test score. The numbers in Panel C suggest that although secondary schools are more likely to follow the teacher's assessment, even when it is higher than the test score, the number of switchers is relatively low when children are allocated based on TA_i. In addition, if children are

[14] These numbers are supported by the Dutch Education Inspectorate who find similar numbers,

placed according to TS_i more children switch up. Finally, the statistics in Panel B of Figure 6 reveal that children switch less often if $TP_i = TA_i = TS_i$. The overall picture that emerges from Figure 6 is that children placed according to TA_i have a lower probability to switch tracks relative to children placed according to TS_i. This conclusion seems to hold regardless of the difference between TA_i and TS_i. In addition, children allocated in accordance with the highest (lowest) of the two assessment measures seem to have a higher probability to switch down (up) a track, which is consistent with overassessment (underassessment) of a child's ability.

5.2. Determinants of track switching

We continue by presenting the results of analyzing probit models in which we estimate the probability of switching. Columns (1)-(3) in Table 5 present marginal effects for overall switching, columns (4)-(6) present marginal effects for switching down and columns (7)-(9) present marginal effects for switching up. The third specification in all three models includes control variables and secondary school fixed effects. Furthermore, standard errors are clustered at the secondary school level.

The estimated coefficients in column (3) suggest that children who are placed in a track according to teacher assessment (and not according to the test score) are 10.4 percent more likely to switch tracks, children who are placed according to the test score (and not according to the teacher assessment) are 15.1 percent more likely to switch tracks and children who are placed not according to the teacher assessment or the test score are 17.5 percent more likely to switch tracks compared to the base level.[15] The coefficients displayed in column (6) suggest that children placed in a track according to teacher assessment (and not according to their test score) are more likely to switch down, whereas children who are placed according to their test score (and not according to their teacher's assessment) or according to neither of the assessment measures are not more likely to switch down. Finally, from the estimated coefficients shown in column (9) we observe that children who are placed in a track according to their test score (and not according to their teacher assessment) and children who are placed according to neither of the assessment measures are more likely to switch up, whereas children placed according to their teacher assessment (and not according to their test score) are not more likely to switch down. Finally, children for whom $TP_i = TA_i = TS_i$ are least likely to switch tracks. For these children there is only little uncertainty about the most efficient track placement. These results seem to suggest that children are less likely to switch when they are allocated to tracks based on teacher assessment. However, teacher

[15] The base level in all these analyses is when $TP_i = TA_i = TS_i$ or $TP_i = TA_i$ & $TP_i = TS_i$ but $TA_i \neq TS_i$. For example: $TA = T3$, $TS = T2$ and $PL = T2/T3$.

assessment is generally more favorable than the test score about a child's ability. Therefore, children who are allocated according to teacher assessment are more likely to switch down and children allocated to the test score are more likely to switch up. [16]

6. Test scores in 9th grade

For a subsample of children we have data about a math and language test score in 9th grade. Children were randomly assigned questions in either math, language or both. We use the answers to 11 math questions or 8 language questions that have been asked to children in all tracks. This results in test scores of $n = 1{,}812$ children.

Table 6 documents the estimated coefficients of an analysis in which we investigate the effect of teacher assessment and test scores in 6th grade on the test score in 9th grade. The coefficients shown are coefficients from OLS regressions in which the dependent variable is the child's score on the test in 9th grade. The test score and teacher assessment are standardized and standard errors are clustered at the secondary school level. Columns (1) and (2) show that both the test score and teacher assessment in 6th grade are positively and statistically significant related to children's test score in 9th grade. Furthermore, the estimated coefficients suggest that the test score and teacher assessment are not orthogonal. Put together, teacher assessment seems to be able to predict the test score in 9th grade more accurately. This effect is robust to the inclusion of secondary school fixed effects, which suggests that this effect is not specific to certain (characteristics of) schools.[17]

We also estimate the predictive power of the teacher assessment for children's test scores in 9th grade after controlling for the covariation between the teacher's assessment and the test score in 6th grade. This analysis results in an estimated coefficient of 0.0395 that is statistically significant at the 10 percent level. So, even after controlling for the covariation between the teacher's assessment and the test score, the assessment seems to be more predictive of later test scores compared to the 6th grade test score. It seems likely that the teacher's assessment also captures other skills, besides intelligence, that are important determinants of children's school career.

[16] In appendix F we look at the type of school children attend. We see that most of the track switching takes place within the pre-vocational education tracks. This is consistent with findings from the Dutch Education Inspectorate. Furthermore, in the switching analyses we control for the type of school (level of comprehensiveness) children attend.

[17] The test in 9th grade is limited (8 to 11 questions) and is a low-stakes test. The correlation between children's test score in 6th grade and their test score in 9th grade is 0.16 and statistically significant at the 1 percent level. For the analyses in Table 6 one could argue that the teacher assessment is better at predicting children's test score in 9th grade as the teacher assessment likely captures other elements next to children's ability that affect children test score on a low-stakes test more compared to their test score on a high-stakes test. The test score in 6th grade could be a better predictor of children's test score on a high-stakes test but such information is unavailable.

Table 7 shows the results of an analysis in which we explain the effects of switching on 9th grade test scores. We compare children who have switched tracks to children who have not. The coefficients are from OLS regressions with the test score in 9th grade as the dependent variable. The standard errors are clustered at the secondary school level. Table 7 has three panels. The top panel presents estimates for overall switching, the middle panel presents estimates for switching down and the bottom panel presents estimates for switching up. In the top panel the coefficients for switching show that for children in the pre-vocational track (GL in Figure 1), the pre-higher education track and the pre-university track there is a statistically significant negative effect of switching tracks on the 9th grade test score. Second, the coefficients displayed in the middle panel of the table suggest that children in the pre-vocational track (GL) and children in the pre-higher education track experience a statistically significant negative effect of switching down tracks. Finally, in the bottom panel of Table 7 the displayed coefficients suggest that children in the pre-vocational education track (TL) and children in the pre-university track experience a statistically significant negative effect of switching up tracks. Overall, it seems to be the case that children who have switched tracks experience a negative effect on their test score in 9th grade compared to children who have not switched tracks. [18]

7. Policy perspective

The transition from elementary to secondary school is monitored and accompanied by both objective ability assessments, such as test scores, and the subjective elementary school teacher assessment about the child's ability. Secondary schools base their decision about the allocation of children to tracks on these two ability signals. The mistakes that could be made in this system are, among others, the difference between initial track placement and track allocation in 9th grade. The extent to which children switch tracks serves as one indicator. We conduct three types of analyses to address the possible costs of mistakes in initial allocation. They should be seen as thought experiments or what-if analyses in which we ignore heterogeneity across subjects and selection issues.[19]

First, we show by how much switching tracks between 7th and 9th grade would have gone down if children would have been allocated in accordance with the subjective assessment measure, i.e., $TA_i = TP_i$. Our previous analyses have suggested that the subjective assessment seems to be a better measure of effective track placement and seems to be correlated with lower levels of

[18] Here we assume that track switches are caused by poor performance and not related to unobservable characteristics, such as illness or behavioral problems.

[19] These analyses do not serve as an overall assessment of the early tracking system in the Netherlands, because we have no information about the possible benefits of early tracking for our population.

switching between 7th and 9th grade. In our sample 19 percent receives a signal $TS_i \neq TA_i$. The first experiment is to examine by how much track switching would fall if $TA_i = TP_i$ in case of $TA_i < TS_i$.[20] Of the population $TA_i < TS_i$ 42.2 percent switches tracks. Of those children allocated according to $TA_i = TP_i$ 36.4 percent switch tracks. If we assume that those children for which $TA_i \neq TP_i$, the same percentage would apply, switching would go down by 10.7 percent as 47.1 percent of children switch tracks when $TA_i \neq TP_i$. If we conduct the same experiment for $TA_i > TS_i$, switching would go down by 19 percent and for $TS_i = TA_i$ it results in a 24.7 percent fall. Overall switching would be reduced by 22.4 percent to a level of about 18 percent compared to almost 24 percent now.

Second, the difference between both assessment measures varies (i.e., more than one track level distance between both signals) for some children. This induces uncertainty about the child's ability and makes track placement in 7th grade more difficult. We observe that these children have a higher probability of switching in their first three years of secondary education. 32.9 percent of our sample receives such signals based on a split of the data into 11 tracks. The allocation across tracks is the following: 10.1 percent of the children is placed in a track according to the test score, 70.5 percent of the children is placed in a track according to the teacher assessment and 19.4 percent of children is placed in a track neither according to the test score nor the teacher assessment, when $TA_i \neq TS_i$. When $TA_i = TS_i$ 89.7 percent of children is placed according to the assessment measures and 10.3 percent of children is placed in a track not according to the assessment measures. If children in the 'else' category were placed according to their test score switching would have gone up by 11 percent. If these children were placed according to their teacher assessment switching would have gone down by 10 percent. Hence, it seems beneficial to allocate children for which the two ability signals differ according to their teacher's advice.

Finally, if the system would allow switching to take place in the first year of secondary school only, the number of switchers would have been reduced by 64.4 percent. Of course, this is not a fair measure because a fraction of the children who switch during or after the second year of secondary education would have probably switched immediately after or even during 7th grade. The analysis we conduct is the following. The data show that 7th and 8th grade children who are placed in tracks according to "else" are most likely to switch tracks in the first two years of secondary school. If one would introduce a policy allowing children to switch in 7th grade only, this would be easier to accomplish if children are allocated to tracks based on one of the two ability signals. In addition, if we assume that all switching has to take place in 7th grade, we would most likely observe more track

[20] This is based on a count of 1,158 overall switches.

placement according to teacher assessment because this turns out to be the signal in which switching is lowest. This would suggest that putting a higher weight on the teacher's assessment of the child's ability would be beneficial to minimize switching. Of course, this way of reasoning does not take into account possible selection issues.

8. Conclusions

This paper documents and interprets the determinants of track placement in the transition from primary to secondary education and the first three years of secondary education. Our main findings suggest that both objective and subjective assessment measures of ability in 6th grade predict track placement in 7th grade. The subjective assessment measure of the teacher contains more information as the teacher has more knowledge about the child's socio-economic background, the objective test score in 6th grade and previous test scores and other results. Our estimates suggest that the teacher's assessment in elementary school is a better predictor of a child's track placement and subsequent performance in secondary school compared to the 6th grade test score. We also observe that approximately 24 percent of our population of children switches tracks between 7th and 9th grade. We obtain a set of estimates that suggests that children are the least likely to switch when they are allocated to a track based on the teacher's assessment. However, when we look at switching down and up separately our estimates suggest that children placed in tracks according to teacher assessment are more likely to switch down and children placed in tracks according to the test score and children allocated not according to any of these assessment measures are more likely to switch up. Finally, test scores in 9th grade seem to be better predicted by teacher assessment compared to 6th grade test scores. In addition, switchers obtain lower test scores on this 9th grade test relative to children who remain in their initial tracks.

This research uses a straightforward research design and explores a dataset which includes information on assessment measures, track placement and subsequent performance. Future work could extend our analysis by using for example more detailed information about different parts of the objective assessment measure. Test scores could be decomposed in a language and math component, which could benefit the analysis of allocation and performance in secondary school. In addition, the relationship with individual characteristics is interesting to explore further. We have used a limited number of covariates because of data limitations, but future data collection efforts also include measures of behavior and personality traits. This information could help in estimating more precise coefficients and possibly additional determinants of performance and allocation. Finally, the children in our database are followed throughout the rest of their educational careers. This opens avenues for future research about longer term effects of track allocation.

