Anne van Leest

ON THE *Right* TRACK?

Unravelling the Determinants of Primary School Teachers' Track Recommendations

Anne van Leest

ON THE *Right* TRACK?

Unravelling the Determinants of Primary School Teachers' Track Recommendations

The research in this dissertation was carried out in the context of the Interuniversity Centre for Educational Sciences (ICO).

ico

Cover design & lay-out: Karin van der Vegt, Revista ontwerp & illustratie
Printed by: Gildeprint

ISBN: 978-90-393-7709-3
DOI: https://doi.org/10.33540/2474

On the Right Track?

Unravelling the Determinants of Primary School Teachers' Track Recommendations

Op het Juiste Spoor?

De Determinanten van het Basisschooladvies Ontrafeld

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

vrijdag 25 oktober 2024 des middags te 2.15 uur

door

Anna Maria Cornelia van Leest

geboren op 26 september 1986
te Utrecht

Promotor:
Prof. dr. J.W.F. van Tartwijk

Copromotoren:
Dr. T.E. Hornstra
Dr. J.E. van de Pol

Beoordelingscommissie:
Prof. dr. E.J.P.G. Denessen
Prof. dr. G. Dilaver
Prof. dr. L. Elffers
Prof. dr. J.J.H.M. Janssen
Prof. dr. A.G.J. van de Schoot

The research reported in this dissertation was funded by the Dutch Ministry of Education, Culture and Science (grant number OCW/PromoDoc/1065001).

TABLE OF CONTENTS

Preface (Voorwoord)

You've just started reading my dissertation about equal educational opportunities around the transition from primary to secondary education. Writing a dissertation might seem like a natural next step in my educational career, given my track record: I finished pre-university education (*'vwo'*) in secondary school, and after that completed two bachelors, two masters and a teaching degree. I don't mention this to brag about it (and of course, yes, I am privileged), but this information is just the objective data that an outsider sees and interprets, not the real story behind it. It's not always easy to follow the educational path you want to follow, especially if it means climbing the educational ladder, and that's not always visible from someone's track record.

It seems like only yesterday that we had to take the Cito school leavers' test in Grade 6. It was all anyone could talk about. At that time, the school leavers' test was the main way of determining students' secondary school placement.

I did well in primary school. Okay, I talked too much, couldn't sit still, but had good grades and mostly wanted to hold my kindergarten teacher's hand (for the record, this was when I was 5 years old). I got a score of 546 on my Cito school leavers' test, which meant I could go to pre-university education. However, my Grade 6 teacher had a different opinion. My teacher said I was forgetful. For example, I didn't always bring the right materials for my paper on the brass band from home. She often rescheduled working on this paper from Thursday afternoon to Thursday morning without letting us know in advance. Unlike most of my classmates, I had the books I needed for my paper at home because I was working on it on the computer with my mom (it was not very common at the time to work on your paper on the computer, at least not in primary school).

My teacher recommended me to go to the theoretical track of vocational education (*'mavo/vmbo-t'*). Despite this advice, my parents and I decided I should try pre-university education: "I have never tried that before, so I think I should definitely be able to do that" (this famous quote isn't from Pipi Longstocking by the way, as many people think). Unfortunately for me, the teacher had a good relationship with the secondary school I applied to. She managed to convince the school to place me in the track of her choice instead of the one my parents and I had chosen.

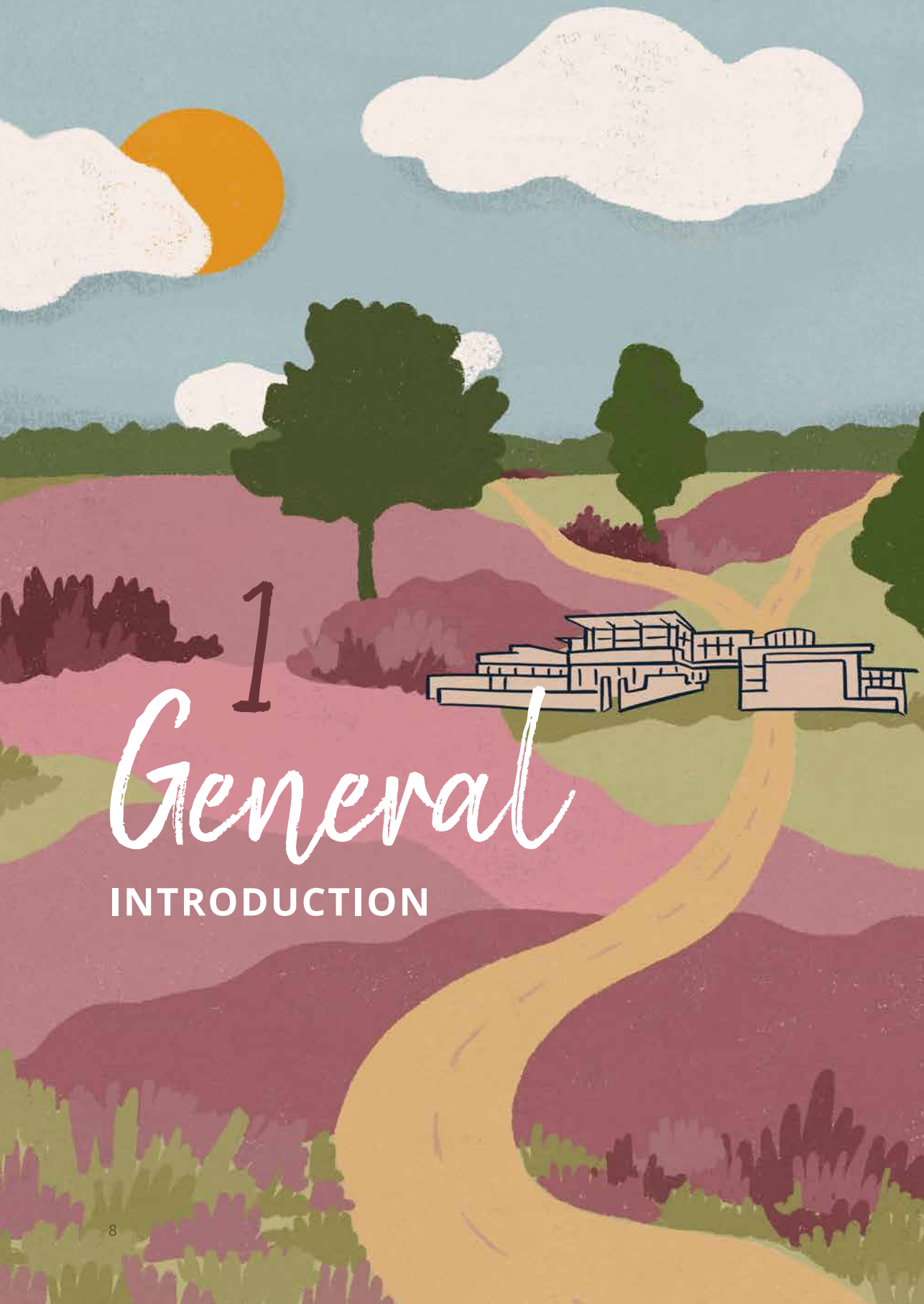
The point I am trying to make with this example is not that we should only look at test scores, or that teachers' expectations are not accurate, but that such key moments in students' lives can leave a deep impression and determine the rest of their lives, not just their (educational) careers. It shows that how we allocate students to a secondary school track does matter. And although this is just a pretty straightforward example that worked out just fine in the end, what if someone did not have the opportunity to follow the educational path that he or she wanted or was able to follow?

In my working life, I have come across this topic many times. As a secondary school teacher of Dutch language and culture, I often saw students having a difficult (or even an unfair or unequal) transition from primary to secondary education, especially students with a different background than myself. This did not only entail 'real examples' of students in my lower secondary school classes, but also 'literature examples' in the Dutch books I discussed with my upper secondary school students. For example, in Murat Isik's *Wees Onzichtbaar*, Karin Amatmoekrim's *Het gym*, Robert Vuijsje's *Alleen maar nette mensen* and Mano Bouzamour's *De belofte van Pisa*. All of these are worth reading, and to share my love for literature with you (I am a Dutch teacher, after all), please see the Appendix on page 206 of this dissertation for these literary examples). What these examples have in common is how the transition from primary to secondary education affects students, particularly in terms of their life path, self-image, and personal growth. And more recently, I met Loes Ypma while I was working as a teacher educator. She started the debate on track recommendations and equal educational opportunities in the Dutch House of Representatives ('Tweede Kamer') in 2015, just before I started my PhD. She was working as a member of the parliament at the time. Her opening speech of this famous debate inspired me to do the research you will read about in Chapter 2. You can read the opening of her (translated) speech in the Appendix on page 214.

As you can imagine, this topic is really close to my heart. While I focus on a macro perspective of the topic in this dissertation, I hope to also work on more student-centred perspectives on equal educational opportunities at the transition from primary to secondary education in my future work. I want to explore how we can make sure that every student is on their right track, however we want to define that.

To be continued!

The quotes at the end of each chapter were part of interviews with Grade 6 teachers. These interviews were part of a larger research project. Teachers checked the manuscripts and gave permission to use (quotes from) them. The original Dutch quotes can be found in the Appendix on page 216.



1 General

INTRODUCTION

Introduction

Do we recommend students to the 'right' secondary school track at the end of primary education? That is a question that has dominated the debate on equal educational opportunities for students in the Netherlands since halfway the 20th century (Lek & van de Schoot, 2019; OECD, 2016b; Timmermans et al., 2018). Tracks differ in the educational qualifications students can acquire, which determine their options for tertiary education (Boone & Demanet, 2020; Glock et al., 2015; Korpershoek et al., 2016), their job prospects and, as a consequence, their future social economic positions (Reed et al., 2015; Rodrigues et al., 2018; G. M. Strand, 2020). A track recommendation is considered 'right' when it gives an optimal reflection of students' potential for educational performance in secondary education (Boone & Van Houtte, 2013; Klapproth et al., 2012; Korpershoek et al., 2016). Because of the consequences of the track placement, the decision which track is the right track for a student should be taken extremely careful (Glock et al., 2012; Korpershoek et al., 2016; Timmermans, Kuyper, et al., 2015).

From a meritocratic perspective, it is important that all students have equal chances to receive a particular track recommendation, taking into account their achievement levels (OECD, 2016b; Themelis, 2008). Hence, two students with similar levels of achievement in primary school but different backgrounds should be given similar track recommendations. In fact, the reason why many countries introduced (high-stakes) standardised tests was to ensure that students were evaluated based on their abilites rather than their background (Au, 2013; Faasse et al., 1987; Luijkx & De Heus, 2008; for a full overview of the historical background of the Dutch educational system and the allocation process, see the Appendix of this Introduction on pages 24 to 29). However, research showed that teachers' track recommendations are sometimes biased by students' background characteristics (Boone & Van Houtte, 2013; Klapproth et al., 2013; Timmermans et al., 2018). In different educational systems, different ways of formulating track recommendations are used to refer students to the right track and avoid bias based on students' backgrounds.

The debate on the allocation process mainly centers around the question to what extent track recommendations should be based entirely on (objective) standardised tests versus (more subjective) teacher judgements. To better understand to what extent track recommendations should be based on standardised tests or on teacher judgements, and to be able to estimate how student background characteristics play a role in track recommendations, more insight is needed into the interplay between student achievement, their background characteristics, and other student attributes, such as classroom behaviour and

teacher-student relationship. Earlier research has mainly focused on the separate contribution of these factors on track recommendations. However, this may not tell the complete story. Questions regarding the interplay between student achievement, perceived student attributes, and student background in predicting track recommendations remain unanswered.

Therefore, the first aim of this dissertation was to investigate the extent to which teachers consider student achievement, perceived student attributes, and student background when formulating track recommendations, and how these factors jointly predict teachers' track recommendations. Since previous research mainly examined the role of individual factors in track recommendations, this dissertation contributes to the existing knowledge on track recommendations by also including the interplay between these factors. Additionally, the studies in this dissertation also examine differences among teachers in their consideration of these factors when formulating track recommendations.

Furthermore, since the goal of a track recommendation is to give an optimal reflection of students' potential educational performance in secondary education (Boone & Van Houtte, 2013; Glock et al., 2012; Klapproth et al., 2012; Korpershoek et al., 2016; Timmermans, Kuyper, et al., 2015), it is also important to know which factors are predictive of students' secondary school performance. Therefore, the second aim of this dissertation was to examine how predictive these factors actually are for students' performance in secondary school.

Literature Overview

Hierarchically tracked educational systems for secondary education, based on ability level, exist in, for example, Hong Kong, Singapore, Great-Britain, Hungary, Germany, Austria, Switzerland, Luxembourg, France, Belgium, and the Netherlands (Le Métais, 2003; Naayer et al., 2016). In most of these systems, the academically more challenging tracks are referred to as 'higher' levels of education.[1] When and how students are selected for these tracks varies from one educational system to another, depending on differences in the structure and typology (Le Métais, 2003; LeTendre et al., 2003; Motiejunaite-Schulmeister, Akvile Sicurella & Birch, 2022). Although the distinction is not black-and-white but rather a continuum, two main ways of recommending students to different tracks can be distinguished: test-based track recommendations and judgement-based track recommendations.

Test-Based Track Recommendations

In countries with test-based track recommendations, students' test results are used as entry requirements for specific levels of education (Ho & Lee, 2022). Countries, such as China, Singapore, and Latvia, have such test-based recommendation systems (Glock et al., 2012; Le Métais, 2003; Naayer et al., 2016). In many of these countries, students complete

so-called school leavers' tests at the end of primary school. The results on such a test automatically determine the secondary school track that is considered as the 'right' track for a student (Boone & Van Houtte, 2013; Driessen et al., 2008; Timmermans, Kuyper, et al., 2015). Until 2014, before a policy reform, the Dutch allocation system could be considered as a test-based system as well.

The OECD (2016)[2] argued that using the results of a standardised national test as primary indicator for track recommendations will have a positive impact on educational equality, because all students make the same standardised test, ideally administered under similar conditions (Boone & Van Houtte, 2013; Knoester & Au, 2017). Hence, for all students, the standard is set at the same level to ensure that students are judged solely on their achievement, regardless of other student characteristics such as their background.