References

Bernardi, M., M. Bratti and G. De Simone (2014). "I wish I knew…" – Misperceived ability, school track counselling services and performances in upper secondary education, IZA Discussion Paper no. 7940, http://ftp.iza.org/dp7940.pdf.

Bosch, N., A. Deelen and R. Euwals (2010). "Is part-time employment here to Stay? Working hours of Dutch women over successive generations, *Labor*, vol. 24, no. 1, pp. 35-54.

Brunello G. and D. Checchi (2007). Does school tracking affect equality of opportunity? New international evidence, *Economic Policy*, vol. 22, no. 52, pp. 781-861.

Burgess, S. and E. Greaves (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities, *Journal of Labor Economics*, vol. 31, no. 3, pp. 535-576.

Cito (2012). *Terugblik en resultaten 2012. Eindtoets Basisonderwijs Groep 8*, Cito, Arnhem.

Cornwell, C., D.B. Mustard and J. van Parys (2012). Noncognitive skills and the gender disparities in test scores and teacher assessments: evidence from primary school, *Journal of Human Resources*, vol. 48, no. 1, pp. 236-264.

Currie, J. and D. Thomas (1999). Early test scores, socioeconomic status and future outcomes, NBER Working Paper no. 6943, http://www.nber.org/papers/w6943.

Dee, T.S. (2005). A teacher like me: does race, ethnicity or gender matter? *American Economic Review*, vol. 95, no. 2, pp. 158-165.

Diris, R. (2012). *The economics of the school curriculum*, PhD thesis, Maastricht University.

Education Inspectorate (2007). Transition from primary to secondary education, http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2007/Aansluiting+vo+bo+printversie.pdf

van Elk, R., M. van der Steeg and D. Webbink (2009). The effect of early tracking on participation in higher education, CPB Document no. 182, http://www.cpb.nl/en/publication/effect-early-tracking-participation-higher-education.

Fairlie, R.W., Hoffmann, F and Oreopoulos, P. (2014). A community college instructor like me: race and ethnicity interactions in the classroom. *American Economic Review,* vol. 104, no. 8, pp. 2567-2591.

Gibbons, S. and A. Chevalier (2008). Assessment and age 16+ education participation, *Research Papers in Education*, vol. 23, no. 2, pp. 113-123.

Golsteyn, B.H.H. and T. Schils (2014). Gender gaps in primary school achievement: A decomposition into endowment and returns to IQ and non-cognitive factors, *Economics of Education Review*, 41: 176-187

Hanushek, E. and L. Woessmann (2006). Does early tracking affect educational inequality and performance? Differences-in-differences evidence across countries, *Economic Journal*, vol. 116, no. 510, pp. c63-c76.

Kautz, T., J.J. Heckman, R. Diris, B. ter Weel and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success, OECD, Paris, http://www.oecd.org/edu/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-

Cognitive-Skills-to-Promote-Lifetime-Success.pdf.

Lavy, V. (2008). Do gender stereotypes reduce girls' human capital outcomes? Evidence from a natural experiment, *Journal of Public Economics*, vol. 92, no. 10-11, pp. 2083-2105.

Lindahl, E. (2007). Comparing teachers' assessments and national test results - evidence from Sweden, IFAU Working Paper no. 2007:24.

Murnane, J., J.B. Willett and F. Levy (1995). The growing importance of cognitive skills in wage determination, *Review of Economics and Statistics*, vol. 77, no. 2, pp. 251-266.

OECD (2007). *Thematic review of tertiary education - The Netherlands*, Country Note, OECD, Paris.

Segal, C. (2008). Classroom behavior, *Journal of Human Resources*, vol. 43, no. 4, pp. 783-814.

Timmermans, A., H. Kuyper and G. van der Werf (2012). *Schooladviezen en onderwijsloopbanen. Voorkomen, risicofactoren en gevolgen van onder- en overadvisering*, Gronings Instituut voor Onderwijs van Onderwijs, Rijksuniversteit Groningen.

Woessmann, L. (2004). Institutional comparisons in educational production, *Journal for Institutional Comparisons*, vol. 2, no. 4, pp. 3-6.

Table 1: The distribution of test scores, teacher assessment and track placement

	Track										
	T1a	T1a/b	T1b	T1b/c	T1c	T1c/d	T1d	T1d/2	T2	T2/3	T3
TS_i	2.00	2.62	3.87	2.76	6.33	9.07	7.11	13.98	20.49	18.96	12.82
TA_i	4.00	3.27	4.53	0.31	0.11	5.18	10.96	12.33	18.02	16.62	24.67
TP_i	2.69	3.11	7.27	0.00	1.53	1.84	10.73	10.84	11.07	25.93	24.98
SW_i	3.59	0.21	11.30	0.00	7.29	5.28	19.01	10.35	12.99	9.93	20.06
TS_i 9th	1.21	2.21	4.14	0.00	1.88	1.38	11.70	11.31	9.82	29.14	27.12

Note: All rows add up to 100. $n = 4{,}500$ except for SW_i where $n = 4{,}019$ and TS_i 9^{th} where $n = 1{,}812$. See Figure 1 for a schedule of the different educational tracks. TS_i is the objective assessment measure based on the *Cito Eindtoets* (test score) in 6th grade. TA_i is the subjective assessment measure based on the teacher advice in 6th grade. TP_i is track placement in 7th grade. SW_i is a measure of switching. The numbers in this row show the fraction of children that switch away from this track after initial placement in 7th grade. TS_i 9th is the test score in 9th grade on math or language.

Table 2: Descriptive statistics of the main variables in the empirical analysis

	N	Mean	SD	Min	Median	Max
Female	4,422	0.51	0,50	0	1	1
Age (years)	4,500	12.06	0.55	10	12	14.75
Region of birth child 1. Limburg 2. The Netherlands 3. Abroad	3,943	1.15	0.46	1	1	3
Region of birth father	3,923	1.36	0.67	1	1	3
Region of birth mother	3,935	1.35	0.67	1	1	3
Education father 1. Lower education 2. Vocational education 3. Higher education 4. University	3,752	2.38	1.10	1	2	4
Education mother	3,763	2.22	1.03	1	2	4
Labor market position father 1. Employed 2. Unemployed 3. Sick/unable to work 4. Other	3,858	1.12	0.51	1	1	4
Labor market position mother	3,869	1.48	1.04	1	1	4
Workdays per week father 1. Part-time 1-2 days 2. Part-time 3-4 days 3. Fulltime	3,249	2.91	0.29	1	3	3
Workdays per week mother	2,848	2.31	0.62	1	2	3
Living with both parents	3,608	0.79	0.41	0	1	1
Teacher assessment	4,500	5.60	2.15	1	6	8
Track placement	4,500	5.75	2.16	1	7	8
Test score	4,500	538.67	8.03	504	540	550
Test score (short)	4,500	5.21	2.02	1	6	8
Switching	4,019	0.24	0.42	0	0	1

Note: The same description of region of birth of the child applies to the father and mother. Similarly, the same description of Education father, Labor market position father and Workdays per week father apply to the mother.

Table 3: Correlation coefficients

	Teacher assessment	Test score	Track placement	Switching
Teacher assessment	1			
Test score	0.8495***	1		
Track placement	0.8896***	0.8040***	1	
Switching	-0.2247***	-0.1973***	-0.2309***	1

Note: Bivariate correlation coefficients. ***, ** represents significance at the 1 and 5 percent level respectively. Teacher Assessment, track placement and test score (short) are all measured on a scale from 1-8. All other variables are measured as described in table 2.

Table 4: The determinants of track placement (dependent variable TP_i)

	Track placement											
	7th grade						9th grade					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Test score (TS)	1.517***		0.687***	0.762***	1.102***	0.980***	1.007***	1.501***	1.285***	0.854***	1.037***	1.086***
	(0.093)		(0.106)	(0.087)	(0.142)	(0.263)	(0.072)	(0.115)	(0.280)	(0.076)	(0.110)	(0.199)
Teacher assessment (TA)		1.998***	1.490***	1.463***	1.847***	1.941***	1.690***	2.173***	2.097***	1.030***	1.612***	1.481***
		(0.246)	(0.249)	(0.246)	(0.279)	(0.278)	(0.206)	(0.209)	(0.273)	(0.249)	(0.183)	(0.204)
Cut1	-2.994***	-4.029***	-4.162***	-4.255***	-5.875***	-11.266***	-6.168***	-9.515***	-11.529***	-4.152***	-7.459***	-11.350***
Cut2	-2.027***	-2.410***	-2.553***	-2.796***	-4.115***	-9.060***	-3.564***	-6.732***	-8.574***	-2.863***	-5.378***	-8.905***
Cut3	-1.712***	-1.890***	-2.031***	-2.011***	-3.419***	-8.408***	-3.009***	-5.391***	-7.637***	-2.181***	-3.968***	-7.841***
Cut4	-0.962***	-0.889***	-0.994***	-0.811	-1.895***	-7.245***	-1.246**	-3.476***	-5.669***	-0,705	-2.703***	-6.233***
Cut5	-0.365*	-0.160	-0.216	0.062	-0.523	-5.483***	1.001*	-0,601	-2.575***	0.785**	-0,629	-3.885***
Cut6	0.151	0.432**	0.414**	0.651	0.323	-4.719***						
Cut7	1.322***	1.651***	1.734***	2.199***	1.720**	-3.319**						
Test score/(C5-C1)	0,351		0,117	0,118	0,145	0,123	0,140	0,168	0,144	0,173	0,152	0,145
Teacher assessment/(C5-C1)		0,352	0,253	0,227	0,243	0,244	0,236	0,244	0,234	0,209	0,236	0,198
Difference TS-TA (p-value)			0,012	0,011	0,030	0,025	0,001	0,002	0,056	0,562	0,018	0,229
Controls	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes
Primary school FE	no	no	no	no	yes	yes	no	yes	yes	no	yes	yes
GPAt-1,2,3	no	no	no	no	no	yes	no	no	yes	no	no	yes
Observations	4,500	4,500	4,500	2,509	1,100	572	1,708	872	469	2,509	1,100	572

Note: Coefficients are from ordered probit regressions with track placement as the dependent variable. The dependent variable in columns1-6 is track placement in 7th grade for all children. The dependent variable in columns 7-12 is track placement in 9th grade, where columns 7-9 are for children who have not switched tracks and columns 10-12 are for all children. The dependent variable in 9th grade consists only of 6 tracks as over time the combination tracks disappear and children get allocated to a final track. Standard errors are clustered at the secondary school level. ***, **, * represent 1, 5 and 10 percent significance levels respectively. TS and TA are standardized. The control variables are gender, age, birth region of the child, birth region of father and mother, education level of father and mother, labor market position of father and mother, work days per week of father and mother, if the child lives with both parents, area code of residence and region. $GPA_{t-1,2,3}$ is the child's GPA in elementary school in previous years.

Table 5: Switching between tracks

	Switch			Switch down			Switch up		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PL=TA≠TS	0.099***	0.093***	0.098***	0.103***	0.091***	0.085***	-0.017	-0.010	-0.000
	(0.029)	(0.030)	(0.032)	(0.026)	(0.025)	(0.027)	(0.012)	(0.010)	(0.007)
PL=TS≠TA	0.319***	0.161***	0.179**	0.102***	0.024	0.025	0.252**	0.096***	0.084**
	(0.101)	(0.049)	(0.078)	(0.038)	(0.043)	(0.038)	(0.117)	(0.036)	(0.042)
PL≠TA≠TS	0.269***	0.197***	0.167**	0.131**	0.090	0.079	0.164*	0.105***	0.059*
	(0.100)	(0.065)	(0.081)	(0.060)	(0.073)	(0.062)	(0.097)	(0.038)	(0.035)
Controls	no	no	yes	no	no	yes	no	no	yes
Secondary school FE	no	yes	yes	no	yes	yes	no	yes	yes
Observations	4,019	3,836	2499	4,019	3,664	2281	4,019	3,031	1750

Note: Probit regressions with marginal effects. Standard errors are clustered at the secondary school level. ***, **, *, at 1 percent, 5 percent and 10 percent significance levels respectively. The base level in these regressions is when PL=TA=TS and when PL=TA & PL=TS but TA≠TS. This can happen for example when $TA = T3, TS = T2$ and $PL = T2/T3$. In column 1, the base level represents 52,97 percent of the children. The three other categories shown in the table represent 30,28 percent, 5,8 percent and 10,95 percent respectively. We control for gender, age, birth region of the child, birth region of the father and mother, education level of the father and mother, labor market position of father and mother, if the child lives with both parents, area code of residence and region.

Table 6: Determinants of scores on math or language tests in 9th grade

	math or language test 9th grade			
	(1)	(2)	(3)	(4)
Test score	0.027*		0.006	0.002
	(0.014)		(0.015)	(0.016)
Teacher advice		0.071***	0.067***	0.060***
		(0.016)	(0.018)	(0.018)
Controls	yes	yes	yes	yes
Secondary school FE	no	no	no	yes
Observations	1,116	1,116	1,116	1,116
Adjusted R-squared	0,072	0,082	0,082	0,085

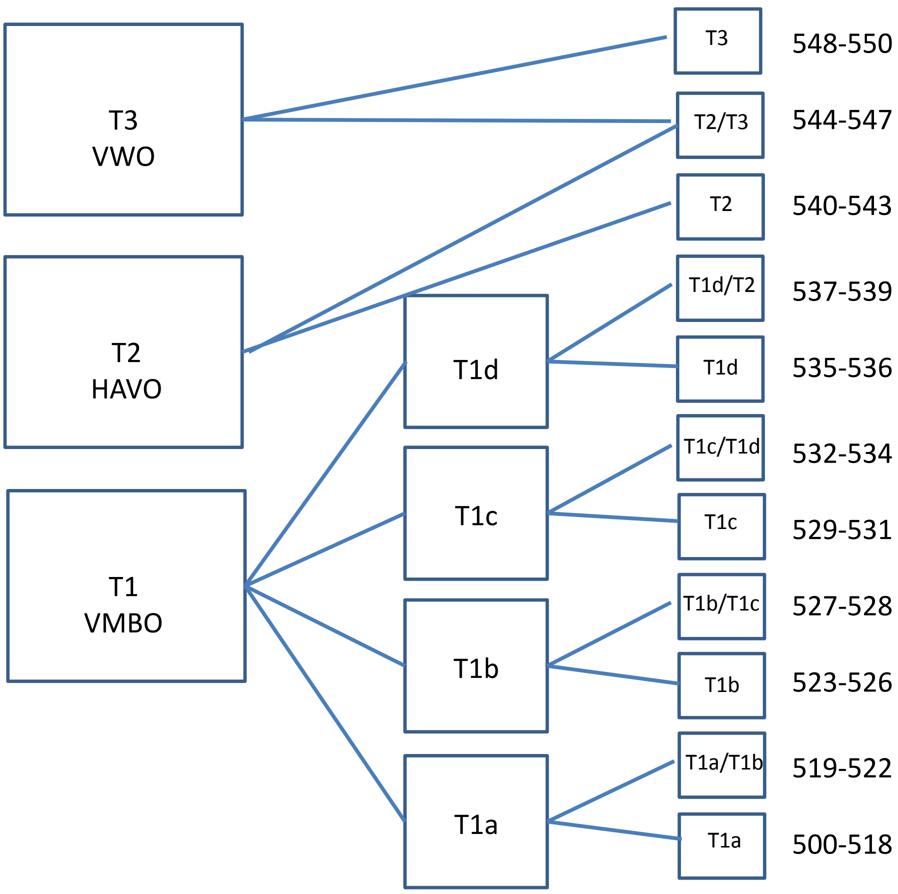
Note: Estimates are coefficients from OLS regressions. The dependent variable is the children's score on a math or language test in 9th grade. Test score and teacher advice are standardized. Standard errors are clustered at the secondary school level. ***, **, *, at 1 percent, 5 percent and 10 percent significance levels respectively. We control for track placement in 7th grade, gender, age, region of birth of child, father and mother, education level of father and mother, labor market position of father and mother, days worked of father and mother, if the child lives with both parents, area code of residence and region.

Table 7: The impact of switching tracks on math and language tests in 9th grade

	Test in 9th grade: math or language											
	Pre-vocational education (GL)			Pre-vocational education (TL)			Pre-higher education			pre-university education		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Switch	-0.142*	-0.528**	-0.543**	0.027	-0.000	0.023	-0.032	-0.097***	-0.102**	-0.073	-0.240**	-0.183*
	(0.045)	(0.119)	(0.123)	(0.030)	(0.091)	(0.106)	(0.020)	(0.029)	(0.035)	(0.043)	(0.089)	(0.089)
Controls	no	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes
Secondary school FE	no	no	yes	no	no	yes	no	no	yes	no	no	yes
Observations	88	38	38	317	155	155	603	303	303	732	415	415
Adjusted r-squared	0.059	0.092	0.022	0.002	0.048	0.081	0.002	0.042	0.054	0.004	0.034	0.041
Switch down	-0.103**	-0.528**	-0.543**	-0.001	-0.000	0.023	-0.039	-0.160**	-0.162**			
	(0.018)	(0.119)	(0.123)	(0.025)	(0.091)	(0.106)	(0.025)	(0.060)	(0.064)			
Controls	no	yes	yes	no	yes	yes	no	yes	yes			
Secondary school FE	no	no	yes	no	no	yes	no	no	yes			
Observations	88	38	38	317	155	155	603	303	303			
Adjusted r-squared	0.032	0.092	0.022	-0.003	0.014	0.078	0.003	0.054	0.064			
Switch up				0.044*	-0.166**	-0.204***	0.022	0.079	0.110	-0.073	-0.240**	-0.183*
				(0.023)	(0.055)	(0.040)	(0.045)	(0.071)	(0.084)	(0.043)	(0.089)	(0.089)
Controls				no	yes	yes	no	yes	yes	no	yes	yes
Secondary school FE				no	no	yes	no	no	yes	no	no	yes
Observations				317	155	155	603	303	303	732	415	415
Adjusted r-squared				0.061	0.049	0.079	-0.001	0.031	0.051	0.004	0.034	0.041

Note: Estimates are coefficients from OLS regressions. The dependent variable is the children's score on a math or language test in 9th grade. Standard errors are clustered at the secondary school level. ***, **, *, at 1 percent, 5 percent and 10 percent significance levels respectively. We control for test score and teacher advice in 6th grade, track placement in 7th grade, gender, age, region of birth of child, father and mother, education level of father and mother, labor market position of father and mother, days worked of father and mother, if the child lives with both parents, area code of residence and region.

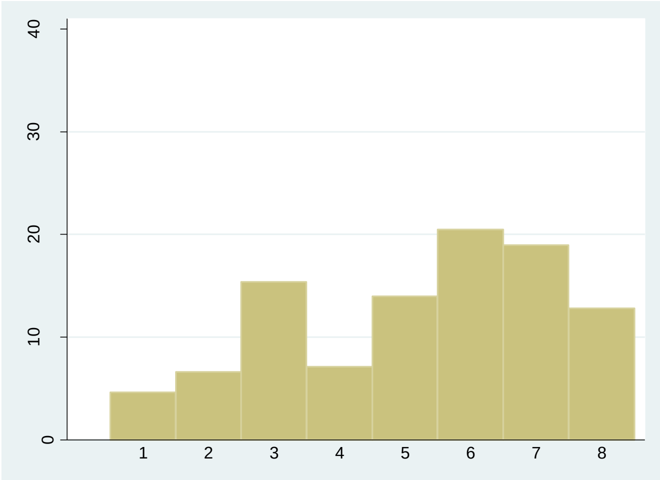
Figure 1: Tracks in secondary education in The Netherlands



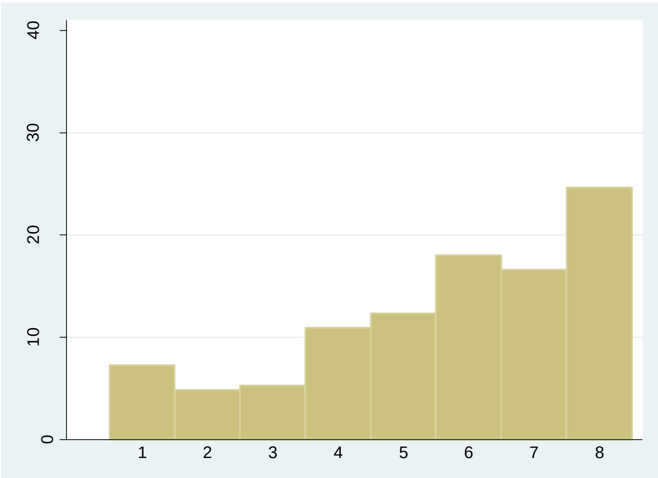
Note: The left-hand side shows the three major tracks from high (T3) to low (T1). The T1 track is subdivided into four sub-tracks. The right-hand side of the figure shows the 11 tracks to which children can be allocated in 7th grade and the objective assessment measure (test score) in the brackets that belongs to each of these tracks.