Judgement-Based Track Recommendations

Countries, such as Germany, Luxembourg, France, and (recently) the Netherlands, have judgement-based recommendation systems, meaning that track recommendations are based on teacher expectations at the end of primary education about students' potential achievement (Babad et al., 1982; Klapproth et al., 2012). Research has shown that judgement-based track recommendations are usually primarily based on students' (prior) achievement in terms of test results as well. Test results have been found to explain up to 80% of the variance in track recommendations (Feron et al., 2013; Luyten, 2004; Timmermans, Kuyper, et al., 2015). Students' prior achievement does not only refer to their test scores on the school leavers' test, but also, for example, on their achievement on standardised tests throughout primary school (Driessen, 2005; Feron et al., 2016; Scheerens et al., 2011). In the Netherlands, teachers can obtain this information from a student monitoring system.

Proponents of a judgement-based system argue that a broader view of student achievement is important for selecting the 'right' level of secondary education (Boone & Van Houtte, 2013; Driessen, 2005). It is argued that students' results on a school leavers' test do not give a full and accurate view of students' abilities, because, for example, students may perform rather inconsistently in different subject domains (Luyten, 1998) or may have different achievement growth curves during primary school (Caro, Lenkeit, et al., 2009; Curby et al., 2009; Helbling et al., 2019). Furthermore, a judgement-based system also gives teachers the opportunity to consider other, more 'subjective', factors when formulating track recommendations, such as the student attributes work habits, classroom behaviour

1 Words like 'higher' and 'lower' in the context of education can have the connotation of more or less valuable. In this dissertation, the use of higher and lower are chosen for the sake of readability, but it is not intended to pass judgement on the value of different educational tracks.

2 The OECD is the Organization for Economic Co-operation and Development.

and parental support or involvement, that can also be predictive of future school success (Fredricks et al., 2004; Geven et al., 2018; van Nuland, 2011). These aspects could be valuable complements to the use of standardised test scores in the formulation of track recommendations (Feron et al., 2013), as teachers have in-depth knowledge of students due to their daily interaction in the classroom (de Boer et al., 2010). Research has shown that teachers indeed include their perceptions of different student attributes in their judgement-based track recommendations (Driessen et al., 2008; Sneyers et al., 2018; Timmermans et al., 2016, 2019).

Opponents of the judgement-based system argue that teachers may not be able to accurately assess the importance or direction of the 'subjective' attributes of students. Therefore, their perceptions may be inaccurate or even biased toward certain groups of students (Brandmiller et al., 2020; Geven et al., 2018; Timmermans et al., 2016). If teachers' perceptions are inaccurate, track recommendations may not predict students' future school success. If teachers are biased towards certain groups of students, differences in track recommendations between students arise.

Differences Between Groups of Students

While most differences in track recommendations between (groups of) students are due to differences in achievement levels (Luyten & Bosker, 2004; Timmermans, Kuyper, et al., 2015; van Rooijen et al., 2016), there is a small portion of these differences in track recommendations that cannot be explained by differences in achievement. This means that students with similar achievement but different backgrounds, such as gender, SES, and migration background, receive different track recommendations (Boone & Van Houtte, 2013; de Boer et al., 2010; Klapproth et al., 2013). This contradicts the meritocratic ideal that students with the same potential (ability) should be able to achieve the same educational outcomes regardless of their background (Themelis, 2008).

Research on teacher expectations can help to understand the role of student background characteristics in teachers' track recommendations. Teachers' expectations are teachers' inferences about students' potential behaviour or achievement (Good & Brophy, 1997). Teachers, both as individuals and together with colleagues, typically form expectations for their individual students, as well as for the class as a whole (Brophy, 1983). Researchers point out that biased teacher expectations in class are generally small (Brophy, 1983; de Boer et al., 2010; Jussim & Harber, 2005), but can be powerful among certain stigmatised subgroups (Klapproth et al., 2012; McKown & Weinstein, 2002; Sorhagen, 2013). Particularly in tracked educational systems where teacher judgements determine track recommendations, the effects of biased teacher expectations can have an important impact on students' track recommendation, and, in turn, their educational opportunities (Hopwood et al., 2016; Inspectie van het Onderwijs, 2018b; Pietsch & Stubbe, 2007).

Especially for SES, prior research has repeatedly found a bias in track recommendations: higher-SES students received higher track recommendations than lower-SES students with similar achievement levels (Batruch et al., 2023; Feron et al., 2016; Pit-ten Cate et al., 2016;

Timmermans et al., 2018). Bias in track recommendations was also found for gender and migration background, although these findings were not as strong and clear as the bias found in relation to SES, as research reported mixed results for the effects of gender and migration background on track recommendations. For gender, some studies found no evidence of bias in track recommendations (Boone & Van Houtte, 2013; Driessen, 2005; Krolak-Schwerdt et al., 2017), whereas some studies did in favour for girls (Feron et al., 2016; Jürges & Schneider, 2011; Timmermans et al., 2018). Also for migration background, some studies found no evidence of bias in track recommendations (Boone et al., 2018; de Boer et al., 2010; Driessen, 2012), while others found evidence that students with a migration background received higher track recommendations than students without a migration background in case of similar achievement (Caro, Lenkeit, et al., 2009; Luyten & Bosker, 2004; Timmermans et al., 2019), or conversely, that students with a migration background received lower track recommendations than students without a migration background in case of similar achievement (Klapproth et al., 2013; Korpershoek et al., 2016; Lüdemann & Schwerdt, 2013).

Differences Between Teachers

The extent to which teachers weigh students' achievement, student attributes, and student background characteristics in their track recommendations may vary between teachers.[3] This enables them to select, either consciously or subconsciously, which factors they consider in their track recommendations, how they weigh them and in what direction (i.e. positive or negative). Teachers' consideration of factors also depends on the guidelines of the allocation procedure and the type and amount of (standardised) student information available (Vanlommel & Schildkamp, 2019). In terms of achievement, for example, teachers may find it difficult to formulate track recommendations for students who perform inconsistently between different subject domains or have fluctuating achievement growth curves, as their achievement will not point directly to one specific secondary school track. As a result, some teachers may opt for more careful (i.e. lower) track recommendations (i.e. based on students' lowest achievement domain), whereas others opt for average (i.e. based on the mean of different achievement levels), or higher (i.e. based on students' highest achievement domain) track recommendations.

Furthermore, Timmermans et al. (2019) has shown that there are small differences between teachers in how they weigh different student attributes in their track recommendations. In other words, the weighting of some attributes in track recommendations

3 Chapters 2 to 4 did not distinguish between teacher and school level due to unavailability of teacher data. The descriptive statistics of the data also revealed that most primary schools had only one Grade 6 teacher. Therefore, differentiating between teachers and schools would not have added value as these effects would overlap.

differed from teacher to teacher. For instance, when teachers perceived their relationship with the student as more conflictual, some teachers weighed this positively while others weighed it negatively in their track recommendations. This resulted in higher and lower track recommendations, respectively. In this sense, it might seem that student-level factors have no effect on track recommendations, when in fact positive and negative effects cancel each other out.

There may also be differences between teachers in how they weigh student background characteristics in their track recommendations. Timmermans, de Wolf, et al. (2015), for example, found that characteristics of the school population affect teachers' track recommendations. That is, students in classes with few lower-SES students were more likely to be recommended to the highest secondary school track than students in classes with more lower-SES students. This could be due to a frame of reference effect: teachers form a frame of reference based on the student composition of their class (Boone et al., 2018; Geven et al., 2018; Timmermans, Kuyper, et al., 2015) and this, in turn, has an (unintended) impact on students' track recommendations, as individual students are compared to other students in the class. However, a study conducted in Flanders, Belgium by Boone et al. (2018) found no significant effect of SES class composition on school track recommendations. Boone et al. proposed that this may be attributed to the non-binding nature of the school track recommendations in Flanders, where track recommendations did not serve as formal entrance criteria for secondary education. Consequently, teachers may experience less parental pressure to formulate a track recommendation for the highest (academic) secondary school track (Dronkers et al., 1998). In line with this, Boone et al. (2018) argued that the impact of SES class composition would be more pronounced in educational systems with binding track recommendations, where secondary schools must follow this track recommendation formulated by teachers when students are allocated to a secondary school track, as is in the study by Timmermans, de Wolf, et al. (2015).

Most research on the impact of students' background on teacher judgements, expectations, and/or track recommendations focused on the effects situated at student level (e.g. Ready & Wright, 2011; Sorhagen, 2013), while disregarding the teacher or school level (but for an exception see Timmermans et al. (2016, 2019)). Yet, if there are significant differences between teachers or schools in the factors they include in their track recommendations, this may affect the predictive value of these factors.

Secondary School Success

After receiving a track recommendation, students are assigned to different secondary school tracks. In most tracked systems, students take all secondary school subjects at one level of education corresponding to the level of the track, leading to a qualification or diploma at that level (Korthals & Dronkers, 2016). Research indicated that after three or four years of secondary education, 60% to 75% of the Dutch students follow the same secondary school track as their track recommendation indicated (Driessen et al., 2005; Timmermans et al., 2013; van Rooijen et al., 2017). This suggests that track recommendations are not

always optimally predictive of students' future performance. The question that then arises is which factors are most predictive of students' secondary school success and thus should be used when formulating track recommendations.

The Role of Student Achievement

Students' achievement in primary school, which largely determines their track recommendations, also seems to largely determine their secondary school success. Research indicated a strong correlation between primary and secondary school achievement (de Boer et al., 2010; Poncelet & Metis Associates, 2004). For example, approximately 75% of the Dutch students showed more or less similar reading comprehension and mathematics achievement levels in Grade 6, final year of primary school, and Grade 9, third year of secondary school (van Rooijen et al., 2016). Likewise, it could also be that achievement growth in primary education is predictive of continued growth in achievement in secondary education. That is, a student who learns quickly and acquires new knowledge or skills quickly may have faster achievement growth in both primary and secondary education. This might suggest that it is important to take students' achievement growth into account when formulating track recommendations. At the other hand, as mentioned earlier, tracking could ensure that students' achievement growth is limited. Allocation to a particular secondary school track affects students' achievement (development) (Borghans et al., 2019; Dockx et al., 2019; Hanushek & Wössmann, 2006).

The Role of Student Attributes

Besides students' achievement, teachers may also consider other student attributes than students' background characteristics in their track recommendations. Examples of these attributes are students' work habits, classroom behaviour or even parental involvement (Inspectie van het Onderwijs, 2018d; Sneyers et al., 2018; Timmermans et al., 2019).

Research in the Netherlands compared the predictive value of different types of track recommendations for students' secondary school success (Dijks et al., 2020; Feron et al., 2016; Lek & van de Schoot, 2019). The results showed that judgement-based track recommendations were better predictors of students' secondary school success than test-based track recommendations. This may be due to teachers' consideration of such students' attributes in their track recommendation on top of student achievement. Prior research investigating the importance of student attributes from students' own perspective (or from their parents' perspective with regard to parental involvement) has shown that these attributes indeed predict students' future school performance (Boonk et al., 2018; Roorda et al., 2017). However, it is unknown whether this is also true for teachers' perceptions of these attributes. Due to the 'subjective' nature of these attributes and the fact that these student attributes are not stable over time (Praetorius et al., 2017; Vanlommel & Schildkamp, 2019; Zhu & Urhahne, 2020), teachers may not always accurately assess these attributes. Hence, teacher perceived student attributes in primary school may have limited predictive value for

students' performance in secondary school, as students may have developed, for example, different work habits or classroom behaviour in the meantime. This may particularly occur after the transition to secondary education, as students have entered a new educational environment with different requirements and different teachers.

On the Right Track?

Given the significant impact that track recommendations have on students' future educational careers and beyond (Boone & Demanet, 2020; Glock et al., 2015; Korpershoek et al., 2016), it is important that the transition procedure from primary to secondary education is as 'best' and 'fair' as possible. Therefore, it is important to investigate the factors considered by teachers when formulating track recommendations.

Unanswered Questions

Bias in track recommendations based on students' background is a cause for concern as students from different backgrounds do not have equal opportunities to receive track recommendations that match their abilities. Although this bias in track recommendations, especially for SES, has been repeatedly found (Batruch et al., 2023; Krolak-Schwerdt et al., 2017; Timmermans et al., 2019), a number of questions about this topic remain unanswered.

First of all, if standardised achievement tests would weigh more heavily in track recommendations, would this reduce the teacher bias of students' background characteristics? Because educational systems of different countries differ in many ways, it is difficult to make a valid comparison between allocation systems which differ in the extent to which standardised tests play a role. However, a policy reform of the Dutch allocation procedure in 2014 has provided us with the unique opportunity to examine the effects of the extent to which standardised tests are weighed in school track recommendations. That is, the Dutch educational policy regarding track recommendations shifted from relying mostly on students' school leavers' test results (i.e. test-based recommendations) to relying more strongly on teachers' judgements (i.e. judgement-based recommendations; Ministerie van Onderwijs Cultuur en Wetenschap, 2014). The weight of achievement and teacher bias in track recommendations is addressed in Chapter 2 of this dissertation.

Second, previous research examining the role of student achievement in track recommendations usually takes into account students' most recent test scores, and more specifically, often averages across domains or in the form of a school leavers' test score (de Boer et al., 2010; Glock et al., 2012; Luyten & Bosker, 2004). However, teachers may also look at other aspects of achievement, such as the degree of achievement inconsistency between multiple subject domains or achievement growth over several years. It can sometimes be difficult for teachers to properly consider these factors in their track recommendations because, unlike results on a school leavers' test, they do not clearly point to one secondary school track. Consequently, when teachers are in doubt about the most optimal track, they

might be somewhat cautious in their track recommendation to a student's disadvantage. The relationship between achievement inconsistency and achievement growth, on the one hand, and track recommendations, on the other, are addressed in Chapters 3 and 4 of this dissertation, respectively.

Furthermore, besides the unique contributions of prior achievement and student background characteristics, there may be a complex interplay between these factors that has not yet been examined in prior research on track recommendations. Although teachers seem to base their track recommendation primarily on students' achievement (Driessen, 2005; Feron et al., 2013; Luyten & Bosker, 2004), prior research showed that there may be achievement differences between different groups of students, which, in turn, can impact teachers' track recommendations (Timmermans, de Wolf, et al., 2015; Wang et al., 2018). The degree of achievement inconsistency and differences in achievement growth may contribute to the level of differences in track recommendations between groups of students. To illustrate, if teachers give higher track recommendations to students who have always performed high rather than to students who started with lower initial achievement such as lower-SES students (Caro, Lenkeit, et al., 2009; Caro, McDonald, et al., 2009; Helbling et al., 2019), but who showed more rapid growth over time, this may result in lower track recommendations for the students with lower initial achievement. Similarly, if teachers give lower track recommendations to students who perform inconsistently across subject domains because they base their track recommendation on students' weakest domain (i.e. with the lowest level of achievement), this may lead to lower track recommendations for groups of students with more inconsistent achievement. The interplay between student achievement and background characteristics is addressed in Chapters 2, 3 and 4 of this dissertation.