Figure 2: The distribution of objective and subjective assessment measures in 6th grade and track placement in 7th grade in 8 brackets (from low to high)

A: Objective assessment measure (test score) in 6th grade



B: Subjective assessment measure (teacher assessment) in 6th grade



C: Track placement in 7th grade

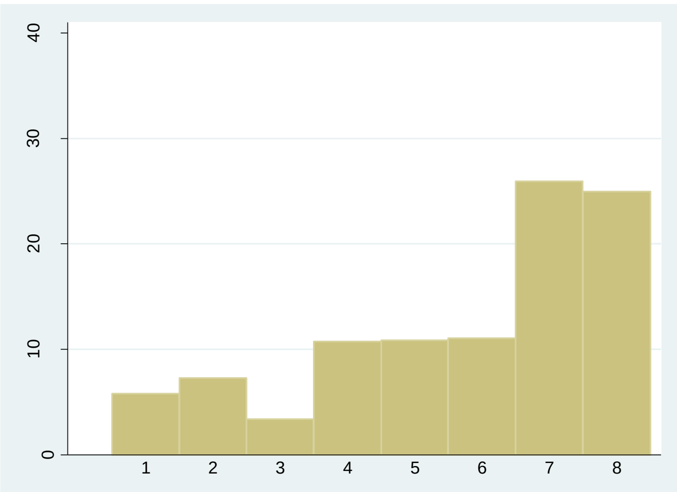
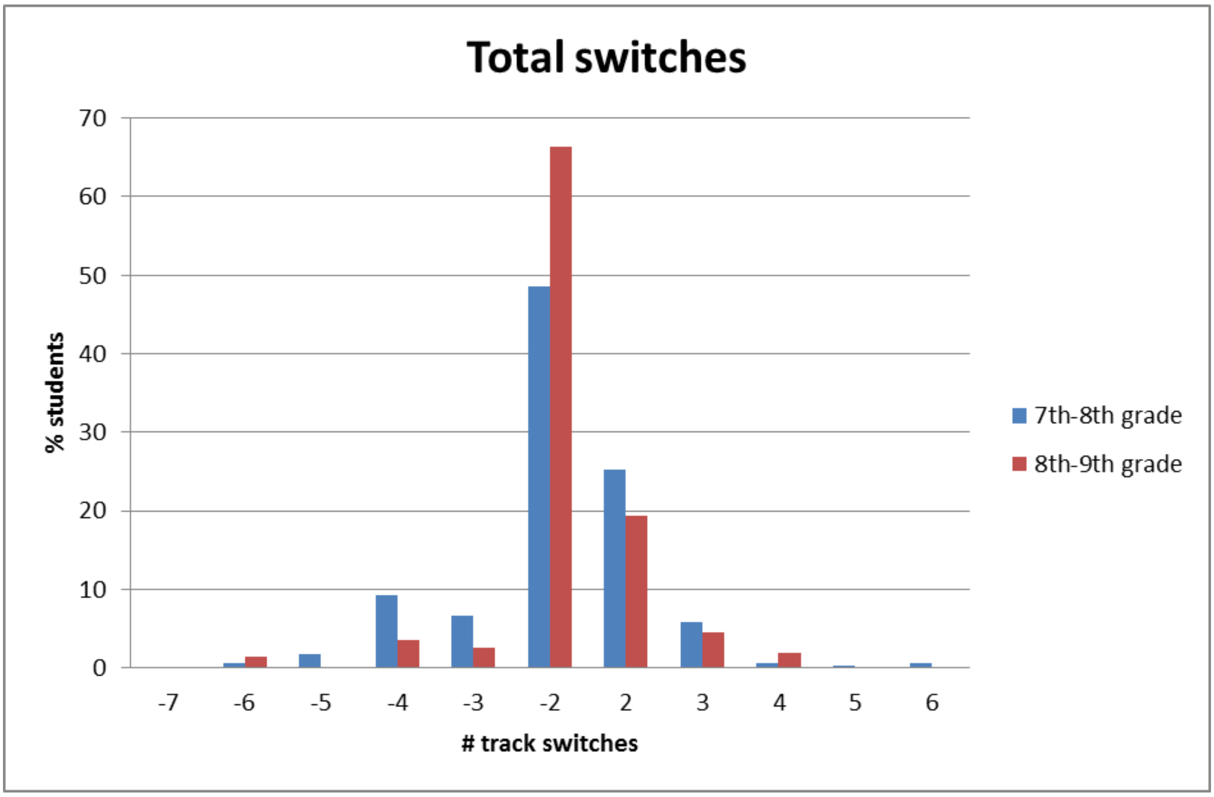
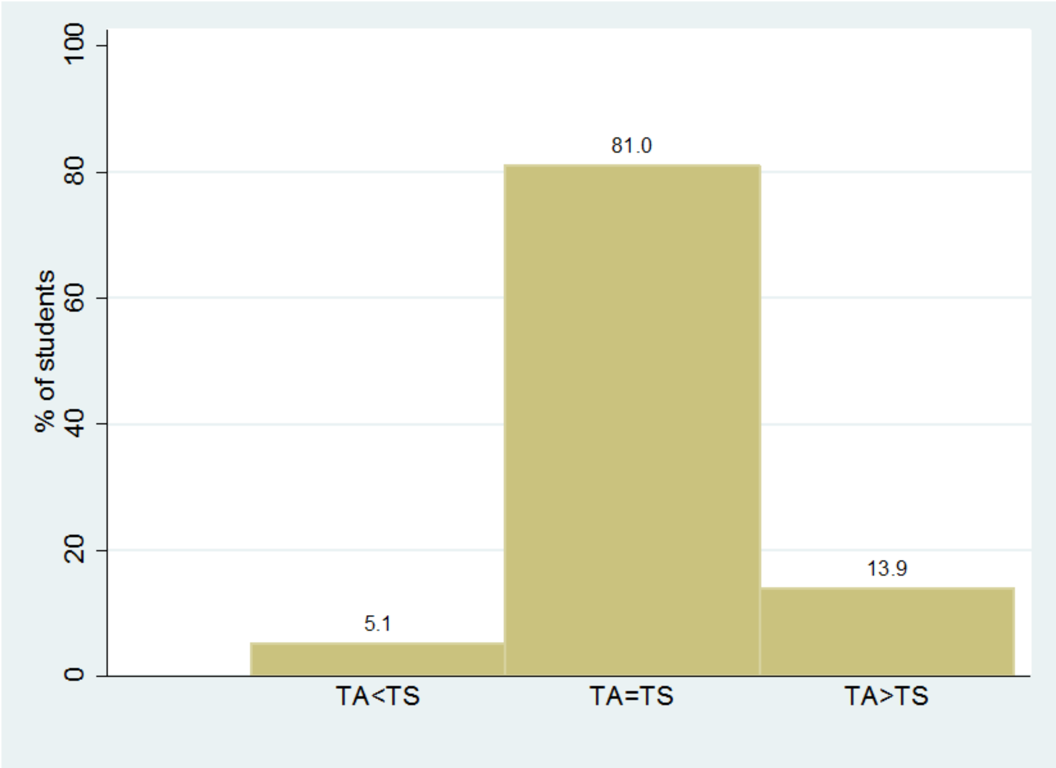


Figure 3: Track switching between initial track allocation in 7th grade and 9th grade



Note: The figure shows the total number of switchers. Negative numbers on the horizontal axis are defined as switches down and positive numbers as switches up. The blue and red bars add up to 100 percent individually.

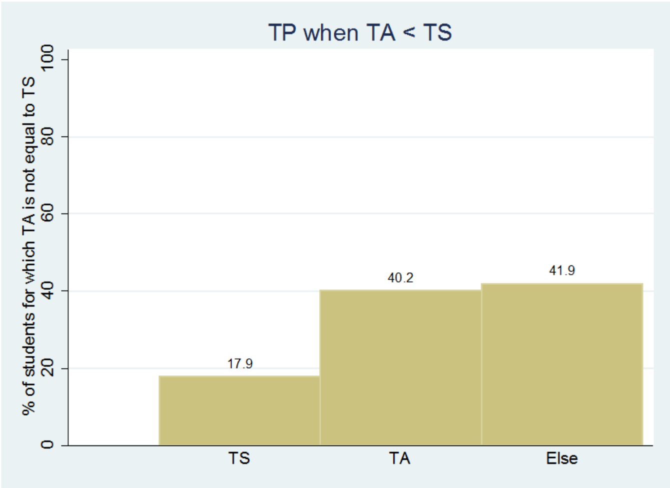
Figure 4: Similarities of and differences between objective and subjective assessment measures



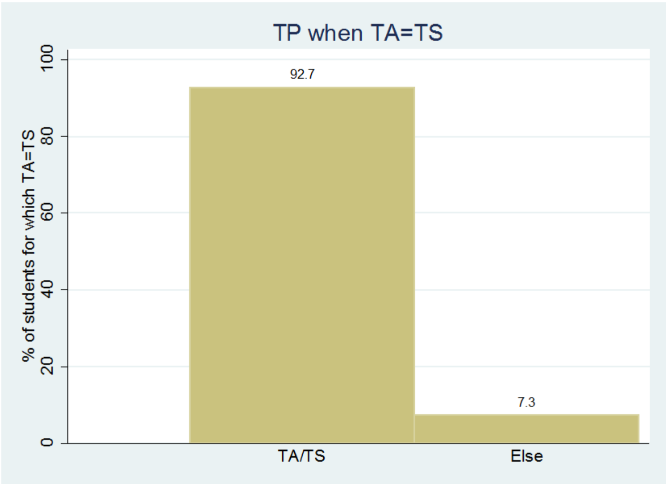
Note: $n = 4{,}500$. The three columns add up to 100 percent. Objective assessment measures are test scores in 6th grade (TS_i); subjective assessment measures are teacher assessments in 6th grade (TA_i).

Figure 5: Track placement for different objective and subjective assessment measures

A: $TA_i < TS_i$



B: $TA_i = TS_i$



C: $TA_i > TS_i$

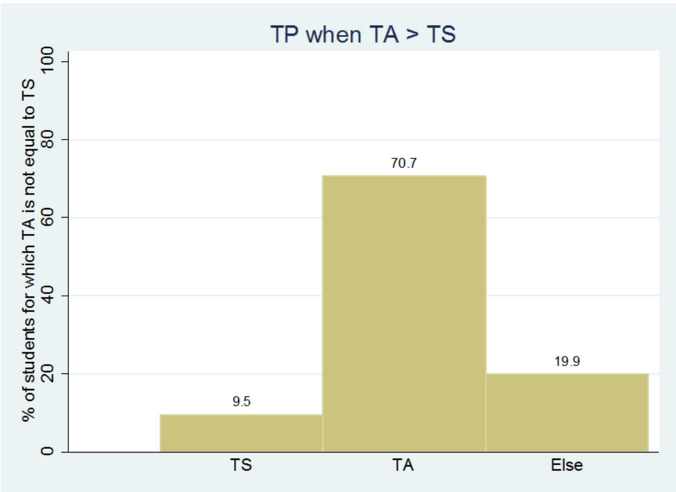
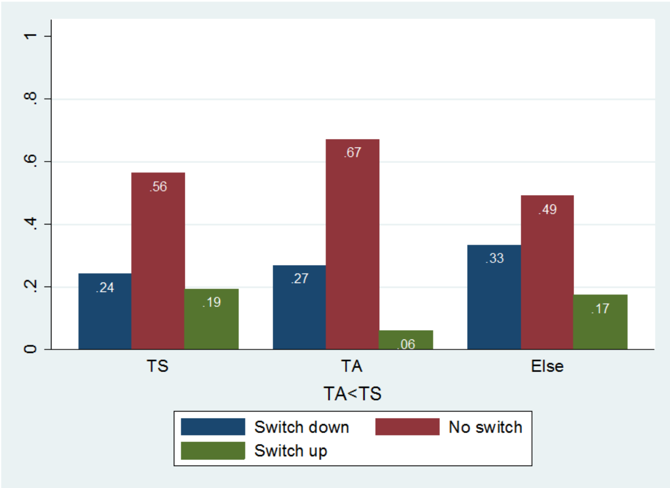
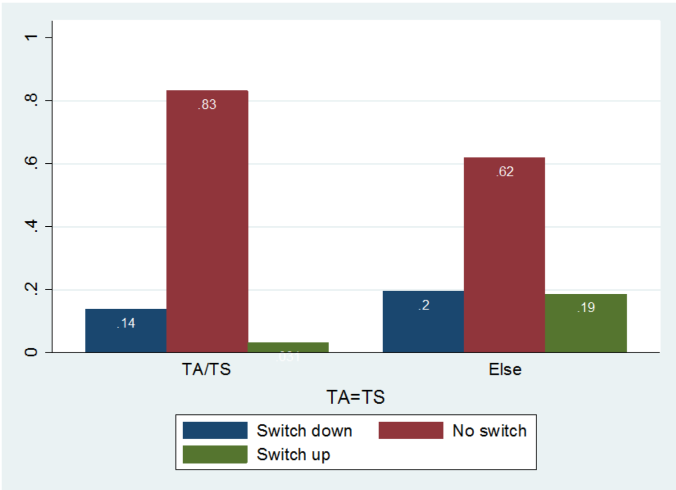