Third, similarly, there may be an interplay between student attributes and student background on track recommendations, besides the unique contributions, that has not yet been investigated in previous research. That is, teachers may potentially weigh certain student attributes more strongly for certain groups of students. For example, if parents of students from lower-SES families are perceived to be less involved (Bakker et al., 2007) and if their lower perceived parental involvement is considered more often in their track recommendations compared to students from higher-SES families (Sneyers et al., 2018; Timmermans et al., 2016), these individual effects could lead to a negative cumulative effect on track recommendations for lower-SES students. Thus far, it is unclear how these factors interact in predicting students' track recommendations. The interplay between perceived student attributes and background characteristics is addressed in Chapter 5 of this dissertation.

Fourth, to make matters even more complex, the extent to which achievement (including achievement inconsistency and achievement growth), student attributes, and student background characteristics are weighed may differ across teachers or schools. Most prior research (but for an exception see Timmermans et al. (2016, 2019)) has disregarded these differences. If there are substantial differences, this may lead to an underestimation

of the predictive value of these factors. To illustrate, some teachers may give students with less involved parents a lower track recommendation because they fear that the parents will not be able to help their child with their schoolwork if they need it. Other teachers, however, may give these students a higher track recommendation because they were able to reach the particular educational level without much help from their parents so far. In this example, where the effects of parental involvement on track recommendations vary from negative to positive across teachers, the overall net effect across teachers may not be significant. The difference between teachers in which factors they consider when formulating track recommendations is addressed in Chapters 2, 3, and 5 of this dissertation.

Research Questions

With the answers resulting from the questions introduced in the previous section, the aim was to address two overarching research questions that are introduced underneath. Prior research has mainly focused on the unique contribution of student achievement, student attributes, and student background characteristics on track recommendations, but this may not tell the full story. Questions regarding the interplay between student achievement, perceived student attributes, and student background in predicting track recommendations remain unanswered. A more comprehensive understanding of the role of background characteristics on track recommendations can contribute to the ongoing debate on how best to shape the allocations process based on track recommendations. Therefore, the first research question of this dissertation is as follows:

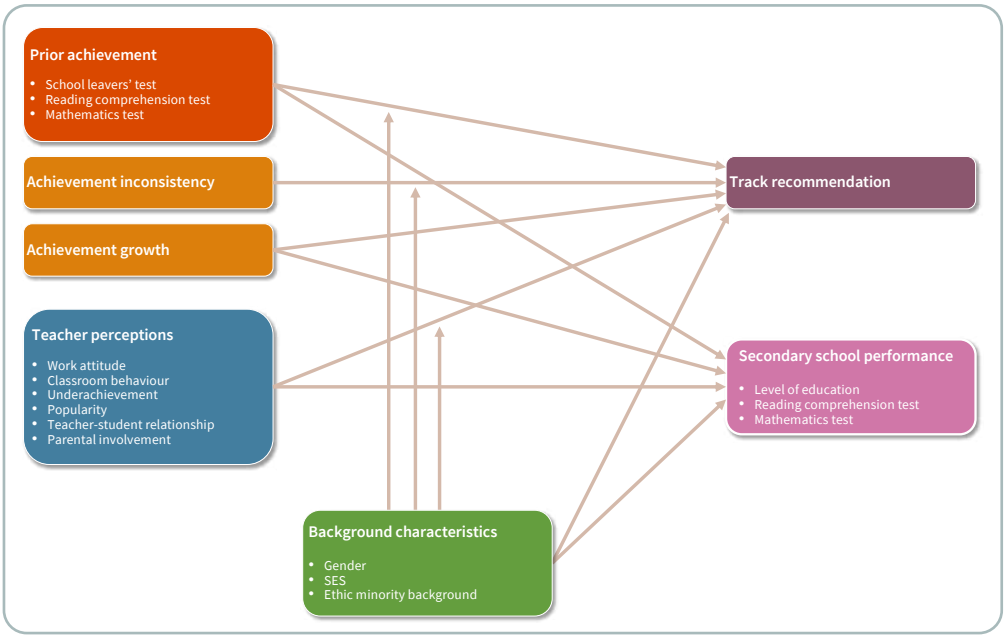
What aspects of student achievement, perceived student attributes, and student background characteristics do teachers consider when formulating track recommendations and how do these factors conjointly predict teachers' track recommendations?

Additionally, the studies in this dissertation also examine differences among teachers in their consideration of these factors when formulating track recommendations. Furthermore, since the goal of a track recommendation is to give an optimal reflection of students' potential educational performance in secondary education (Boone & Van Houtte, 2013; Glock et al., 2012; Klapproth et al., 2012), it is also important to know which factors are predictive of students' secondary school performance. Whether or not specific aspects should be taken into account in track recommendations may thus also depend on how predictive they are for students' future school success. While research indicates a strong correlation between primary and secondary school achievement (de Boer et al., 2010; Poncelet & Metis Associates, 2004; van Rooijen et al., 2016), it is not clear to what extent other factors are also predictive of secondary school success. Hence, the second research question of this dissertation is:

To what extent are aspects of student achievement in primary school, perceived student attributes and student background characteristics predictive of students' secondary school performance?

In Figure 1, an overview of the relations between the factors considered in this dissertation is presented.

Figure 1



Relations Between Factors Examined in This Dissertation

Research Method

The studies that are described in this dissertation use data from two longitudinal datasets containing data on the educational development of Dutch students in primary and secondary education. Chapters 2, 3, and 4 used a dataset which contains information on students in the city of Utrecht. This dataset was collected from two cohorts of students who received their track recommendation before the policy reform of 2014-2015, which changed the allocation decision process from a test-based to a judgement-based system, and two cohorts of students who received their track recommendation after the policy reform. Chapter 5 is based on data of the COOL$^{5\text{-}18}$ cohort study and provides information on a national level. This dataset contains data from before the policy reform of 2014-2015.

Data from two cohorts of students were included. Further details on the cohorts used in this dissertation can be found in Table 1.

In the present studies, the factors considered by teachers in track recommendations that are examined are students' achievement (i.e. the school leavers' test, most recent achievement in reading comprehension and mathematics, achievement inconsistency and achievement growth), teachers' perceptions of student attributes (i.e. work habits, classroom behaviour, underachievement, popularity, teacher-student relationship, and parental involvement), and student background characteristics (i.e. gender, SES and migration background).

Table 1

	Grade		National dataset (COOL$^{5-18}$) Cohort		Utrecht dataset Cohort			
	Dutch grade	International equivalent	1	2	1	2	3	4
Primary education	Groep 6	Grade 4	2005/2006	2008/2009	2010/2011	2011/2012	2012/2013	2013/2014
	Groep 7	Grade 5	2006/2007	2009/2010	2011/2012	2012/2013	2013/2014	2014/2015
	Groep 8	Grade 6	2007/2008	2010/2011	2012/2013	2013/2014	2014/2015	2015/2016
Secondary education	Klas 1	Grade 7	2008/2009	2011/2012	2013/2014	2014/2015	2015/2016	2016/2017
	Klas 2	Grade 8	2009/2010	2012/2013	2014/2015	2015/2016	2016/2017	2017/2018
	Klas 3	Grade 9	2010/2011	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019

Overview of the Different Cohorts of Both Datasets Used in this Dissertation

Overview of Chapters

In Table 2, a schematic overview of the chapters in this dissertation is presented. In the research project described in Chapter 2, it is investigated what the role of students' SES in track recommendations is on top of students' achievement. In addition, it is examined whether the bias in track recommendations is reduced when standardised achievement tests play a greater role in teachers' track recommendations. The policy reform of the Dutch educational system in 2014 has provided a unique opportunity to compare the effects of students' background characteristics in two situations: (1) before the policy reform when teachers had access to the results of the school leavers' test to base their track recommendations on (a test-based track recommendation), and (2) after the policy reform when teachers did not have access to the results of the school leavers' test, and had

to base their track recommendations on standardised tests from the student monitoring system throughout primary education and on their own expectations of students' future abilities (a judgement-based track recommendation). In addition, the direct and indirect effects of students' SES on both types of track recommendations were examined, as well as differences between teachers in how they weighed achievement and SES in their track recommendations.

The research described in Chapters 3 and 4 focuses on various aspects of student achievement that may or may not be considered in teachers track recommendations. These included the extent to which students show differences in achievement across subject domains (achievement inconsistency) and students' achievement growth over time. The focus in Chapter 3 is on the challenges teachers face when formulating track recommendations for students with varying achievements in reading comprehension and mathematics. The research in Chapter 4 examines if and how achievement growth in reading comprehension and mathematics throughout primary education was considered in teachers' track recommendations. In both chapters, it was examined whether students from different backgrounds (gender and SES) received different track recommendations based on, respectively, achievement inconsistency or growth. In Chapter 3, differences between teachers in how they considered achievement inconsistency is included, whereas Chapter 4 also focuses on the predictive value of students' achievement growth for secondary school performance.

After investigating factors related to students' achievement in primary education, with the research in Chapter 5 the focus shifts to the more 'subjective' aspect of track recommendations. With the research described in this chapter, the interaction effects of perceived student attributes, such as perceived work habits and parental involvement, and student background on track recommendations is explored. In addition to considering students' gender and SES, this study also takes into account students' migration background. Furthermore, differences between teachers in how they weighed the various student attributes in their track recommendations were investigated, as well as the predictive value of student attributes for secondary school performance.

Finally, in Chapter 6, a summary and general discussion of the main research findings of this dissertation is presented. This is followed by implications for practice and educational policy, a discussion of the limitations of the studies in this dissertation, and scientific implications and suggestions for future research.

"We voelen ons wel tekort gedaan *als er in het nieuws komt dat opleidingsniveau leidt tot lagere/hogere adviezen, omdat we naar zoveel meer kijken."*

Table 2

Chapter	Predictor variables	Outcome variables	Additional focus	Sample – *N*
2.	Students' SES + Prior achievement	Track recommendations	Before + after policy change Differences between teachers	Utrecht – 8,639
3.	Inconsistency of achievement	Track recommendations	Gender + SES Differences between teachers	Utrecht – 4,248
4.	Achievement growth	Track recommendations + Secondary school performance	Gender + SES	Utrecht – 4,738
5.	Perceived student attributes	Track recommendations + Secondary school performance	Gender + SES + migration background Differences between teachers	Nationwide COOL[5-18] – 17,953

Schematic Overview of the Topics Adressed in the Chapters of This Dissertation

When the Teacher is in Doubt: Some Illustrative Cases

In many cases, students' prior achievement in primary education will indicate which level of secondary education gives an optimal reflection of students' future abilities. However, this is not always very clear, for instance when students perform inconsistently between subject domains or when they have different growth rates but similar most recent achievement.

To illustrate, examples of students - related to the main topics of the chapters in this dissertation - of students for whom teachers might and might not find it difficult to formulate a track recommendation are presented. These examples are real students, and their data is derived from our own 'Utrecht' data. Their names were pseudonymised. For each study in this dissertation, two cases are selected that are illustrative of the factors considered in that chapter. These examples show that in some cases, although students have (more or less) similar characteristics, except for one, it is difficult for teachers to formulate similar track recommendations.

In the discussion, we return to these examples and describe which track recommendation these students received and discuss these cases in light of the overall findings of this dissertation.

Chapter 2: Socioeconomic Status

Two girls, Destiny and Joanne, have almost similar primary school achievement. Destiny's reading comprehension and mathematics test results from the student monitoring system LVS are higher than those of Joanne, but both have a similar score on the school leavers' test at the end of primary education, namely 547. Destiny is from a lower-SES background, Joanne from a higher-SES background. For both Destiny and Joanne, their achievement on the school leavers' test could suggest the pre-university track (*'vwo'*), but Joanne's reading comprehension and mathematics test results are lower. Therefore, Joanne's teacher may be in doubt whether to formulate a senior general (*'havo'*) or pre-university (*'vwo'*) track.

Chapter 3: Inconsistency in Achievement

Two boys, Paul and Mark, both from a middle-SES background, have similar primary school achievement. Their reading comprehension and mathematics test results at the end of primary education are similar. Both boys show substantial inconsistencies in their achievement with higher mathematics than reading comprehension achievement. They have not yet taken the school leavers' test yet, so these results are not known by their primary school teachers. Since the achievement of both Paul and Mark is rather inconsistently, it is difficult for their teachers which track to recommend.

Chapter 4: Achievement Growth

Two boys, Olivier and Thomas, have similar reading comprehension and mathematics achievement at the end of primary education. Both started primary education with similar reading comprehension achievement and thus had a similar growth curve for reading comprehension. However, for mathematics this was not the case. Olivier started primary education with high initial mathematics achievement and showed slower achievement growth during primary education, while Thomas started with low initial achievement and showed faster achievement growth. The boys have not yet taken the school leavers' test yet, so these results are not known by their primary school teachers. Furthermore, Thomas is from a middle-SES background, while Olivier is from a higher-SES background. Since the boys show different growth curves for mathematics, their teachers may be in doubt which track to recommend.

Chapter 5: Teachers' Perception of Student Attributes

Two girls, Emma and Kelly, both have similar reading comprehension and mathematics achievement at the end of primary education, and results on the school leavers' test, namely 539. Emma is from a higher-SES background, Kelly from a lower-SES background. According to the teacher's perceptions, Emma's parents are assumed to be highly involved in her schooling, while Kelly's parents aren't. For both Emma and Kelly, their achievement on the school leavers' test are known and suggest the senior general track (*'havo'*). However, since Kelly's parents are less involved in her schooling and cannot help her with future schoolwork, her teacher may be in doubt which track to recommend.

Appendix: The Dutch Educational System

The Structure of Secondary Education

The studies described in this dissertation were situated in the Netherlands. Therefore, an overview of the Dutch educational (tracking) system, focusing on secondary education, is provided here. The Dutch secondary school system is characterised by three main aspects that make optimal track recommendations important.