Figure 6: Track switching for different objective and subjective assessment measures

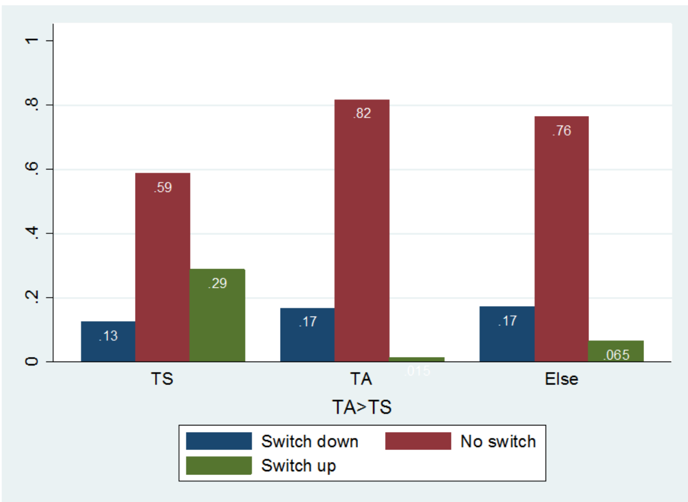
A: $TA_i < TS_i$



B: $TA_i = TS_i$



C: $TA_i > TS_i$



Appendix

This appendix presents supporting material. The data come from a unique dataset on the educational development of children in a given region (Limburg) of the Netherlands, collected by researchers from Maastricht University. These data are collected in a cooperative project with (primary and secondary) schools, school boards and municipalities to analyze school performance in order to foster educational improvement. After selecting children for whom we have information on their test score, teacher assessment and track placement we retain a working sample of $n = 4{,}500$ children.

In Section A we discuss the bivariate correlation coefficients of all of our outcomes measures and covariates used in our analyses. In section B we investigate whether or not individual characteristics are able to predict whether children end up in our sample of analyses. Section C1 presents an initial overview of whether the differences between the test score, the teacher assessment and track placement are related to individual characteristics. Section C2 follows up on Section C1 and presents the marginal effects of the differences between the test score, the teacher assessment and track placement. In Section D1 we move to track switching where we first provide an overview of how switching is related to individual characteristics. We distinguish between switching up and down tracks. Afterwards in Section D2 we observe the marginal effects of probit models for switching.

A. Correlations

Table A shows the bivariate correlation coefficients of our main variables of interest, TS_i , TA_i and TP_i as well as the covariates used in our analyses. We observe that girls seem to have a slightly lower test score compared to boys. Second, older children seem to have a lower teacher assessment, test score and track placement while they seem to switch tracks more. This could be due to repetition. Third, children who have a father or mother that was born further away from Limburg seem to have a lower teacher assessment, test score and track placement. Furthermore, children whose mother was born further away from Limburg seem to switch track more. Fourth, the education level of the father and mother is positively related to the teacher assessment, test score and track placement of the child and negatively related to switching tracks. Fifth, it seems that there is a negative relationship between the father and mother not doing paid work and teacher assessment, test score and track placement of the child. Sixth, it seems that there is a positive relationship between the child living with both parents and the teacher assessment, test score and track placement. Seventh, workdays per week seems to be negatively correlated with teacher assessment, test score and track placement. The group of children that receives the highest teacher assessment, test score and track placement is largest when the father and/or mother work 3-4 days per week. Correlations among teacher assessment, test score and track placement are high and positive.

Table A: Correlation coefficients

	Female	Age (years)	Region of birth child	Region of birth father	Region of birth mother	Education father	Education mother	Labour market position father	Labour market position mother	Living with both parents	Workdays per week father	Workdays per week mother	Teacher assessment	Track placement	Test score (short)	Test score	Switching
Female	1																
Age (years)	-0.0931***	1															
Region of birth child	-0.0299	0.0540***	1														
Region of birth father	-0.0025	0.0722***	0.4563***	1													
Region of birth mother	-0.0125	0.0847***	0.4498***	0.5683***	1												
Education father	-0.0229	-0.1100***	0.0962***	0.0633***	0.0506***	1											
Education mother	-0.0127	-0.1285***	0.0724***	0.0568***	0.0719***	0.5400***	1										
Labour market position father	0.0369**	0.0889***	0.0549***	0.1272***	0.1066***	-0.0675***	-0.0618***	1									
Labour market position mother	0.0062	0.0596***	0.0837***	0.1163***	0.1416***	-0.0148	-0.1004***	0.0810***	1								
Living with both parents	0.0052	-0.0656***	-0.0402**	-0.0733***	-0.0288	0.0760***	0.0637***	-0.1123***	0.0168	1							
Workdays per week father	-0.0231	0.0449**	-0.0076	-0.0250	-0.0491***	-0.0751***	-0.0631***	-0.0679***	0.0531***	0.0062	1						
Workdays per week mother	-0.0304	0.0693***	0.0324	0.1124***	0.1112***	-0.0674***	-0.0117	0.0762***	-0.0945***	-0.1722***	0.0426**	1					
Teacher assessment	-0.0019	-0.2290***	-0.0040	-0.0532***	-0.0487***	0.2936***	0.3187***	-0.0857***	-0.0742***	0.1011***	-0.0439**	-0.0829***	1				
Track placement	-0.0005	-0.2138***	-0.0083	-0.0528***	-0.0377**	0.2832***	0.3097***	-0.0892***	-0.0742***	0.1060***	-0.0508***	-0.0836***	0.8896***	1			
Test score (short)	-0.0389***	-0.1954***	-0.0200	-0.0719***	-0.0576***	0.2539***	0.2860***	-0.0795***	-0.0782***	0.0998***	-0.0411**	-0.0835***	0.8495***	0.8040***	1		
Test score	-0.0382**	-0.2058***	-0.0191	-0.0739***	-0.0566***	0.2537***	0.2887***	-0.0903***	-0.0853***	0.0929***	-0.0383**	-0.0860***	0.8569***	0.8183***	0.9727***	1	
Switching	-0.0248	0.0887***	0.0088	0.0403	0.0419**	-0.1337***	-0.1324***	0.0320	-0.0108	-0.0658***	0.0313	0.0323	-0.2247***	-0.2309***	-0.1973***	-0.1780***	1

Note: Bivariate correlation coefficients. ***, ** represents significance at the 1 and 5 percent level respectively. Teacher Assessment, track placement and test score (short) are all measured on a scale from 1-8. All other variables are measured as described in Table 2.

B. Selection into the sample

We address selection issues with regard to the sample we use for our analyses in Table B. Table B shows OLS estimates of whether characteristics of children are able to predict whether the child ends up in our sample for analyses or has missing data. The dependent variable is whether the child ends up in our sample (1) or has missing data (0). Children who end up in our sample for analyses are children for whom we have data on teacher assessment, test score and track placement in 7th grade.

Table B shows that after controlling for primary and secondary school fixed effects and clustering the standard errors at the secondary school level there are almost no individual characteristics that predict whether children end up in our sample for analysis or not. We use primary and secondary school fixed effects since we are concerned mainly about selection within schools, where schools only provide us with data about their best performing children and might leave out data about lower performing children. We expect that such selection would show up on children's SES data, such as education levels of parents. We cluster standard errors at the secondary school level since observations of children are not independent within the secondary school they attend. We do not find evidence at a five percent significance level that any of the characteristics of children is able to predict whether the child ends up in our sample for analyses. Since we do not find any observable characteristics of children that predict whether the child ends up in our sample, we assume that there are also no unobservable characteristics that distinguish the children in our sample from the children who are not in our sample. Note that there are missing observations in our analyses. This is due to the fact that we do not have information on all covariates for all children. Since we are specifically interested in whether any of these covariates predict selection into our sample of analyses we lose observations when data on covariates is missing. Furthermore, we only have elementary school fixed effects for children in the south of the region we obtained the data from.

Table B: Selection into the sample

	Selection into sample				
	(1)	(2)	(3)	(4)	(5)
Female	-0.005	-0.009	0.002	-0.006	-0.006
	(0.013)	(0.017)	(0.008)	(0.009)	(0.008)
Age 2nd quarter	-0.002	0.022	0.003	0.021	0.021*
	(0.019)	(0.024)	(0.011)	(0.013)	(0.011)
Age 3rd quarter	-0.032*	-0.002	-0.029***	0.020*	0.020*
	(0.018)	(0.024)	(0.011)	(0.012)	(0.011)
Age 4th quarter	0.033*	-0.003	0.005	0.014	0.014
	(0.019)	(0.025)	(0.011)	(0.013)	(0.011)
Student born in other region NL	-0.055*	-0.007	-0.028	-0.020	-0.020
	(0.029)	(0.045)	(0.018)	(0.023)	(0.023)
Student born abroad	-0.015	-0.037	-0.087***	-0.025	-0.025
	(0.044)	(0.060)	(0.027)	(0.031)	(0.029)
Father born in other region NL	-0.026	0.034	-0.020	0.029*	0.029*
	(0.022)	(0.032)	(0.013)	(0.017)	(0.016)
Father born abroad	-0.037	0.012	-0.032*	0.007	0.007
	(0.029)	(0.036)	(0.018)	(0.019)	(0.015)
Mother born in other region NL	0.013	-0.014	-0.005	-0.022	-0.022
	(0.022)	(0.032)	(0.013)	(0.017)	(0.021)
Mother born abroad	0.050*	0.014	0.015	-0.010	-0.010
	(0.030)	(0.037)	(0.018)	(0.019)	(0.017)
Father has vocational education	0.040**	0.020	0.019*	-0.011	-0.011
	(0.017)	(0.023)	(0.011)	(0.012)	(0.013)
Father has higher education	0.071***	0.079***	0.016	-0.002	-0.002
	(0.022)	(0.029)	(0.013)	(0.015)	(0.013)
Father has university education	0.080***	0.105***	0.031**	0.006	0.006
	(0.023)	(0.030)	(0.014)	(0.016)	(0.014)
Mother has vocational education	0.034**	0.079***	0.019*	0.025**	0.025*
	(0.017)	(0.023)	(0.010)	(0.012)	(0.014)
Mother has higher education	0.071***	0.096***	0.012	0.016	0.016
	(0.023)	(0.030)	(0.014)	(0.016)	(0.016)
Mother has university education	0.092***	0.095***	0.004	0.020	0.020
	(0.025)	(0.033)	(0.015)	(0.017)	(0.018)
Father is unemployed	-0.224**	-0.352**	-0.035	-0.191**	-0.191
	(0.111)	(0.168)	(0.068)	(0.088)	(0.137)
Father is sick/unable to work	0.081	-0.116	0.008	-0.030	-0.030
	(0.111)	(0.152)	(0.067)	(0.080)	(0.027)
Father other labour market position	-0.201	-0.086	-0.098	0.003	0.003
	(0.127)	(0.148)	(0.077)	(0.078)	(0.040)
Mother is unemployed	-0.014	0.089	0.006	-0.119	-0.119
	(0.100)	(0.169)	(0.061)	(0.089)	(0.086)
Mother is sick/unable to work	0.066	-0.013	0.033	-0.065	-0.065
	(0.126)	(0.152)	(0.077)	(0.079)	(0.049)
Mother other labour market position	0.108**	-0.059	0.053	-0.035	-0.035
	(0.054)	(0.069)	(0.033)	(0.036)	(0.035)
Student lives with both parents	0.027	0.037*	0.006	0.010	0.010
	(0.017)	(0.021)	(0.010)	(0.011)	(0.010)
Father works parttime 3/4 days per week	0.138	0.193	0.115	0.186*	0.186
	(0.120)	(0.211)	(0.073)	(0.110)	(0.183)
Father works fulltime	0.124	0.181	0.130*	0.198*	0.198
	(0.118)	(0.209)	(0.071)	(0.109)	(0.182)
Mother works parrtime 3/4 days per week	-0.014	-0.011	0.002	0.004	0.004
	(0.024)	(0.034)	(0.015)	(0.018)	(0.016)
Mother works fulltime	-0.056**	-0.049	-0.018	0.002	0.002
	(0.025)	(0.035)	(0.015)	(0.018)	(0.019)
Student has repeated a grade in secondary school	0.258***	0.087	0.044**	0.013	0.013
	(0.028)	(0.120)	(0.018)	(0.064)	(0.019)
Constant	-0.162	0.533	-0.554*	0.893***	0.893***
	(0.474)	(0.337)	(0.291)	(0.204)	(0.236)
Primary school FE	no	yes	no	yes	yes
Secondary school FE	no	no	yes	yes	yes
SEs clustered at the SS	no	no	no	no	yes
Observations	4,861	1,872	4,861	1,872	1,872
R-squared	0.186	0.561	0.707	0.884	0.884