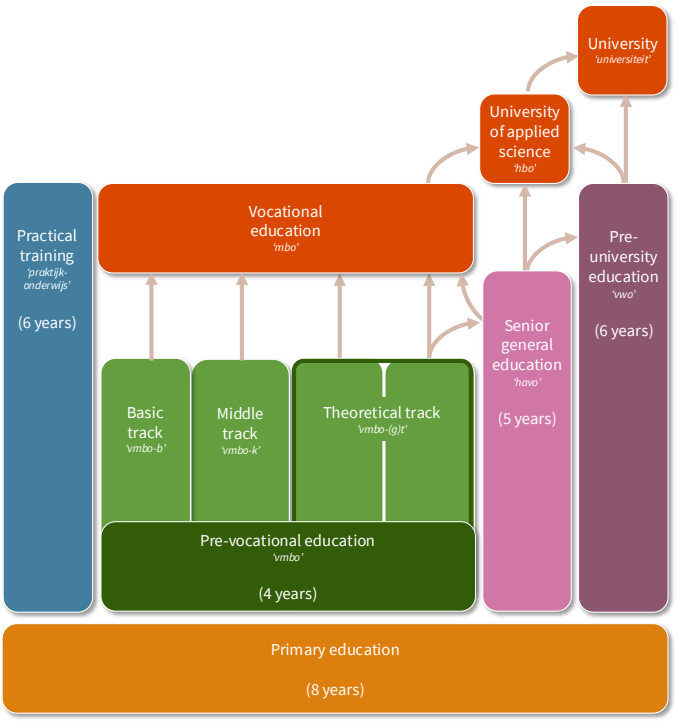
First, the Dutch educational system has relatively early tracking for secondary education. In the Netherlands, students are tracked into different ability levels at the age of twelve, which is relatively early compared to most other countries (Inspectie van het Onderwijs, 2007; OECD, 2020; Strello et al., 2021). By comparison, Austrian students make this transition even earlier at the age of 10 (Schütz et al., 2005), while Finnish students are allocated to different tracks around the age of 16 (Heiskala & Erola, 2019). A disadvantage of early tracking is that students may not have sufficient time to show their full potential prior to being separated into different tracks (van de Werfhorst & Heath, 2019). Early tracking also results in divergent achievement between students in lower and higher tracks. That is, research has shown that early tracking results in higher achievement among those in higher tracks, and to lower achievement for those in lower tracks (Horn, 2013; Miller, 2018; van Elk et al., 2009). This may be the reason why early tracking has been found to reinforce socioeconomic related inequalities in educational attainment (Bauer & Riphahn, 2006; Meghir & Palme, 2005; Pekkala Kerr et al., 2013). It needs to be mentioned here that an increasing number of Dutch secondary schools have a bridging period in place of one, two, or even three years (commonly known as 'brede brugklassen' in Dutch) aiming to mitigate the effects of early tracking. In this bridging period, students with adjacent track recommendations are not placed in specific tracks right away but are kept in a class with multiple tracks for this period (heterogeneous classes).

Second, the Dutch secondary school system is considered to be rather rigid, because path dependency plays a major role in students' educational careers. This means that the track to which students are allocated largely determines their future (school) careers (Glock et al., 2012; Korpershoek et al., 2016; Timmermans, Kuyper, et al., 2015). Not only because students may perform according to the track they are in (i.e. self-fulfilling prophecy effects; Brophy, 1983; De Boer et al., 2010; Jussim & Harber, 2005), but also because in general mobility between the hierarchical tracks is rather limited, especially switching upwards (Feron et al., 2016). This is due to factors, such as the types of tracks offered by secondary schools, as not all schools offer all tracks (Brunello & Checchi, 2007; Gamoran, 2009), or strict requirements for upward switching (de Winter-Koçak & Reches, 2022). Often students also need an extra school year to transfer to a higher track, especially in upper secondary education and in schools with only homogeneous tracks in lower secondary education (Exalto et al., 2018; Visser et al., 2022). Because of the reasons mentioned above, the more rigid an educational system is, the worse it is in terms of equal educational opportunities, especially for lower-SES students. Lower-SES students are more likely to switch to lower

tracks (and are less likely to switch to higher tracks), repeat a grade or drop out of school, and are less likely to make the transition to tertiary education and have a lower completed level of secondary education (Feskens et al., 2016; Inspectie van het Onderwijs, 2017b; Scheerens et al., 2019).

Third, the Dutch educational system consists of many secondary school tracks (i.e. horizontal stratification). Students are allocated to six different secondary school tracks, ranging from practical training via different types of pre-vocational education to pre-university education (Naayer et al., 2016; OECD, 2020; Strello et al., 2021), see Figure 2. In contrast, in Greece there are only two secondary school tracks: vocational education and general education. Research has shown that having more secondary school tracks, for example, results in lower achievement in reading comprehension (OECD, 2020).

Figure 2



Note. This figure presents the six different hierarchical tracks in Dutch secondary education, including the duration of each track and the possible intra-secondary transitions, and the possible pathways to tertiary education. The scale from 1 (lowest track: practical training) to 6 (highest track: pre-university education) corresponds to the six secondary school tracks.

Overview of the Dutch Educational System

History of the Dutch Educational System

Until 1960

Although these three main aspects of the Dutch secondary system did not always exist, the system was already hierarchically tracked for a longer time. Until the 1960s, students commonly followed their parents' career choices, with boys often following their fathers into professions, such as shoemaker or teacher, while girls typically followed their mothers into the role of housewife. Moreover, many students only attended primary education and went working rather than attending secondary education (van de Ven, 2007). Students who did enrol into secondary education, mostly students from affluent families, could choose the secondary school track themselves, although in practice this was often decided by their parents. Nevertheless, they also did receive a track recommendation from their primary school teacher, sometimes supplemented by an ability test, and/or a compulsory admission test (Faasse et al., 1987). At that time, the different types of secondary education[4] were more disconnected from each other than they are today. It meant that schools usually offered only one type of education, and that the choice of a particular secondary school track was a definite one (Faasse et al., 1987; Luijkx & de Heus, 2008). This was also the case for tertiary education, which also consisted of strictly separated tracks based on secondary school qualifications (UNESCO, 1985). Switching downwards or upwards between different levels of secondary or tertiary education was therefore not common or even possible (Luijkx & de Heus, 2008). Since mostly students from affluent families had the opportunity to continue their education, family background seemed to predetermine an individual's future (educational) career.

Post-Second World War: The Rise of the Meritocracy

The educational situation in the Netherlands, which was gradually developing as a welfare state, changed after the Second World War. An increasing number of students attended secondary school, and a greater emphasis was attached to diplomas and certificates. Diplomas and certificates were seen as a gateway to many (new) interesting job opportunities as modernisation of society demanded more highly skilled workers (van de Ven, 2007; van der Ploeg, 1993). At the same time, not only in the Netherlands, but also in other countries, a debate about the equal educational opportunities of students emerged. To create equal educational opportunities for students from different backgrounds, reforms that improved access to education were implemented in many countries, including the Netherlands. Meritocracy had a prominent role in these reforms: students with the same

4 The types of secondary education, ranked from low to high, were continued primary school, domestic science or technical school, advanced primary school, and classical and modern grammar school (Faasse et al., 1987; Luijkx & de Heus, 2008).

potential (their 'merit') should be able to achieve the same educational outcomes, such as educational qualifications or diplomas, and subsequent labour market positions, regardless of their background (Themelis, 2008).

1968: The Mammoth Act

In 1968, the post-primary educational system in the Netherlands underwent a reform, known as the Mammoth Act (Stiggins & Stiggins, 2002), which was the basis of the current tracked educational system. The different secondary school tracks that existed were streamlined into six educational levels, each designed to prepare students for different hierarchical levels of tertiary education, as can be seen in Figure 2 (Luijkx & de Heus, 2008; van de Ven, 2007; van Lutsenburg Maas, 1964). This reform resulted in higher social mobility, as it became possible to switch between secondary and tertiary school tracks (i.e. intra-secondary and -tertiary transitions) to some extent. After completion of a lower level of education, students had the opportunity to continue their education at a higher educational level. For example, after completing pre-vocational education students could continue their education either in the fourth grade of senior general education (secondary education), or enter vocational education (tertiary education), and completion of either of these options opened the possibility to continue to go to college (van de Ven, 2007; van der Ploeg, 1993).

The Mammoth Act also brought changes to the allocation procedure for the transition from primary to secondary education. To ensure that students with similar levels of achievement were able to achieve similar educational success, compulsory testing was implemented. This resulted in the establishment of the Central Institute for Test Development (Cito[5]), a national organisation responsible for developing standardised achievement tests (Cito, n.d.-a; Faasse et al., 1987; Luijkx & de Heus, 2008). In 1968, Cito introduced the school leavers' test ('de Citotoets') in the Netherlands, which was derived from the Amsterdam School Test developed a few years earlier by Adriaan de Groot. This school leavers' test was intended to provide teachers with an objective measure of students' abilities to inform their track recommendations, and thereby to limit the role of students' background characteristics. Likewise, (high-stakes) standardised tests were introduced in schools in the Netherlands and other countries during the late 20th century (Au, 2013; Faasse et al., 1987; Luijkx & de Heus, 2008).

5 Cito is the Central Institute for Test Development, established in 1967 (Cito, n.d.-a; Faasse et al., 1987; Luijkx & de Heus, 2008). Cito is a national organisation responsible for developing standardised achievement tests, such as a school leavers' test and the student monitoring system which includes standardised tests in, for example, reading comprehension, mathematics, and English for almost all grades throughout primary and secondary school (Cito, n.d.-c, n.d.-b).

Late 20th Century: Introduction of a Test-Based System

Although the school leavers' test was initially widely criticised as being an inappropriate measure of students' abilities, it became increasingly popular in the late 20th century, and was used by most primary schools across the country (Cito, n.d.-a). Based on the school leavers' test, primary school teachers could formulate a track recommendation indicating the most optimal track for a student. Although, in theory, teachers could consider other information in their recommendation, in practice, the track recommendations became almost entirely based on the school leavers' test results (Luyten & Bosker, 2004). Moreover, secondary schools also used the test scores as admission requirements, with a given range of test scores corresponding to specific tracks. As secondary schools used the school leavers' test to allocate students to a secondary school track, the school leavers' test functioned as a high-stakes test, thus bypassing teachers' track recommendations (Driessen, 2006; van Nuland, 2011). Given this strong focus on the school leavers' test, this system could be considered as a test-based system.

2015: Policy Reform Towards a Judgement-Based System

The strong focus on the school leavers' test triggered a renewed discussion about providing equal educational opportunities. Opponents of this way of allocating students to a secondary school track argued that too much weight was given to the school leavers' test, as it was only administered at one moment in time. If a student performs lower on this test, for example due to personal circumstances, performance pressure, or language barriers, the student will receive a lower track recommendation or is placed in a lower track. Opponents also argued that teachers had a better and accurate view of students' abilities, since teachers have deep knowledge of their students due to their daily interactions in class (de Boer et al., 2010; 'Plenair Debat 56-4: Omzeilen van Het Schooladvies', 2015).

As a result, in the 2014-2015 academic year, the allocation process was reformed to a judgement-based system. This policy change meant that teachers' track recommendations played a more significant role in students' track placement in secondary education (Ministerie van Onderwijs Cultuur en Wetenschap, 2014). Although the nationwide school leavers' test became mandatory, teachers no longer had access to its results when formulating track recommendations due to a changed time schedule (Korpershoek et al., 2016; Oomens et al., 2019). Instead, teachers had access to other (standardised) test results and also relied on their own expectations of students' future achievement and development when formulating track recommendations (Oomens et al., 2019). Additionally, the judgement-based track recommendation became binding: secondary schools were required to allocate students to the educational level of the track recommended by the teacher rather than the test score (Dutch Ministry of Education, 2014).

However, teachers were allowed to adjust their initial track recommendation if the school leavers' test results indicated a higher level of secondary education than the initial track recommendation (Korpershoek et al., 2016; Oomens et al., 2017). The school leavers' test results thereby served as a correction function to avoid allocating students to a lower

track based on biased track recommendations due to students' background characteristics. Previous studies indicated that especially during the early years of this policy reform, it was very uncommon for track recommendations to be adjusted for students who were eligible for upward corrections (Hebbink et al., 2022; Oomens et al., 2019; van Look et al., 2018). In more recent years, the track recommendations were adjusted for about 25% of the students who were eligible for upward corrections (Hartgers et al., 2021).

2024: Introduction of the Transition Test

While the changed transition procedures in 2014 were intended for the school leavers' test to serve as a second opinion or correction function, in practice this did not happen. As a result, a new policy reform was just implemented in the 2024-2025 academic year. The school leavers' test has now been replaced by the so-called transition test ('doorstroomtoets' in Dutch). The new name signifies that the test is only a marker of students' ongoing development from primary to secondary education rather than a marker of the end of their development in primary school (*Kamerstukken II, 35671, Nr. 13, 2021*). The biggest change compared to the previous policies is that teachers' track recommendations need to be adjusted if the transition test indicates a higher secondary school level than the track recommendation, except in cases of compelling reasons not to do so (Ministerie van Onderwijs Cultuur en Wetenschap, 2023). These new regulations are intended to strengthen the correction function of the test. In addition, the timing of the test was changed. Previously, many students who received upward corrections were unable to enrol in a secondary school that offered the level of education of the adjusted track because the adjustments were made after the students were already registered for secondary education (*Kamerstukken II, 35671, Nr. 13, 2021*). Due to the new timing, students enrol in secondary education after the potential adjustments have been made. In all, these policy changes need to promote students' equal educational opportunities.

"Er komt eigenlijk zoveel bij kijken voordat je zegt: dit is het advies dat bij dit kind past."

"De harde gegevens gebruiken we om te bekijken of een kind in potentie iets zou kunnen, maar een schooladvies gaat ook altijd samen met zicht op de 'zachte' gegevens."

2 Test- or Judgement-Based

School Track Recommendations: Equal Opportunities for Students With Different Socioeconomic Backgrounds?

Abstract

There are concerns that school track recommendations that are mostly based on teachers' judgements of students' achievement ('judgement-based recommendations') are more biased by students' SES than school track recommendations that are mostly based on standardised test results ('test-based recommendations'). A recent policy reform of the Dutch educational system has provided us the unique opportunity to compare the effects of students' SES on these two types of track recommendations. Therefore, the aim of this study was to examine the differences between test-based and judgement-based recommendations regarding the direct and indirect effect of students' SES at student level and school level. The sample consisted of 8,639 Grade 6 students from 105 Dutch primary schools. Data were analysed using multilevel mediation models. Results showed that track recommendations were higher for higher-SES students. This was mostly due to differences in students' prior achievement. SES also had a small direct effect on judgement-based but not on test-based recommendations. The effects were partly situated at school level. Overall, these findings indicated that teachers were capable of appropriately formulating school track recommendations relying mostly on students' prior achievement without being biased by students' SES.

Keywords

equal educational opportunities, transition from primary to secondary education, school track recommendation, SES, prior achievement, reading comprehension, mathematics, school leavers' test, teacher judgement

This chapter is based on: van Leest, A., Hornstra, L., van Tartwijk, J., & van de Pol, J. (2020). Test- or judgement-based school track recommendations: Equal opportunities for students with different socio-economic backgrounds? *British Journal of Educational Psychology, 91*(1), 193-216. https://doi.org/10.1111/bjep.12356

Acknowledgement of author contributions: All authors contributed to the design of the study. AvL collected the data and prepared the data for analyses; AvL analysed the data with help from LH; AvL drafted the manuscript; All authors contributed to critical revision of the manuscript; LH, JvdP, and JvT supervised the study.

Introduction

In many countries, such as Great-Britain, Germany, Luxembourg, Singapore, France, and the Netherlands, the transition from primary to secondary education involves students' allocation to specific school tracks. These tracks differ in the educational qualifications students can acquire and determine students' educational career (Contini & Scagni, 2011; Glock et al., 2012; Korpershoek et al., 2016). A substantial body of research (e.g. Driessen et al., 2008; Dutch Inspectorate of Education, 2018c; Klapproth et al., 2012; OECD, 2016; Pietsch & Stubbe, 2007) has reported that students' background characteristics, including their families' socioeconomic status (SES), have an impact on students' allocation to specific secondary school tracks. To create equal opportunities for all students, it is important that students are allocated to the secondary school track that is most appropriate based on their abilities, regardless of their SES (Tieben & Wolbers, 2010). In most tracked educational systems, students' allocation is typically based on the 'school track recommendation' students receive at the end of primary education (Contini & Scagni, 2011; Glock et al., 2012; LeTendre et al., 2003). The degree to which school track recommendations are based on standardised test results or on teachers' judgements differ per educational system. Because it is difficult to make a valid comparison between these distinct types of track recommendations, given that within one country, one type of recommendation is used (LeTendre et al., 2003), little research has addressed this topic. A recent policy reform of the Dutch educational system enables us to compare two types of track recommendations: recommendations that are primarily based on results of a school leavers' test (test-based recommendations) and recommendations that are primarily based on teacher judgements (judgement-based recommendations). In the present study, the aim was to investigate for both types of recommendations (1) whether students' SES has an effect on track recommendations, and (2) whether these effects are mediated by students' prior achievement.

Test-Based Versus Judgement-Based Track Recommendations

School track recommendations are formulated at the end of primary education, but how these track recommendations are developed differs per educational system (LeTendre et al., 2003). In some countries, such as Great-Britain and Singapore, track recommendations are based on students' results on a standardised, multisubject school leavers' test (Boone & Van Houtte, 2013; Driessen, 2011b; Le Métais, 2003). In other countries, such as Belgium, Germany, Luxembourg, and France, track recommendations are based on teachers' expectations about the most optimal level of secondary education for students to develop and perform successfully (Boone & Van Houtte, 2013; de Boer et al., 2010; Glock et al., 2012; Klapproth et al., 2012; Le Métais, 2003; Timmermans, Kuyper, et al., 2015). These expectations are teachers' inferences about students' potential achievement, usually based on teachers' judgements of students' current achievement as well as other characteristics, such as their motivation or behaviour (Boone & Van Houtte, 2013; Feron et al., 2016; Klapproth et al., 2012; Riley & Ungerleider, 2012). In short, track recommendations can be mostly based

on test results, hereafter referred to as 'test-based recommendations', or on teacher judgements, hereafter referred to as 'judgement-based recommendations' (Boone & Van Houtte, 2013; Glock et al., 2012; Klapproth et al., 2012). Since teachers' judgements are partly based on (standardised) test results as well, a strict distinction between test-based and judgement-based recommendations cannot be made. The degree to which track recommendations are based on standardised tests or on teachers' judgements should rather be seen as a continuum. In this continuum, the two types of recommendations primarily differ in the extent to which teachers may consider students' (background) characteristics, such as SES, motivation, or behaviour.

A recent policy reform in the Netherlands provides the opportunity to compare the two types of track recommendations (Ministerie van Onderwijs Cultuur en Wetenschap, 2014). Before the policy reform, test-based recommendations were used. These recommendations were mostly based on students' standardised school leavers' test score (Luyten & Bosker, 2004). This test result directly indicated a level of secondary education that was most appropriate for a student. Approximately 75% to 80% of the students received a track recommendation that matched their result on the school leavers' test, while the remaining 20% to 25% of the students received a track recommendation that did not match their result on the school leavers' test (Oomens et al., 2019). After the policy reform, judgement-based recommendations were used. The school leavers' test scores were no longer available to teachers when formulating track recommendations (Korpershoek et al., 2016; Oomens et al., 2019). Instead, these recommendations were based on teachers' expectations of students' future achievement and development during secondary education (Oomens et al., 2019). Even though both types of track recommendations are not situated at the far end of the continuum, they substantially differ in the extent to which standardised test results and other student characteristics can or will be considered.

There are valid arguments in favour of both types of recommendations. A central argument in favour of using test-based recommendations is that using a school leavers' test as primary indicator improves educational equality. A school leavers' test consists of the same set of questions and is ideally administered under similar conditions (Knoester & Au, 2017). Additionally, for all students, the standard is set at the same level to ensure that they are judged solely on their achievement, neglecting other student characteristics. Consequently, test-based recommendations should be similar for all students with comparable achievement levels, regardless of their background characteristics. In turn, students from different backgrounds will have equal opportunities to be assigned a certain track recommendation (OECD, 2016b). However, a school leavers' test is administered at one specific moment, which makes it impossible to take students' (cognitive) development throughout primary school into account (Driessen, 2005; OECD, 2016b). When a student does not perform as well as he or she normally does, the results of the school leavers' test are not in line with the student's actual abilities (Boone & Van Houtte, 2013).

A central argument in favour of using judgement-based recommendations is that teachers have the opportunity, in addition to standardised test results, to include (non-

cognitive) information that may be predictive of students' future secondary school success, such as work habits, classroom behaviour, motivation, talents, development and school engagement (Boone & Van Houtte, 2013; Driessen, 2011b; Driessen et al., 2008; Feron et al., 2016; Jungbluth, 2003; Klapproth et al., 2012). Consequently, in educational systems that use judgement-based recommendations, students who perform at the same educational level may receive different track recommendations. Moreover, teacher judgements are susceptible to bias, which may cause an undesired effect: lower track recommendations for students with more disadvantaged backgrounds (Boone & Van Houtte, 2013; Driessen, 2005, 2011b; Driessen et al., 2008; OECD, 2016b).

Impact of SES

On average, track recommendations are more positive for higher-SES students than for lower-SES students (Inspectie van het Onderwijs, 2018a). This can primarily be explained by students' prior achievement (Caro, Lenkeit, et al., 2009; Sirin, 2005): because lower-SES students generally perform lower than higher-SES students, their track recommendations are lower (Luyten & Bosker, 2004; Timmermans, Kuyper, et al., 2015). These findings point to an indirect effect of SES on track recommendations through students' achievement levels. For example, Caro and colleagues (2009) demonstrated that families' SES was indirectly related to students' track recommendation through their mathematics achievement and growth. Higher-SES students performed better in mathematics than lower-SES students and, consequently, received higher track recommendations. In addition, in a meta-analysis of 74 studies, Sirin (2005) concluded that lower-SES students obtained lower scores on standardised literacy tests than higher-SES students. Explanations for these indirect effects are related to students' social, cultural and financial capital or resources, which are predictive of school achievement (Milne & Plourde, 2006; Sirin, 2005). For example, Bradley and Corwyn (2002) reported that lower-SES children were less likely to visit museums, libraries, theatrical events or educational institutions. Moreover, research of Constantino (2005) indicated that higher-SES children had access to more books than lower-SES children and, consequently, were more likely to read regularly. Because of these indirect effects, we can draw the conclusion that both test-based and judgement-based recommendations may be lower for lower-SES students than for higher-SES students.

However, previous studies (e.g. Driessen, 2011; Dutch Inspectorate of Education, 2018a; Luyten & Bosker, 2004; OECD, 2016) that included students' prior achievement as well as their SES identified that students' prior achievement could not entirely explain the effect of SES on track recommendations. This finding indicates that there is also a direct effect of SES, which cannot be accounted for by students' prior achievement. This direct effect may be caused by track recommendation bias (Driessen et al., 2008; Jungbluth, 2003). Research (e.g. Driessen et al., 2008; Dutch Inspectorate of Education, 2018a; OECD, 2016) emphasised that students' SES was a stronger source of track recommendation bias compared to other student background characteristics, such as gender or migration background. There are multiple considerations of teachers, both explicit and implicit, that can explain this

finding. An example of an explicit consideration is when teachers consider the parents of a lower-SES student to be less able to provide (educational) support (e.g. Bakker et al., 2007; Hoover-Dempsey et al., 1992). Consequently, teachers may be more careful when formulating a track recommendation. Implicit considerations may occur if stereotypes are activated when teachers perceive a student as belonging to a particular subgroup (Fiske & Neuberg, 1990; Peterson et al., 2016). Consequently, information about that student will be interpreted in terms of the activated stereotype and, in turn, used for forming judgements (Krolak-Schwerdt et al., 2013). For example, students' SES can activate such a stereotype and can cause teachers to perceive lower-SES students as less capable than higher-SES students, which could subsequently lead to lower track recommendations. In addition, research (Glock & Krolak-Schwerdt, 2013) demonstrated that the stronger the stereotype was, the stronger the impact on judgements was. As teachers' recommendations were overall quite accurate (De Boer et al., 2010), the effects of biased track recommendations were, on average, small (Klapproth et al., 2012; Pietsch & Stubbe, 2007). However, the effects may increase among certain stigmatised subgroups, such as lower-SES students.

In addition, because judgement-based recommendations rely more strongly on teachers' own judgements of their students compared to test-based recommendations, non-cognitive student characteristics may have a substantial impact. Indeed, prior research (Bol et al., 2014; Caro, Lenkeit, et al., 2009) indicated that judgement-based recommendations tended to be more biased by students' SES than test-based recommendations. More specifically, Luyten and Bosker (2004) reported that the influence of students' background characteristics on track recommendations was stronger at schools that did not administer a school leavers' test compared to schools that did administer a test. Hence, because teachers may (unintendedly) formulate biased track recommendations (Driessen et al., 2008; Jungbluth, 2003), the Dutch policy reform shifting from test-based to judgement-based recommendations can result in stronger effects of students' SES on track recommendations. To further examine this assumption, students' prior achievement is included in the present study.

Student and School Level

Most research on the impact of SES on teacher judgements, expectations and/or track recommendations focused on the effects situated at student level (e.g. Ready & Wright, 2011; Sorhagen, 2013). Because track recommendations are generally made during staff or decision council meetings within a school (Barg, 2013; Bonizzoni et al., 2016; Boone & Van Houtte, 2013; Dollmann, 2016; Driessen, 2011b; Klapproth et al., 2012, 2013; PO-raad & VO-raad, 2014), the effects of SES on school track recommendations may also be situated at school level. In other words, perhaps the differences between track recommendations for lower- and higher-SES students are not situated *within* schools, but *between* schools. Timmermans, Kuyper, et al.(2015) reported that students in higher-SES classes more frequently received a school track recommendation for the highest secondary school track than students in lower-SES classes. This can occur since students are grouped in

classes and teachers experience these classes daily. It is likely that teachers form a frame of reference based on the composition of the class (Boone et al., 2018; Klapproth & Fischer, 2019), perceiving students in higher-SES classes as possessing more advanced academic skills than students in lower-SES classes (Ready & Wright, 2011).

However, a study in Flanders, Belgium (Boone et al., 2018) reported no significant effect of SES class composition on school track recommendations. They suggested that this might be due to the fact that school track recommendations were non-binding in Flanders, i.e. track recommendations did not serve as formal entrance criteria for secondary education. Consequently, teachers may experience less parental pressure to get a track recommendation for the highest (academic) secondary school track (Dronkers et al., 1998). In turn, Boone and colleagues (2018) speculated that the impact of SES class composition would be more pronounced in educational systems with binding track recommendations, as is in the study of Timmermans, Kuyper, et al.(2015). In general, previous research presented mixed evidence regarding the effect of SES on track recommendations at class composition level.

Present Study

Because of the ongoing (political) debate addressing educational equality at the transition from primary to secondary education, it is of significant importance to examine what type of track recommendation leads to the most equal educational opportunities for students (OECD, 2016b). Previous research (Bol et al., 2014) compared countries with different educational systems to investigate the impact of SES on test-based and judgement-based track recommendations. However, educational systems tend to vary from one another in many aspects, which makes it difficult to ensure a valid comparison and to determine the exact impact of SES on different types of track recommendations. A recent policy reform of the Dutch educational system has provided us the unique opportunity to compare the effects of students' SES on two types of track recommendations. The policy shifted from relying primarily on students' school leavers' test results (test-based recommendations) to relying more strongly on teacher judgements (judgement-based recommendations) (Ministerie van Onderwijs Cultuur en Wetenschap, 2014). Both types of recommendations substantially differ in the degree to which teachers consider standardised test results and other student characteristics. To summarize, in the present study, it was investigate for both types of track recommendations (1) whether students' SES has an effect on track recommendations at student and school level, and (2) whether these effects are mediated by students' prior achievement in reading comprehension and mathematics, and in case of test-based recommendations also by students' results on the school leavers' test.

Based on previous research (Inspectie van het Onderwijs, 2018a; Luyten & Bosker, 2004; Timmermans, Kuyper, et al., 2015), it was expected that higher-SES students would have higher track recommendations than lower-SES students. Second, because students from different SES backgrounds differ in achievement (Caro, Lenkeit, et al., 2009; Sirin, 2005), it was expected that for both types of track recommendations the effect of SES on

track recommendations would primarily be explained by differences in prior achievement (i.e. an indirect effect) between students with different SES. However, because test-based recommendations might be less biased by teacher judgements (Bol et al., 2014; Caro, Lenkeit, et al., 2009; Luyten & Bosker, 2004), a weaker direct effect of SES on test-based recommendations was expected than on judgement-based recommendations. Finally, no specific expectation was formulated at which level the effects of SES on track recommendations will be situated, given the conflicting findings of prior research (Boone et al., 2018; Timmermans, Kuyper, et al., 2015).