Note: Coefficients are from OLS regressions. The dependent variable has value 1 when the child is included in the analyses sample and value 0 when the child has missing data on teacher assessment or test score in 6th grade or track placement in 7th grade.

C1. Similarities of and differences between assessment measures and allocation

Table C1 shows a number of statistics. The left-hand panel shows the relation between different individual characteristics and the objective and subjective assessment measure, the middle panel shows the relation between individual characteristics and the subjective assessment measure and track placement, the right-hand panel shows the relation between individual characteristics and the objective assessment measure and track placement.[21]

We observe that girls more often seem to receive a teacher assessment that is higher compared to the test score, while they less often seem to receive a teacher assessment that is lower compared to the test score. In addition, girls more often receive a track placement equal to the subjective assessment measure and more often have a test score that is lower compared to their track placement. Second, the education level of the child's mother and father has a similar relation to the assessment measures and track placement. We observe that when a child's parents have a higher education level the child more often receives a subjective assessment that is higher than the objective assessment and less often receives a subjective assessment that is lower than the objective assessment compared to when the child's parents have a lower level of education. The middle and right-hand panel show that children of parents who have a higher education level less often receive a subjective and objective assessment that is higher than track placement. This suggests that girls and children from parents who are highly educated not only get a more favorable assessment, compared to boys and children from parents who are lower educated, from the elementary school teacher but also from the secondary school with regard to track placement. Third, from the labor market position of the mother we observe that children who have mothers that are unemployed more often receive a subjective assessment that is lower than the objective assessment and less often receive a subjective assessment that is higher than the objective assessment. Furthermore, children of mothers who are sick/unable to work less often receive a subjective assessment that is lower than track placement and more often receive an objective and subjective assessment that is higher than track placement. In our data over 70 percent of the mothers who are unemployed and over 85 percent of the mothers who are sick\unable to work completed only lower education or vocational education. Fourth, we observe a pattern related to the region of birth of the children and their parents. Children or parents born in another part of the Netherlands outside of Limburg or born abroad more often receive a subjective assessment higher than the objective assessment and more often receive a subjective and objective assessment that is not equal to track placement. In

[21] In Table 4 we disregard the labor market position of the father and the workdays per week of the father. Almost all fathers are employed and work fulltime. There are too little observations in the other categories to make any statements about this information.

most of the cases when the assessment measures are not equal to track placement these children receive assessment measures that are lower than track placement. This is especially the case for children or parents born abroad. Finally, we observe a pattern in children's birth quarters. There is an unofficial cut-off date for children's birthdays where children born after 1 October often have to wait another year to start primary education. Therefore, the oldest children are born in the fourth quarter. We observe that younger children more often receive a subjective assessment that is higher than the objective assessment. These children also more often receive a subjective assessment that is higher than their track placement and an objective assessment that is lower than their track placement. When we look at the 6th grade test scores of these children we see that test scores rise with children's age. Workdays per week mother and whether the child lives with both parents does not show any pattern.

Table C1: Similarities of and differences between assessment measures and allocation

		Objective and subjective assesment measures			Subjective assessment measures and track placement			Objective assessment measures and track		
		TAi < TSi	TAi = TSi	TAi > TSi	TAi < TPi	TAi = TPi	TAi > TPi	TSi < TPi	TPi = TSi	TSi > TPi
Female	Boys	6.19	81.70	12.11	7.81	87.34	4.85	18.48	75.60	5.91
	Girls	4.21	80.34	15.46	7.17	89.33	3.50	21.21	74.84	3.94
Education father	Lower education	8.14	79.04	12.82	9.66	85.96	4.37	19.13	73.96	6.92
	Vocational	5.32	81.31	13.37	7.89	87.62	4.49	20.43	73.92	5.65
	Higher education	3.95	81.21	14.83	5.93	91.38	2.68	19.77	76.69	3.53
	University	3.15	82.61	14.24	6.18	90.20	3.62	20.77	76.66	2.57
Education mother	Lower education	7.93	78.63	13.44	8.12	86.94	4.93	19.83	73.21	6.96
	Vocational	5.44	81.06	13.50	8.26	87.84	3.90	19.88	75.22	4.90
	Higher education	3.72	83.66	12.62	5.66	91.26	3.07	17.96	77.83	4.21
	University	2.57	82.48	14.95	6.59	90.84	2.57	22.35	75.08	2.57
Labour market position father	Employed	5.42	81.07	13.52	7.57	88.50	3.92	19.73	75.46	4.81
	Unemployed	1.32	85.53	13.16	7.89	86.84	5.26	27.63	60.53	11.84
	Sick/unable to work	5.15	81.44	13.40	6.19	90.72	3.09	18.56	78.35	3.09
	Other	5.97	79.10	14.93	5.97	85.07	8.96	16.42	77.61	5.97
Labour market position mother	Employed	5.49	81.02	13.49	7.81	88.33	3.86	19.68	75.25	5.07
	Unemployed	7.06	84.71	8.24	7.06	89.41	3.53	16.47	77.65	5.88
	Sick/unable to work	4.38	80.63	15.00	3.75	89.38	6.88	16.88	75.00	8.13
	Other	4.70	81.39	13.91	6.95	89.37	3.68	22.29	74.44	3.27
Workdays per week father	Fulltime	5.78	80.80	13.42	7.60	88.43	3.97	19.70	75.22	5.08
	Parttime 3-4 days	2.99	82.09	14.93	7.84	89.93	2.24	20.15	77.99	1.87
	Parttime 1-2 days									
Workdays per week mother	Fulltime	6.00	79.75	14.25	8.42	86.65	4.93	19.35	75.27	5.38
	Parttime 3-4 days	5.16	81.98	12.86	7.64	89.42	2.95	19.69	75.22	5.09
	Parttime 1-2 days	5.44	81.17	13.39	8.37	87.87	3.77	18.83	77.82	3.35
Living with both parents	No	5.35	80.81	13.84	6.92	88.38	4.70	19.06	75.07	5.87
	Yes	5.52	81.18	13.30	7.60	88.60	3.80	19.99	75.05	4.96
Region of birth child	Limburg	5.41	81.87	12.72	7.31	89.07	3.61	19.52	75.44	5.04
	Netherlands	5.68	75.38	18.94	8.33	84.09	7.58	18.94	75.38	5.68
	Abroad	5.45	75.15	19.39	10.30	82.42	7.27	24.85	69.70	5.45
Region of birth father	Limburg	5.60	81.59	12.81	7.74	88.72	3.53	19.74	75.10	5.16
	Netherlands	4.21	81.50	14.29	4.95	90.11	4.95	16.48	78.75	4.76
	Abroad	5.54	77.83	16.63	9.24	84.53	6.24	23.79	71.36	4.85
Region of birth mother	Limburg	5.32	81.90	12.78	7.23	89.13	3.65	19.04	75.88	5.09
	Netherlands	4.83	80.69	14.48	7.14	87.45	5.41	19.69	75.10	5.21
	Abroad	6.54	76.87	16.59	9.81	85.05	5.14	24.30	70.79	4.91
Birth quarter child	1st	5.02	68.13	26.85	9.86	84.93	5.21	33.70	60.91	5.39
	2nd	5.81	67.08	27.10	9.66	83.63	6.71	32.47	60.82	6.71
	3rd	5.21	64.75	30.04	8.83	83.48	7.69	35.60	56.98	7.42
	4th	6.23	68.57	25.19	10.22	85.11	4.68	33.25	60.43	6.32

Note: $n = 4{,}500$ in total but some children have missing data on some of the covariates. The numbers in the rows in each section of columns add up to 100 percent. TS_i is the objective assessment measure based on the Cito Eindtoets (test score) in 6th grade. TA_i is the subjective assessment measure based on teacher advice. TP_i is track placement in 7th grade. All are measured in 8 brackets. The last row for 'work days per week father' is missing. There are too little fathers that work 1-2 days per week to draw any conclusions based on this information.

C2. Marginal effects of similarities of and differences between assessment and allocation

Table C2 presents marginal effects of differences between the objective and subjective assessment measures and track placement. The left-hand panel shows the relation between different individual characteristics and the objective and subjective assessment measure, the middle panel shows the relation between individual characteristics and the subjective assessment measure and track placement, the right-hand panel shows the relation between individual characteristics and the objective assessment measure and track placement. We include elementary school fixed effects as we want to observe within school differences.[22] In addition, all analyses in this paper include clustered standard errors either at the elementary school or secondary school level since we believe that the observations are not independent within schools. The marginal effects result from an ordered probit model in which we predict the probability that individual characteristics affect differences between assessment measures of differences between an assessment measure and track placement.[23]

First, we observe that girls are less likely to receive a subjective assessment that is lower than or equal to the objective assessment and more likely to receive a subjective assessment that is higher than the objective assessment compared to boys. Furthermore, girls are more likely to receive an objective assessment that is lower than track placement and less likely to receive an objective assessment that is equal or higher than track placement compared to boys. This suggests that elementary schools tend to give a higher assessment to girls compared to boys and that secondary schools tend to place girls in higher tracks compared to boys. Second, children born abroad or fathers born in another part of The Netherlands outside Limburg are less likely to receive an objective assessment that is lower than track placement and more likely to receive an objective assessment that is equal to track placement. In addition, children who have fathers who were born abroad are more likely to receive a subjective assessment that is equal to the objective assessment and less likely to receive a subjective assessment that is higher than the objective assessment. While children who have mothers that were born abroad are more likely to have a subjective assessment that is lower than track placement and less likely to have a subjective assessment that is higher than track placement. Third, children whose mother is unemployed or sick/unable to work are less likely to receive a subjective assessment that is higher than the objective assessment. Furthermore, these

[22] Specifications without elementary school fixed effects show estimates of similar size and statistical significance. This also suggests that different schools do not systematically over or under assess children or children with different characteristics.