Method

Sample

The data used in this research were part of a larger dataset on the educational development of students in primary and secondary education. From the full dataset, the dataset of the present study was constructed including various variables of students' background and educational achievement in primary school. Schools that gave their approval of using their data were included. These data were retrieved from an online student monitoring platform that primary schools used to upload student information. An institution representing the schools, which had access to this online monitoring platform, downloaded and anonymised the data.

The dataset consisted of a sample of 8,639 Grade 6 students from 105 primary schools of a large city in the Netherlands. Students were from four cohorts: two cohorts before the educational policy reform (having test-based recommendations) and two cohorts after the reform (having judgement-based recommendations), which allowed us to examine differences between both types of recommendations. The cohorts with test-based recommendations consisted of 4,391 (50.8% of the total sample) students who were in Grade 6 of primary education in the academic year 2012-2013 or 2013-2014, and the cohorts with judgement-based recommendations consisted of 4,248 (49.2% of the total sample) students who were in Grade 6 in academic year 2014-2015 or 2015-2016.

Measures

The main focus of the present study was the variable track recommendation. Additionally, students' educational achievement was included, consisting of two main variables: (a) a standardised mathematics test, and (b) a standardised reading comprehension test. For test-based recommendations, a third achievement variable was included: (c) the school leavers' test. School leavers' test results were solely available for test-based recommendations. Finally, students' SES was included to examine whether track recommendations were biased by SES. The descriptive statistics of these variables are presented in Table 1.

Track Recommendation. The Dutch secondary educational system is a tracked system. At the beginning of March in Grade 6, i.e. the final year of primary education, teachers formulate a track recommendation that indicates which of the six secondary school tracks is most appropriate for a specific student. The six tracks are: (1) practical training, (2) basic pre-vocational secondary education, (3) middle pre-vocational secondary education, (4) theoretical pre-vocational education, (5) senior general secondary education, and (6) pre-university education. Because the variable track recommendation consisted of six categories, it was considered as a continuous variable. According to several researchers (Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993) this can be done without any harm to the analyses.

In the present study, the initial track recommendation, i.e. the recommendation without any corrections made at a later stage, was used to make a valid comparison between the recommendations formulated before and after the policy reform. Because a singular recommendation is mandatory and part of the official regulations in the city our data originates from, combined track recommendations of adjacent tracks are not part of this initial track recommendation. Finally, this track recommendation does not necessarily correspond to the actual track placement of a student. Since the actual track placement is influenced by other factors such as regulations that secondary schools apply when allocating students to a specific school track, it is beyond the scope of this research.

Prior Achievement. Students' prior achievement was measured using students' reading comprehension and mathematics scores on standardised tests. These scores were retrieved from schools' monitoring and evaluation system. The monitoring and evaluation system is developed by Cito i.e. the Dutch National Institute for Educational Measurement, and consists of several standardised tests throughout primary school to monitor students' progress in different subject domains. For the present study, students' most recent scores on both subject domains were selected, because these generally provided the most predictive value for track recommendations (PO-raad & VO-raad, 2014). The scores on the tests are converted by Cito into a single test score for each subject domain, resulting in reading comprehension test scores ranging from -87 to 147, and in mathematics test scores ranging from 0 to 168 (Cito, 2016). Prior research (Feenstra et al., 2010; Janssen et al., 2010) indicated that the tests had a high validity and high internal consistency ($\alpha > .80$). To account for potential differences between test versions, the test scores were standardised.

School Leavers' Test. While reading comprehension and mathematics were included as measures for students' prior achievement for both types of recommendations, the school leavers' test is of significant importance as achievement indicator for test-based recommendations. Because of a changed time schedule, the results of this test are not available for judgement-based recommendations. The school leavers' test is a multisubject high-stakes test administered in Grade 6 of primary education. The test is administered at the same time in the whole country. Although there are multiple types of school leavers'

tests available nowadays, the vast majority of schools administer the school leavers' test developed by Cito (College voor Toetsen en Examens, 2015; van Look et al., 2018). Therefore, the Cito school leavers' test, also known as the End of Primary Education Test, was included in the present study. It has been designed as a standardised measurement of students' achievement level to indicate an appropriate secondary school track type. The test consists of 290 multiple-choice items, divided over four subtests: (a) Dutch language (reading comprehension, writing, decoding, spelling/grammar), (b) mathematics (arithmetic, geometry, algebra), (c) study skills (map-reading, interpreting study texts, information sources, graphs, diagrams and tables), and (d) world studies (geography, history, science). The results on these different subject domains are converted by Cito into a single test score ranging from 501 to 550 (van Boxtel et al., 2011). The Cito school leavers' test is calibrated each year to guarantee that students' average test scores are comparable across years. Prior research (College voor Toetsen en Examens, 2015; van Boxtel et al., 2011) reported that these tests had a high validity and high internal consistency ($\alpha > .95$).

Socioeconomic Status (SES). Students' six-digit postal code was used as an approximation of students' families' SES, as it can be an useful marker of SES (Danesh et al., 1999). Indicators of SES to recode postal codes were provided by Statistics Netherlands (CBS), an autonomous organisation that offers statistical data to gain insight into social issues (van Leeuwen, 2019). The SES variable was composed of three indicators: (a) the most recent mean household income after tax, (b) the mean real estate value, and (c) the number of people with unemployment or social welfare benefits, all measured in 2016. The three indicators were recoded into one factor score using principal component analysis (PCA). Because six-digit postal codes referred to smaller areas, the classification of six-digit postal codes was more precise than classification of five-digit or four-digit postal codes (van der Aa et al., 2011). For Dutch cities, the six-digit postal codes were, on average, shared by only 15 to 20 households, providing an accurate impression of those households (Deckers et al., 2016; Guhn et al., 2010; van Hattem et al., 2009). However, the six-digit classification contained missing values (41.13% of the total sample is complete). Therefore, the five-digit postal code classification (in 57.26% of the total sample) or the four-digit postal code classification (in 1.47% of the total sample) was used to account for missing data on the six-digit classification. Both five-digit and four-digit postal code classifications were based on the same indicators as the six-digit postal code classification. High scores on this variable indicated a high SES, while low scores indicated a low SES.

Data analyses

Because of the hierarchical structure of the data with students (level 1) nested in primary schools (level 2), the data were analysed using a two-level multilevel model in SPSS 26 (Burstein, 1980; Hox et al., 2018). The analyses were conducted for the test-based and the judgement-based recommendations separately. Analyses regarding the test-based recommendations included the variable school leavers' test, because teachers had access

to students' results on this test when formulating a track recommendation, whereas these test results were not available when teachers formulated judgement-based recommendations. The independent variables school leavers' test and SES were grand-mean centred prior to their entrance into the multilevel models (Enders & Tofighi, 2007).

Table 1

	Test-based recommendations			Judgement-based recommendations			Total				
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	Min.	Max.
Track recommendation	4,281	4.55	1.23	4,242	4.66	1.26	8,523	4.61	1.24	1	6
SES	4,385	0.20	0.94	4,242	0.31	0.91	8,627	0.25	0.92	-2.66	3.65
School leavers' test	1,554	538.73	7.25	–	–	–	1,554	538.73	7.25	505	550
Prior achievement											
Reading comprehension test	4,373	61.02	20.20	4,245	64.36	20.60	8,618	62.66	20.46	-27	147
Mathematics test	4,379	112.76	12.87	4,243	114.53	14.20	8,622	113.63	13.57	21	168

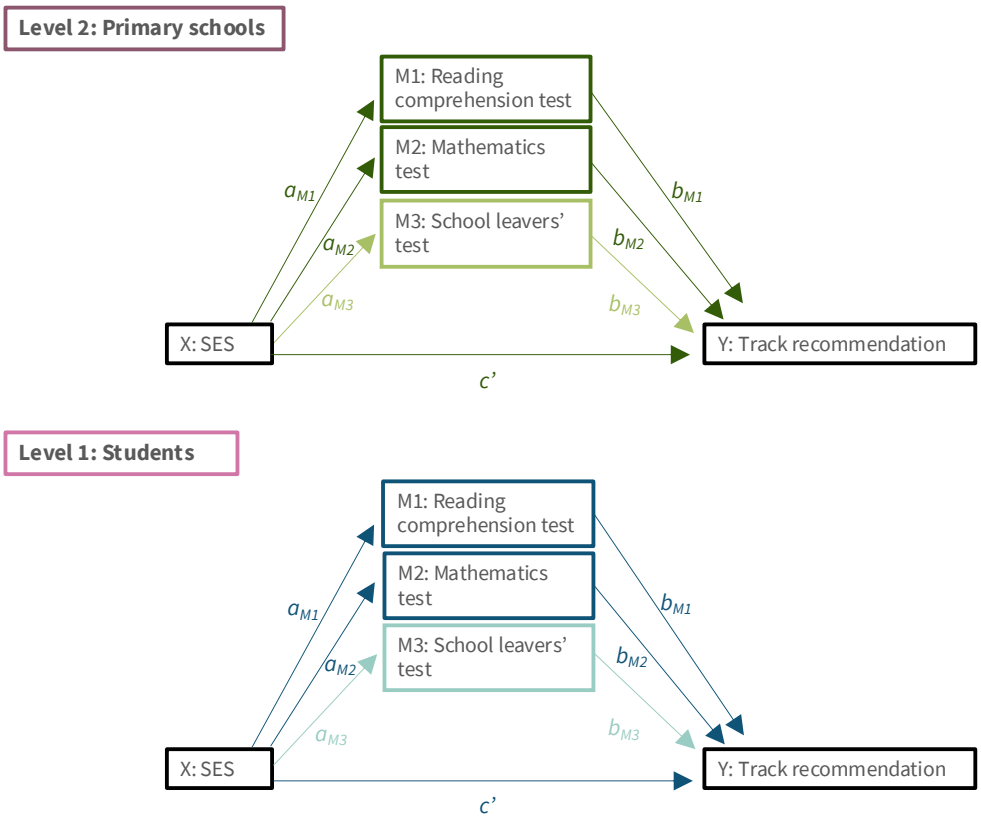
Descriptive Statistics of Track Recommendation, SES, Prior Achievement and School Leavers' Test

The percentage of missing data was extremely low, ranging from 0.1% to 1.3% (exact numbers of *n* of each variable included in the analyses can be seen in Table 1), except for the variable school leavers' test (35.4% complete records of the sample of the first two cohorts). This was due to guidelines of the online student monitoring platform, where our data were downloaded from. Until the policy reform in 2014-2015 it was not mandatory to upload students' school leavers' test scores and, consequently, most schools did not upload them. Comparing data of students who were omitted from the analysis based on missing data on the school leavers' test with data of students who were included revealed some signs of attrition bias. On average, students with missing values on the school leavers' test had a lower SES ($p < .001$), received lower track recommendations ($p < .001$), and performed lower on reading comprehension ($p < .001$) and mathematics ($p < .001$). The effect sizes for these differences were small ($d < 0.26$). In addition, these findings suggested that, even though these differences were small, the missingness was not at random. Given these differences, FIML (Full Information Maximum Likelihood) estimation was applied in the main analyses to account for the missing data (Schafer & Graham, 2002).

First, to examine whether there was an indirect effect of SES on track recommendations through students' reading comprehension and mathematics skills and at which levels the effects were located, multilevel mediation analyses were conducted using the MLmed macro for SPSS (Rockwood, 2017b, 2017a). All parameters for a 1–1–1 mediation model

were estimated in which all variables were measured at level 1, the student level (as can be seen in Figure 1). Monte Carlo stimulations were used to estimate 95% confidence intervals (Rockwood & Hayes, 2017). All independent and mediator variables were automatically grand mean centred by the Mlmed macro prior to their entrance into the multilevel mediation models.

Figure 1



Note. Path a represents the effect of students' SES on students' prior reading comprehension and mathematics achievement. Path *b* represents the effect of students' prior reading comprehension and mathematics achievement on track recommendation. Path *c'* represents the direct effect of students' SES on track recommendation. Path *ab* represents the indirect effect of students' SES on track recommendation through students' prior reading comprehension and mathematics achievement.

Multilevel Mediation Model as Estimated for the Present Study

Second, a multilevel regression model was performed to examine the explained variance of the different independent variables for both types of track recommendations.

Since missing values on the included variables were identified, the number of cases varied according to the model analysed (as can be seen in Table 3 and 4). An unconditional model (Model 0) with school track recommendation as dependent variable was estimated to investigate the distribution of variance at both levels. In Model 1, students' SES was added to the model as a fixed effect to provide information on whether students' SES was a predictor of track recommendations. Model 2 consisted of the indicators of prior achievement as predictors (fixed effects) of track recommendation. For the test-based recommendations, the school leavers' test was also added to this model. In Model 3, all predictor variables were included.

To evaluate the size of differences of the effects of SES between test-based and judgement-based recommendations, effect sizes were calculated by means of Cohen's d, with 0.2, 0.5 and 0.8 as indicative of small, medium, and large effect sizes, respectively (Cohen, 1988). To compare whether the effects of SES on track recommendations for test-based and judgement-based recommendations differed, a Wald test was calculated using the following equation (Brame et al., 1998):

$$Z = \frac{b_{test-based\ recommendation} - b_{judgement-based\ recommendation}}{\sqrt{(SE_{test-based\ recommendation})^2 + (SE_{judgement-based\ recommendation})^2}}$$

Table 2

Test-based / judgement-based track recommendations				
	1	2	3	4
1. Track recommendation				
2. SES	.32***/.29***			
3. Reading comprehension test	.76***/.77***	.26***/.25***		
4. Mathematics test	.76***/.81***	.19***/.23***	.63***/.68***	
5. School leavers' test	.84***/ –	.28***/ –	.64***/ –	.71***/ –

Note. SES = socioeconomic status.
*p < .05; **p < .01; ***p < .001.

Correlations Between Track Recommendation, SES, Prior Achievement and School Leavers' Test

Results

The correlations between track recommendations, students' SES, and prior achievement are presented in Table 2. Overall, high significant positive correlations were reported between students' prior achievement and both types of track recommendations (test-based recommendations: $r_{Reading}$ = .76, r_{Math} = .76; judgement-based recommendations: $r_{Reading}$ = .77, r_{Math} = .81). Additionally, for test-based recommendations, high significant positive correlations were found between the school leavers' test and track recommendations (r = .84). These high correlations indicated that the track recommendations were closely related to students' prior achievement. Finally, for both types of recommendations, low significant positive correlations were reported between students' SES on the one hand, and track recommendations on the other hand (test-based recommendations: r = .32; judgement-based recommendations: r = .27.

Overall Effects of SES

The results of the multilevel regression models, as can be seen in Table 3 (test-based recommendations) and Table 4 (judgement-based recommendations), illustrated that 22.5% of the variance in test-based and 29.2% of the variance in judgement-based recommendations is attributable to factors at school level. As expected, SES was significantly (positively) related to both test-based and judgement-based recommendations, even after controlling for students' prior achievement. The higher students' SES was, the higher the track recommendations they received. However, the impact of SES was small (test-based recommendations: b = .04, p = .016; judgement-based recommendations: b = .05, p < .001). The difference between test-based and judgement-based recommendations was not significant (z = -.50, p = .309), indicating that there was no significant difference between the two types of recommendations regarding to the impact of SES. After accounting for prior achievement, SES explained approximately 0.5% of the variance in track recommendations (test-based recommendations: R_{Model3}^2 = 84.8%; judgement-based recommendations: R_{Model3}^2 = 76.4%).

A substantial part of the differences in track recommendations was explained by students' prior achievement (test-based recommendations: R_{Model2}^2 = 84.4%; judgement-based recommendations: R_{Model2}^2 = 75.9%). For test-based recommendations, results indicated that the school leavers' test explained a larger part of track recommendations than students' reading comprehension and mathematics achievement (school leavers' test: $R_{Model2a}^2$ = 82.4%; reading comprehension and mathematics achievement: $R_{Model2a}^2$ = 70.8%).

Table 3

Test-based recommendations

	Model 0: Empty		Model 1: SES		Model 2: School leavers' test + prior achievement		Model 2a: School leavers' test		Model 2b: Prior achievement		Model 3: SES + school leavers' test + prior achievement	
	b	SE_b	*b*	SE_b	*b*	SE_b	*b*	SE_b	*b*	SE_b	*b*	SE_b
Intercept	4.39***	.06	4.47***	.05	4.66***	.03	4.73***	.03	4.51***	.03	4.67***	.03
SES			.26***	.02							.04*	.02
School leavers' test					.55***	.02	.78***	.01			.55***	.02
Prior achievement												
Reading comprehension test					.23***	.02			.52***	.01	.23***	.02
Mathematics test					.19***	.02			.59***	.01	.19***	.02
Variance												
Student level	1.21***	.03	1.19***	.03	.19***	.01	.22***	.01	.37***	.01	.19***	.01
School level	0.35***	.06	.22***	.04	.05***	.01	.05***	.01	.09***	.01	.04***	.01
R^2			.102		.844		.824		.708		.848	
Model fit												
-2LL	13,213.27		13,075.73		1,921.39		2,152.20		8,052.58		1,914.94	
Number of schools	98		98		88		88		98		88	
Number of students	4,281		4,275		1,506		1,513		4,257		1,504	

Note. The coefficients of prior achievement tests versus school leavers' test differ in magnitude due to differences in scaling.

SES = socioeconomic status.

*p < .05; **p < .01; ***p < .001.

Unstandardised Estimates of Multilevel Models Predicting Test-Based Track Recommendations With SES, School Leavers' Test and Prior Achievement

Table 4

Judgement-based recommendations

	Model 0: Empty		Model 1: SES		Model 2: Prior achievement		Model 3: SES + prior achievement	
	b	SE_b	*b*	SE_b	*b*	SE_b	*b*	SE_b
Intercept	4.41***	.08	4.48***	.07	4.63***	.03	4.64***	.03
SES			.24***	.02			.05***	.01
Prior achievement								
Reading comprehension test					.49***	.01	.49***	.01
Mathematics test					.69***	.01	.69***	.01
Variance								
Student level	1.26***	.03	1.25***	.03	0.36***	.01	.36***	.01
School level	.52***	.08	.39***	.07	0.07***	.01	.06***	.01
R^2			.087		.759		.764	
Model fit								
-2LL	13,294.03		13,185.32		7,869,54		7,838.05	
Number of schools	101		101		101		101	
Number of students	4,242		4,236		4,235		4,229	

Note. The coefficients of prior achievement tests versus school leavers' test differ in magnitude due to differences in scaling.

SES = socioeconomic status.

Unstandardised Estimates of Multilevel Models Predicting Judgement-Based Track Recommendations With SES and Prior Achievement

Student Level

The results of the multilevel mediation analyses are presented in Figure 2 (test-based recommendations) and Figure 3 (judgement-based recommendations). In line with our expectations, at student level, the models revealed significant positive indirect effects (path ab) of students' SES on test-based recommendations through students' prior achievement (school leavers' test: b = .08, p < .001; reading comprehension: b = .03, p < .001; mathematics: b = .03, p < .001). Similar results were reported for judgement-based recommendations (reading comprehension: b = .07, p < .001; mathematics: b = .10, p < .001). The higher students' SES was, the better they performed and the higher the track recommendation they received.

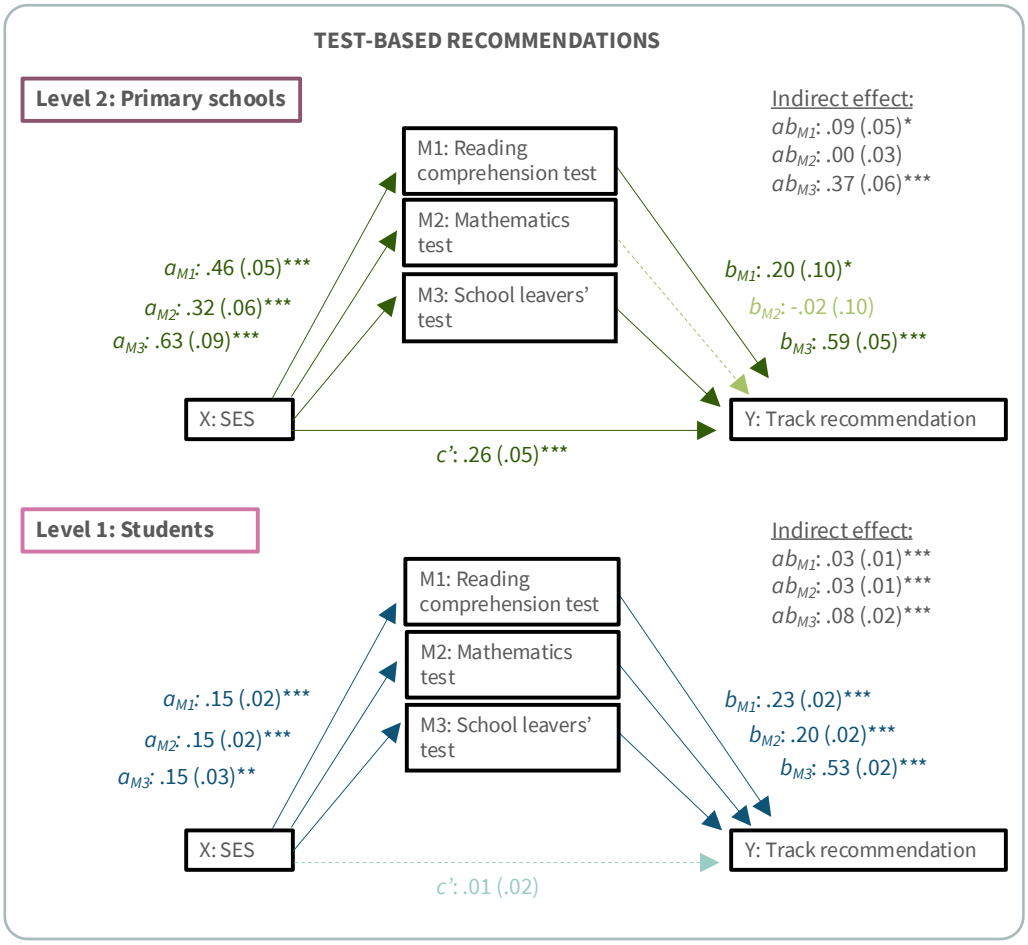
Additionally, in line with our expectations, SES significantly affected judgement-based recommendations (b = .03, p = .009), but did not affect test-based recommendations (b = .01, p = .663) (path *c'*). For judgement-based recommendations, regardless of their prior achievement the higher students' SES was, the higher the track recommendation they received. This finding corresponded to an effect size of 0.03, which can be interpreted as an extremely small effect. This effect indicated that students with a SES of one standard deviation below average received a track recommendation of 0.03 lower than students with a SES of one standard deviation above average.[6] However, the difference between test-based and judgement-based recommendations regarding the direct impact of SES on student level was not significant (z = -1.27, p = .102), indicating that there was no significant difference between the two types of recommendations.

School Level

At school level, the multilevel mediation models presented significant positive indirect effects (path *ab*) of students' SES through students' prior achievement on test-based (reading comprehension: b = .09, p = .044; school leavers' test: b = .37, p < .001) and judgement-based recommendations (reading comprehension: b = .26, p < .001; mathematics: b = .20, p < .001). For test-based recommendations, the indirect effect of students' prior mathematics achievement was not statistically significant (b = .00, p = .885). In addition, a direct effect of SES (path *c'*) on track recommendations at school level was reported for both recommendations (test-based recommendations: b = .26, p < .001; judgement-based recommendations: b = .25, p < .001). These results corresponded to effect sizes of 0.19 for both recommendations, which can be interpreted as small effects. In addition, the difference between test-based and judgement-based recommendations regarding the direct impact of SES at school level was not significant (z = .14, p = .446), indicating that there was no significant difference between the two types of recommendations.

6 The direct effect of SES on track recommendations at student level was significant in cohort 4 (b = .04, p = .016), but not significant in cohort 3 (b = .03, p = .122). Further elaboration of this finding is included in the discussion section.

Figure 2

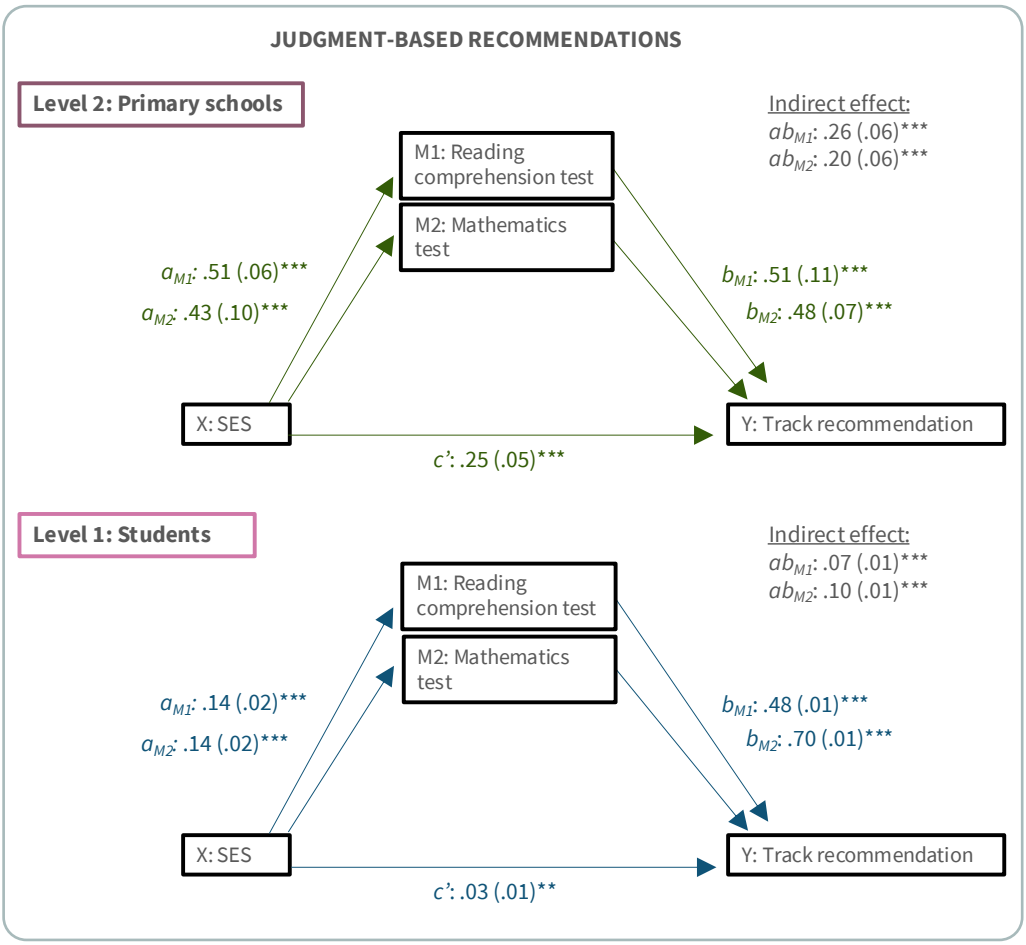


Results of the Multilevel Mediation Model for Test-Based Recommendations

"Vroeger werd er te strikt naar de Cito gekeken en dat is nu minder."

"Verder hebben ouders een mbo-opleiding afgerond: dat is niet relevant voor mij."

Figure 3



Results of the Multilevel Mediation Model for Judgement-Based Recommendations

"Begrijpend lezen en rekenen zijn de belangrijkste factoren en daarin volgen we ook de adviesprocedure die er is."

Discussion

There is an ongoing (political) debate concerning the improvement of educational equality in track recommendations at the transition from primary to secondary education. A policy reform of the Dutch educational system provided us the unique opportunity to compare the direct and indirect effects of students' SES on two types of school track recommendations: test-based and judgement-based recommendations. A multilevel mediation approach was used to examine how these effects of students' SES on track recommendations were distributed across student and school level. Overall, as expected, track recommendations were higher for higher-SES students. This was primarily due to indirect effects, i.e. SES affected prior achievement, and, in turn, students' prior achievement affected their track recommendations. Moreover, a small direct effect of students' SES on judgement-based recommendations was reported, suggesting that students with different SES but similar prior achievement received different track recommendations. In addition, the results indicated that the effects were – at least partly – situated at school level. In general, track recommendations tended to be higher at schools with a population of higher-SES students, irrespective of students' prior achievement. However, these effects were small and the differences between the two types of track recommendations were not significant. These findings did not entirely support the assumption that judgement-based track recommendations are more biased than test-based recommendations (Boone & Van Houtte, 2013; Driessen et al., 2008; OECD, 2016b; van Nuland, 2011). Consequently, the shift to judgement-based recommendations did not seem to lead to a decrease of educational equality. Students' SES played solely a minor role in teachers' track recommendations and the use of the school leavers' test in test-based recommendations did not seem to reduce this effect of SES.