[23] In Table 4 we disregard the labor market position of the father and the workdays per week of the father. Almost all fathers are employed and work fulltime. There are too little observations in the other categories to make any statements about this information.

children are also less likely to receive an objective assessment that is lower than track placement. This suggests that primary and secondary schools seem to value when a child's mother is employed. In our sample mothers who are unemployed or sick/unable to work on average have a lower level of education. Fourth, secondary schools seem to value when mothers work fulltime as their children are more likely to receive a subjective assessment that is lower than track placement and less likely to receive a subjective assessment that is equal to or higher than track placement.

Table C2: Marginal effects of similarities of and differences between assessment and allocation

	Marginal effects								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	TA<TS	TA=TS	TA>TS	TA<PL	TA=PL	TA>PL	TS<PL	TS=PL	TS>PL
	-0.015**	-0.043***	0.058***	0.008	-0.006	-0.002	0.068***	-0.062***	-0.006**
Female	(0.006)	(0.014)	(0.020)	(0.011)	(0.009)	(0.002)	(0.026)	(0.023)	(0.003)
	0.012	0.026**	-0.038*	0.014	-0.012	-0.003	0.012	-0.011	-0.001
Age 2nd quarter	(0.008)	(0.013)	(0.020)	(0.019)	(0.016)	(0.004)	(0.037)	(0.034)	(0.003)
	0.001	0.004	-0.006	-0.000	0.000	0.000	0.035	-0.032	-0.003
Age 3rd quarter	(0.006)	(0.017)	(0.024)	(0.015)	(0.012)	(0.003)	(0.037)	(0.034)	(0.003)
	0.006	0.014	-0.020	0.031	-0.027	-0.005	0.033	-0.030	-0.003
Age 4th quarter	(0.008)	(0.016)	(0.023)	(0.024)	(0.021)	(0.004)	(0.040)	(0.037)	(0.003)
	0.008	0.016	-0.024	0.068	-0.062	-0.006	-0.025	0.022	0.003
Student born in NL	(0.014)	(0.019)	(0.033)	(0.047)	(0.044)	(0.004)	(0.043)	(0.037)	(0.005)
	0.003	0.008	-0.012	-0.007	0.005	0.002	-0.095**	0.077***	0.017
Student born abroad	(0.018)	(0.037)	(0.054)	(0.024)	(0.017)	(0.007)	(0.041)	(0.026)	(0.015)
	0.012	0.022*	-0.033	-0.009	0.007	0.002	-0.063*	0.055*	0.008
Father born in NL	(0.012)	(0.013)	(0.025)	(0.016)	(0.011)	(0.005)	(0.035)	(0.029)	(0.006)
	0.014	0.024***	-0.038*	-0.006	0.005	0.001	-0.038	0.034	0.004
Father born abroad	(0.011)	(0.009)	(0.020)	(0.019)	(0.014)	(0.005)	(0.046)	(0.040)	(0.006)
	-0.001	-0.002	0.002	0.003	-0.002	-0.001	0.009	-0.009	-0.001
Mother born in NL	(0.007)	(0.024)	(0.031)	(0.023)	(0.019)	(0.004)	(0.050)	(0.046)	(0.004)
	0.005	0.012	-0.017	0.046*	-0.041*	-0.005**	0.065	-0.061	-0.004
Mother born abroad	(0.013)	(0.025)	(0.038)	(0.026)	(0.025)	(0.003)	(0.064)	(0.061)	(0.003)
	0.005	0.013	-0.018	-0.024*	0.018*	0.006	-0.035	0.031	0.003
Father has vocational education	(0.007)	(0.016)	(0.022)	(0.014)	(0.009)	(0.005)	(0.035)	(0.032)	(0.004)
	-0.000	-0.001	0.001	-0.004	0.003	0.001	0.036	-0.033	-0.003
Father has higher education	(0.007)	(0.022)	(0.029)	(0.017)	(0.013)	(0.004)	(0.050)	(0.047)	(0.003)
	-0.004	-0.012	0.016	-0.018	0.013	0.005	0.036	-0.033	-0.003
Father has university education	(0.006)	(0.024)	(0.030)	(0.015)	(0.010)	(0.005)	(0.048)	(0.044)	(0.003)
	0.001	0.003	-0.004	0.015	-0.012	-0.003	0.006	-0.005	-0.001
Mother has vocational education	(0.006)	(0.019)	(0.025)	(0.019)	(0.016)	(0.004)	(0.039)	(0.036)	(0.003)
	-0.002	-0.005	0.007	-0.023	0.016*	0.007	-0.052	0.047	0.006
Mother has higher education	(0.007)	(0.023)	(0.030)	(0.017)	(0.009)	(0.007)	(0.040)	(0.035)	(0.006)
	-0.005	-0.017	0.022	-0.012	0.009	0.003	-0.006	0.005	0.001
Mother has university education	(0.006)	(0.027)	(0.033)	(0.017)	(0.012)	(0.005)	(0.048)	(0.043)	(0.004)
	-0.016***	-0.344*	0.360**	0.009	-0.007	-0.002	0.419**	-0.411**	-0.008***
Father is sick/unable to work	(0.004)	(0.179)	(0.180)	(0.032)	(0.027)	(0.005)	(0.179)	(0.179)	(0.003)
Father has other labour market	-0.004	-0.016	0.021	-0.046***	-0.269	0.316	-0.089	0.073	0.016
position	(0.027)	(0.135)	(0.162)	(0.010)	(0.224)	(0.222)	(0.122)	(0.079)	(0.043)
	0.431	-0.342	-0.090***	0.398	-0.391	-0.007*	-0.146***	0.084***	0.062
Mother is unemployed	(0.316)	(0.314)	(0.008)	(0.317)	(0.317)	(0.004)	(0.032)	(0.025)	(0.057)
	0.588	-0.498	-0.090***	0.033	-0.029	-0.004	-0.138***	0.088***	0.050*
Mother is sick/unable to work	(0.373)	(0.371)	(0.008)	(0.059)	(0.055)	(0.004)	(0.022)	(0.008)	(0.029)
Mother has other labour market	0.000	0.001	-0.001	0.056	-0.051	-0.005	0.185	-0.178	-0.007**
position	(0.010)	(0.029)	(0.040)	(0.072)	(0.069)	(0.004)	(0.128)	(0.126)	(0.003)
Father works parttime 3/4 days per	0.001	0.003	-0.004	0.024	-0.021	-0.004	0.018	-0.017	-0.001
week	(0.008)	(0.022)	(0.031)	(0.024)	(0.022)	(0.003)	(0.046)	(0.043)	(0.003)
Mother works parttime 3/4 days per	-0.009	-0.026	0.035	0.041	-0.032*	-0.009	0.012	-0.011	-0.001
week	(0.008)	(0.021)	(0.029)	(0.026)	(0.019)	(0.008)	(0.044)	(0.040)	(0.004)
	-0.009	-0.029	0.038	0.077**	-0.065**	-0.013*	0.039	-0.036	-0.003
Mother works fulltime	(0.007)	(0.026)	(0.033)	(0.035)	(0.029)	(0.007)	(0.046)	(0.042)	(0.004)
	0.004	0.015	-0.019	0.005	-0.004	-0.001	0.003	-0.003	-0.000
Student lives with both parents	(0.005)	(0.019)	(0.024)	(0.014)	(0.011)	(0.003)	(0.033)	(0.030)	(0.003)
Observations	1,100	1,100	1,100	1,100	1,100	1,100	1,100	1,100	1,100

Note: Ordered probit regressions with marginal effects. Every three columns represent one ordered probit regression of three levels. Standard errors are clustered at the elementary school level. All estimates include elementary school fixed effects and control variables for area code of residence and region. TS_i is the objective assessment measure based on the Cito Eindtoets (test score) in 6th grade. TA_i is the subjective assessment measure based on teacher advice. TP_i is track placement in 7th grade. All are measured in 8 brackets. ***, **, * represents significance at the 1, 5 and 10 percent level respectively.

D1. Similarities of and switching between tracks

Table D1 shows the percentages of children who switch tracks between 7th and 9th grade.[24] We observe that girls overall switch less than boys, however, there are more girls that switch up and more boys that switch down. Second, the higher the education level of the parents the fewer children switch between tracks. Children who have parents who have obtained lower education or vocational education are more likely to switch between tracks. As mentioned before this might be an artifact of our data since children's track placement is correlated with their parents' education level. Third, we observe that children whose mother is unemployed switch more. Switching both up and down tracks is more common compared to children whose mother is employed. Fourth, we observe that children whose mother works fulltime switch more compared to children whose mother works part-time. Especially switches down appear more often. Fifth, children who do not live with both parents switch more often, both switches up and down are more common. The last three effects are likely related to parents' education level. In at least two thirds of the cases when the child does not live with both parents the parents have either obtained lower education or vocational education. These findings are consistent with children of higher educated parents more often receiving a subjective assessment compared to the objective assessment. Finally, when the children's parents were born abroad the child more often switches down tracks. This is consistent with the finding that children's whose parents were born abroad more often receive a subjective assessment that is higher compared to the objective assessment. We do not observe a pattern in children's quarter of birth.

[24] In Table 5 we disregard the labor market position of the father and the workdays per week of the father. Almost all fathers are employed and work fulltime. There are too little observations in the other categories to make any statements about this information.

Table D1: Similarities of and differences between switching tracks

		Switch		Switch up		Switch down	
		No	Yes	No	Yes	No	Yes
Female	Boys	76.03	23.97	93.34	6.66	81.46	18.54
	Girls	77.63	22.37	92.57	7.43	83.62	16.38
Education father	Lower education	71.94	28.06	90.93	9.07	78.52	21.48
	Vocational	76.98	23.02	92.33	7.67	84.10	15.90
	Higher education	81.40	18.60	94.15	5.85	86.35	13.65
	University	85.54	14.46	95.54	4.46	89.46	10.54
Education mother	Lower education	71.00	29.00	89.86	10.14	78.65	21.35
	Vocational	77.41	22.59	92.85	7.15	83.85	16.15
	Higher education	85.39	14.61	96.23	3.77	88.86	11.14
	University	84.80	15.20	95.08	4.92	89.00	11.00
Labour market position father	Employed	78.05	21.95	93.01	6.99	83.86	16.14
	Unemployed	80.90	19.10	92.13	7.87	87.64	12.36
	Sick/unable to work	75.83	24.17	90.83	9.17	84.17	15.83
	Other	67.57	32.43	94.59	5.41	71.62	28.38
Labour market position mother	Employed	78.02	21.98	93.08	6.92	83.79	16.21
	Unemployed	71.84	28.16	88.35	11.65	82.52	17.48
	Sick/unable to work	79.03	20.97	91.94	8.06	85.48	14.52
	Other	78.24	21.76	93.70	6.30	83.40	16.60
Workdays per week father	Fulltime	77.76	22.24	92.90	7.10	83.60	16.40
	Parttime 3-4 days	81.97	18.03	94.90	5.10	86.39	13.61
	Parttime 1-2 days						
Workdays per week mother	Fulltime	76.50	23.50	92.50	7.50	82.67	17.33
	Parttime 3-4 days	80.05	19.95	94.42	5.58	84.71	15.29
	Parttime 1-2 days	78.33	21.67	91.67	8.33	85.83	14.17
Living with both parents	No	72.69	27.31	92.71	7.29	78.59	21.41
	Yes	79.27	20.73	93.07	6.93	85.01	14.99
Region of birth child	Limburg	77.75	22.25	93.19	6.81	83.53	16.47
	Netherlands	77.51	22.49	89.36	10.64	84.50	15.50
	Abroad	78.97	21.03	94.36	5.64	84.62	15.38
Region of birth father	Limburg	78.34	21.66	93.37	6.63	84.12	15.88
	Netherlands	78.81	21.19	91.52	8.48	84.77	15.23
	Abroad	73.84	26.16	92.05	7.95	80.43	19.57
Region of birth mother	Limburg	78.62	21.38	93.41	6.59	84.41	15.59
	Netherlands	77.65	22.35	91.68	8.32	82.87	17.13
	Abroad	73.63	26.37	91.89	8.11	80.73	19.27
Birth quarter child	1st	77.74	22.26	93.77	6.23	82.56	17.44
	2nd	76.35	23.65	92.20	7.80	82.78	17.22
	3rd	76.46	23.54	92.42	7.58	82.54	17.46
	4th	75.65	24.35	93.38	6.62	81.29	18.71

Note: $n = 4{,}937$. For the analysis of switching we use all children for whom we have information on track placement from 7^{th}-9^{th} grade to keep sample size as large as possible. For 4,019 of these children we have information on their teacher advice (TA_i) and test score (TS_i) in 6th grade. The numbers in each row section add up to 100 percent. The last row for 'work days per week father' is missing. There are too little fathers that work 1-2 days per week to draw any conclusions based on this information.

D2. Marginal effects of similarities of and differences between track switching

Table D2 reports marginal effects of similarities of and differences between switching. The first column presents estimates from overall switching, while the second and third column specify switching up or switching down. [25] We include secondary school fixed effects as we want to observe within school differences. In addition, standard errors are clustered at the secondary school level as the individual observations are not independent within schools. The marginal effects result from probit models in which we predict the probability that individual characteristics affect switching, switching up and switching down.[26]

From Table D2 we observe that girls are less likely to switch and especially less likely to switch down tracks compared to boys. This indicates that elementary schools are correct to more often award girls with a higher subjective assessment than their objective assessment compared to boys and that secondary schools are correct to more often place girls in a higher track than their objective assessment recommended. Second, children born in the Netherlands but outside Limburg are less likely to switch and especially less likely to switch down tracks compared to children born in Limburg. This indicates that the elementary school was correct in more often giving these children a subjective assessment that was higher than the objective assessment. Third, we observe that children whose fathers a higher educated are increasingly less likely to switch between tracks and especially to switch down tracks. This indicates that elementary schools were correct to less often give these children a subjective assessment that is lower than their objective assessment. Furthermore, secondary schools also seem to have been correct when they less often place children of higher educated parents in a track that is lower than the subjective or objective assessment. Finally, children who live with both parents are less likely to switch and especially less likely to switch down tracks. This effect is likely related to the education level of the children's parents. We do not find a pattern for the other covariates in the model.

[25] The estimated coefficients of switch up and switch down do not exactly add up to the estimated coefficient of overall switching. This is due to the fact that there are 65 children that switch both up and down between 7th and 9th grade. These children are captured twice in the switch up/down but only once with regard to overall switching. We keep these children in our analyses as we want to document all switches.

[26] In Table 5 we disregard the labor market position of the father and the workdays per week of the father. Almost all fathers are employed and work fulltime. There are too little observations in the other categories to make any statements about this information.

Table D2: Marginal effects of similarities of and differences between switching

	(1) Switch	(2) Switch up	(3) Switch down
Female	-0.035* (0.020)	0.005 (0.006)	-0.051*** (0.016)
Age 2nd quarter	-0.007 (0.024)	0.001 (0.006)	-0.026 (0.023)
Age 3rd quarter	0.009 (0.031)	-0.001 (0.007)	-0.005 (0.031)
Age 4th quarter	0.017 (0.028)	-0.001 (0.007)	0.011 (0.028)
Student born in other region NL	-0.066** (0.029)	0.005 (0.007)	-0.053* (0.029)
Student born abroad	-0.015 (0.047)	-0.006 (0.009)	-0.030 (0.024)
Father born in other region NL	0.025 (0.035)	0.005 (0.011)	-0.012 (0.026)
Father born abroad	0.017 (0.035)	0.005 (0.015)	0.015 (0.029)
Mother born in other region NL	0.017 (0.033)	0.002 (0.007)	0.019 (0.032)
Mother born abroad	0.011 (0.037)	-0.006 (0.008)	0.018 (0.036)
Father has vocational education	-0.032 (0.022)	0.001 (0.003)	-0.042** (0.019)
Father has higher education	-0.042** (0.021)	0.003 (0.004)	-0.043*** (0.014)
Father has university education	-0.065*** (0.017)	-0.001 (0.006)	-0.063*** (0.019)
Mother has vocational education	-0.009 (0.018)	-0.004 (0.008)	-0.003 (0.023)
Mother has higher education	-0.041 (0.027)	-0.004 (0.008)	-0.036 (0.026)
Mother has university education	-0.017 (0.028)	0.000 (0.008)	-0.026 (0.029)
Father is sick/unable to work	0.155 (0.277)		0.144 (0.293)
Father other labour market position	-0.017 (0.097)	0.054 (0.090)	-0.058 (0.068)
Mother is unemployed	0.064 (0.148)		0.052 (0.131)
Mother is sick/unable to work	-0.091 (0.063)		-0.042 (0.080)
Mother other labour market position	-0.021 (0.060)	-0.009 (0.010)	-0.029 (0.051)
Student lives with both parents	-0.055** (0.023)	0.006 (0.008)	-0.061*** (0.024)
Father works parttime 3/4 days per week	0.931*** (0.008)	0.996*** (0.005)	0.930*** (0.011)
Father works fulltime	0.266*** (0.020)	0.028 (0.031)	0.205*** (0.010)
Mother works parrtime 3/4 days per week	-0.025 (0.041)	-0.013 (0.012)	-0.002 (0.041)
Mother works fulltime	-0.006 (0.039)	-0.009 (0.009)	0.001 (0.039)
Student has repeated a grade in secondary school	-0.055 (0.037)	-0.011 (0.012)	-0.029 (0.040)
Secondary school FE	yes	yes	yes
Observations	2,019	864	1,853

Note: Probit regressions with marginal effects. Standard errors are clustered at the secondary school level. All estimates include control variables for the test score and teacher advice in 6th grade, track placement in 7th grade, area code of residence and region. TS_i is the objective assessment measure based on the Cito Eindtoets (test score) in 6th grad. TA_i is the subjective assessment measure based on teacher advice. TP_i is track placement in 7th grade. All are measured in 8 brackets. ***, **, * represents significance at the 1, 5 and 10 percent level respectively.

E. Robustness check: taking into account the covariation between the teacher assessment (TA) and the test score (TS).

Table E1 reports a robustness check to show that the teacher assessment (TA) has added value in predicting children's track placement in 7th and 9th grade on top of what is already explained by the test score (TS). The table shows estimated coefficients of the predictive value of the teacher assessment for ordered probit models after correcting for the predictive power of the test score. In the first step, we regress track placement on the test score. In the second step, we fix the parameter for the test score that we estimated in the first step. Afterwards, we regress track placement on the teacher assessment, all control and dummy variables and the fixed parameter for the test score. This provides us with the predictive power of the teacher assessment after correcting for the covariation between the teacher assessment and the test score. Columns 1 and 3 show the estimated parameter for the test score that we keep fixed in columns 2 and 4. In columns 2 and 4 we observe that after correcting for the covariance between the teacher assessment and the test score the teacher assessment still has a statistically significant positive coefficient. This indicates that the teacher assessment has added value for track placement in both 7th and 9th grade on top of the predictive power of the test score. Analysis with OLS provide similar results.

Table E1: Robustness check

	Track placement			
	7th grade		9th grade	
	(1)	(2)	(3)	(4)
Test score (TS)	2.104***	2.104	1.960***	1.960
	(0.137)	(0.000)	(0.111)	(0.000)
Teacher assessment (TA)		1.237***		1.028***
		(0.257)		(0.169)
cut1	-2.275	-4.637***	-5.276***	-8.269***
cut2	-0.994	-2.830*	-3.876***	-6.133***
cut3	-0.579	-2.143	-2.858**	-4.649***
cut4	0.469	-0.500	-1.855	-3.288***
cut5	1.545	1.017	-0.003	-0.955
cut6	2.262	1.942		
cut7	3.504**	3.523**		
Controls	yes	yes	yes	yes
Primary school FE	yes	yes	yes	yes
Observations	1,100	1,100	1,100	1,100

F. Robustness check: is switching related to the availability of school types in the region?

Tables F1 and F2 report descriptive statistics about the percentage of children that switch tracks and the level of comprehensiveness of schools. It is possible that it is easier for children to switch tracks when they go to a school that offers all tracks, a comprehensive school. First, the tables show that the most switching happens in schools that only offer pre-vocational tracks and that the least switching happens in schools that offer pre-higher and pre-university education. This can be a mechanical effect from our data since we observe children until 9^{th} grade which means that it is likely that we capture almost all of the switching in the pre-vocational tracks but not all of the switching in the pre-higher education and pre-university tracks. Students in the pre-higher education and pre-university tracks can still switch in later grades. Second, table F2 shows that more children switch in a comprehensive school (both in absolute as well as in relative terms) compared to children in non-comprehensive schools. It is likely that this effect is driven by track switches in pre-vocational education.

Table F1: robustness check, is switching related to the availability of school types in the region? (column totals)

	Switch				
	pre-vocational tracks	pre-higher/pre-uni	theoretical pre-vocational/pre-higher/pre-uni	comprehensive	Total
No	481	1,575	952	771	3,779
	58.16	90.62	77.84	67.10	76.54
Yes	346	163	271	378	1,158
	41.84	9.38	22.16	32.90	23.46
Total	827	1,738	1,223	1,149	4,937
	100.00	100.00	100.00	100.00	100.00

Table F2: **robustness check, is switching related to the availability of school types in the region? (row totals)**

	Switch				
	pre-vocational tracks	pre-higher/pre-uni	theoretical pre-vocational/pre-higher/pre-uni	comprehensive	Total
No	481	1,575	952	771	3,779
	12.73	41.68	25.19	20.40	100.00
Yes	346	163	271	378	1,158
	29.88	14.08	23.40	32.64	100.00
Total	827	1,738	1,223	1,149	4,937
	16.75	35.20	24.77	23.27	100.00

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (070) 3383 380

January 2015 | ISBN 978-90-5833-676-7