Contrary to previous research (Bol et al., 2014; Caro, Lenkeit, et al., 2009; Luyten & Bosker, 2004), there was no significant difference between test-based recommendations and judgement-based recommendations regarding the impact of students' SES. This finding revealed that, in this context, both types of recommendations were comparable with regard to the (small) impact of students' SES on track recommendations. Luyten and Bosker (2004) reported that not administering a school leavers' test strengthened the impact of SES on track recommendations, whereas the present study provided no support for this finding. There were no differences with regard to the effect of SES on track recommendations between teachers who had access to the results of the school leavers' test and teachers who had no access to these results. The results of the school leavers' test did not add additional information when students' prior achievement in reading comprehension and mathematics was already available to teachers. These findings suggested that teachers relied more strongly on students' prior achievement in reading comprehension and mathematics when results of the school leavers' test were not available. However, in situations where all achievement indicators were available (i.e. when using test-based recommendations), teachers relied more strongly on the results of the school leavers' test when formulating

a track recommendation instead of relying on reading comprehension and mathematics test results. Students' prior achievement explained only 70.8% of the variance in test-based recommendations and 75.9% of the variance in judgement-based recommendations. When the school leavers' test was included, 84.4% of the variance in test-based recommendations was explained by students' prior achievement and the school leavers' test together. Overall, regarding educational equality based on SES, it did not seem to matter whether test-based or judgement-based recommendations were used.

Finally, by using multilevel analyses, it was possible to examine the effects of SES on track recommendations at student and school level, while previous research (Boone et al., 2018; Timmermans, Kuyper, et al., 2015) focused primarily on student or class level. The present study reported no direct effect of students' SES on test-based track recommendations at student level but did report a significant direct effect at school level. These results suggested that there were no differences between test-based recommendations of students with different SES but similar achievement in the same school, whereas these differences were present between schools. In other words, schools with similar achievement but a different population provided different track recommendations. Consequently, it did seem to matter which school students attended. Contrary, although the effect was extremely small, a direct effect of SES on judgement-based recommendations existed at student level. This is in line with results of prior research (Bol et al., 2014; Caro, Lenkeit, et al., 2009; Schneider & Tieben, 2011) revealing that track recommendations were less influenced by SES when they were based on standardised test results instead of teachers' judgements. Because a direct effect of SES on both types of track recommendations existed at school level, no evidence was found for the speculations of Boone and colleagues (2018) that the impact of SES class composition would be more pronounced in educational systems with binding track recommendations as a result of more (higher-SES) parent pressure. They suggested that teachers who were held accountable more strongly for their recommendations, for example when track recommendations were binding, were less prone to be biased by students' background. There was no evidence found for supporting this hypothesis.

Limitations and Future Research

One limitation of the present study is the specific context of this research. The present study was conducted in a large city in the Netherlands, which might affect generalisability to other Dutch regions and countries with different educational systems. School track recommendations may be formulated differently in other countries. In addition, even within the Netherlands, regions have different regulations with regard to formulating track recommendations. The city our data originates from allowed solely singular track recommendations. In other regions it is possible to formulate track recommendations of adjacent tracks. Previous research (Oomens et al., 2019) reported that 20% to 25% of the primary schools used such combined recommendations.

We focused on initial track recommendations formulated in March, not on the actual track placement. These initial recommendations reflect primarily how teachers formulate

a track recommendation. Moreover, including the initial track recommendation allowed us to compare both types of track recommendations. Because the complete procedure of formulating track recommendations and the actual track placement is much more complex, and both types of track recommendations differ in more aspects than included (Oomens et al., 2019), this is beyond the scope of the present research. For example, for test-based recommendations, results of the school leavers' test were available to teachers, whereas for judgement-based recommendations, these results were not available because of a changed time schedule. However, students still make the nationwide school leavers' test but the results primarily serve as a second opinion. Primary school teachers are allowed to adjust their initial recommendations when the school leavers' test results are higher than the initial track recommendation (Korpershoek et al., 2016; Oomens et al., 2019). Prior research (Ministerie van Onderwijs Cultuur en Wetenschap, 2019; Oomens et al., 2019; van Look et al., 2018) has indicated that in most situations in which students were eligible for upward corrections, the initial track recommendation was not corrected. Moreover, a policy evaluation report of Oomens et al. (2019) indicated that corrections of the initial track recommendations did not increase educational equity. Based on parental education, as can be seen as an indicator of SES, equal percentages of corrections were reported for students with lower-educated and higher-educated parents. In addition, the degree of urbanity seemed to be related to these track recommendation corrections. In larger cities, the number of corrections was higher, as well as parental pressure (Oomens et al., 2019). This may also decrease educational equity. Future research could include these topics regarding track recommendation corrections.

In addition, the percentage of missing values across the variables was considerably low (between 0.1% and 1.3% of the total sample), except for the variable school leavers' test. Of the school leavers' test there were only 35,4% complete records in the sample of the first two cohorts. As explained before, this was due to guidelines of the online student monitoring platform, where the data of the present study were downloaded from. Until the policy reform, it was not mandatory to upload students' scores on the school leavers' test, and consequently, most teachers did not upload them. Analyses of the missing values on the variable school leavers' test indicated some signs of attrition bias, which may have impacted the findings of the present study. Yet, although there were indications of bias, it is important to note that these effects were small (i.e. $d < 0.26$). These missing values may be due to schools not having students with lower achievement levels, which are more often lower-SES students, or do not upload the school leavers' test results because it may impact the overall results of the school (Inspectie van het Onderwijs, 2012; Swart et al., 2019). Remarkably, similar missing values for the school leavers' test were identified in other research (Boone & Van Houtte, 2013; Timmermans, Kuyper, et al., 2015). Future research could investigate the underlying causes of this missingness further, as well as the effects on students' track recommendations.

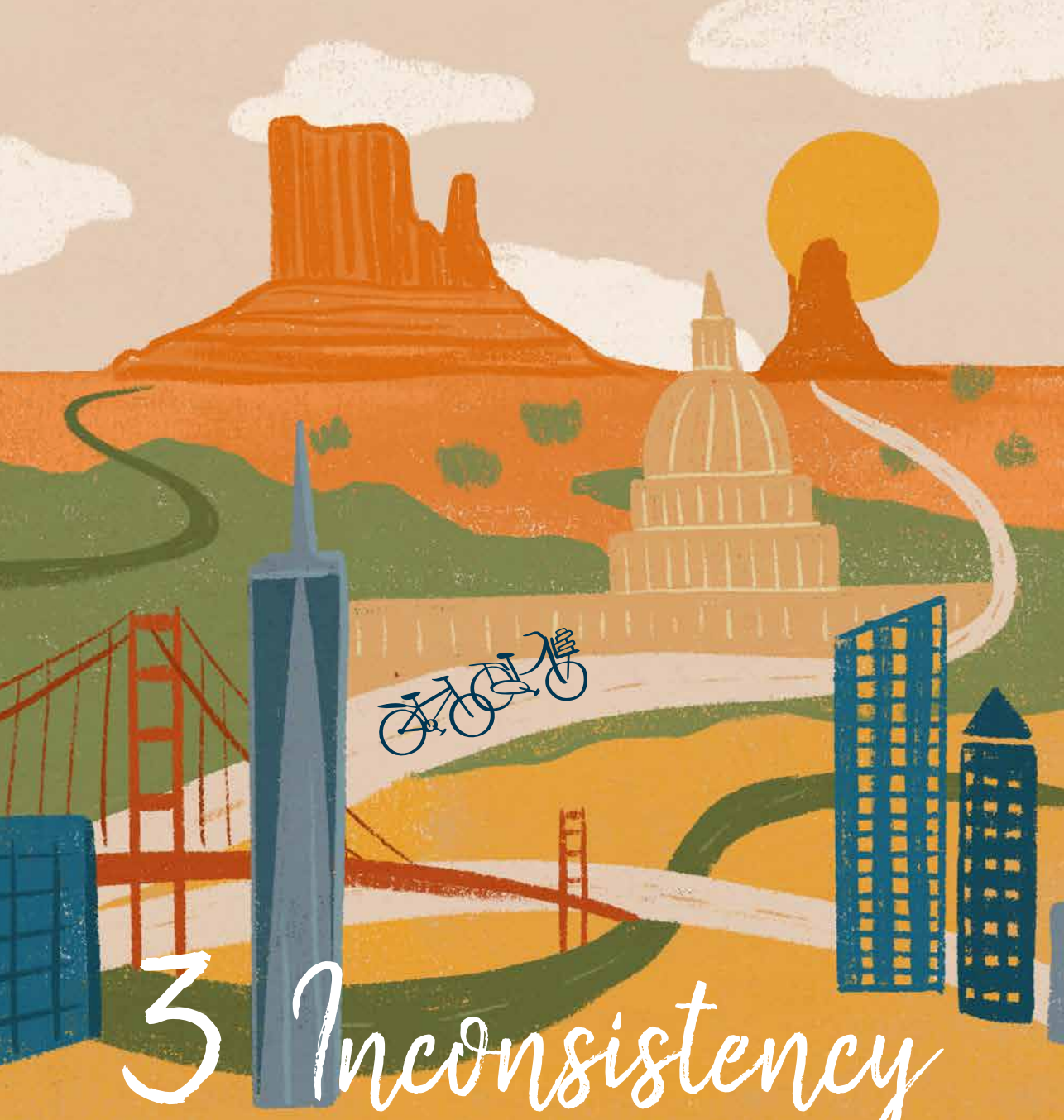
Another limitation of the present study was that a small significant difference between the two cohorts after the policy reform (i.e. cohort 3 and 4) was found. The direct effect of SES on track recommendations at student level was significant in cohort 4 and not significant in cohort 3. Before the policy reform (i.e. cohort 1 and 2), this direct effect on student level was also not significant. It might be that this effect was not yet visible in cohort 3 but emerging in cohort 4 (and potentially in later cohorts). This leads to an interesting question for future research.

Finally, in the present study, SES was determined using this six-digit postal code. For Dutch cities, the six-digit postal codes are, on average, shared by only 15 to 20 households, providing an accurate impression of those households (Deckers et al., 2016; Guhn et al., 2010; van Hattem et al., 2009). However, the six-digit classification contained missing values and, therefore, the missing values were replaced with the five-digit classification that is less precise. Moreover, using the postal codes for measuring SES rather than an individual measure of SES, could have resulted in a small overlap between students' individual SES and school SES. This could affect the results: there was the school level effect of SES on track recommendations, while a part of this effect may actually be situated at the individual level.

For both recommendations, despite the strong effect of students' prior achievement on track recommendations, still 20% of the variance was unexplained either by students' SES or prior achievement. Previous research (Geven et al., 2018; Hornstra et al., 2013; Kaiser et al., 2013; Lleras, 2008) indicated that teachers could also take students' non-cognitive characteristics into account, such as work habits and motivation. In order to decrease the achievement gap based on SES, further research on this topic is needed. Above that, revealing the psychological mechanisms of recommendation bias are important for understanding the formulation of track recommendations and creating equal opportunities to be assigned to a certain secondary school track for all students. This could be included in future research.

Conclusion

Altogether, the findings suggested that teachers were capable of appropriately formulating school track recommendations relying mostly on students' prior achievement without being biased by students' SES. Students' results on the school leavers' test did not seem to be of added value to formulate appropriate school track recommendations. Moreover, it was not the type of track recommendation that was important for educational equality at the transition from primary to secondary education, but rather which and how much (objective) information was available to the teacher when formulating a recommendation.



3 Inconsistency

in Student Achievement Across Subject Domains: Examination of Associations With Students' Gender, Socioeconomic Status, and Teachers' Track Recommendations

Abstract

For students who perform inconsistently across subjects, teachers face challenges in formulating track recommendations, as students' achievement will not point to one secondary school track. This issue may be more prominent for students from diverse backgrounds, given the achievement differences between specific subject domains within these groups. Therefore, the aim of this study was to examine the impact of achievement inconsistency (by comparing standardised achievement levels between reading comprehension and mathematics within students) on students' track recommendations. The sample consisted of 4,248 Grade 6 students from 101 Dutch primary schools. Data were analysed using a t-test, chi-square test and a multilevel moderation model. Results showed that most students performed rather consistently. Approximately 20% of the students performed inconsistently (>1 *SD* difference between subjects). While the overall effect of inconsistency on track recommendations was small, achievement inconsistency primarily seemed to affect track recommendations when the inconsistency was moderate to large. Teachers formulated more 'careful' (i.e. lower) track recommendations when the inconsistency was large. This effect was slightly more pronounced for higher-SES students, with no gender differences. Overall, these findings indicated that a tracked educational system, in which students follow all their courses at the same level, may not be appropriate for the rather substantial group of students whose achievement differs between subject domains.

Keywords

equal educational opportunities, transition from primary to secondary education, school track recommendation, SES, gender, achievement inconsistency, reading comprehension, mathematics, teacher judgement

This chapter is based on: van Leest, A., van de Pol, J., van Tartwijk, J., & Hornstra, L. (2024). Inconsistency in student achievement across subject domains: Examination of associations with students' gender, socioeconomic status, and teachers' track recommendations. *Educational Research and Evaluation, 29*(5), 274-298. https://doi.org/10.1080/13803611.2024.2345197

Acknowledgement of author contributions: All authors contributed to the design of the study. AvL collected the data and prepared the data for analyses; AvL analysed the data with help from LH; AvL drafted the manuscript; All authors contributed to critical revision of the manuscript; LH, JvdP, and JvT supervised the study.

Introduction

In some tracked educational systems, teachers' track recommendations at the end of primary education are used to allocate students to different hierarchical secondary school tracks that determine their future educational career (Glock et al., 2012; G. M. Strand, 2020). Prior research (Feron et al., 2013; Timmermans, Kuyper, et al., 2015; van Leest et al., 2020) has indicated that these track recommendations are mostly based on students' (prior) achievement in multiple subject domains, where correlations between subjects – for example, reading comprehension and mathematics – imply relatively consistent achievement within a student across different subject domains. However, some students may show discrepancies between these different subject domains (Luyten, 1998). As secondary school tracks typically offer the same educational level for all subject domains, teachers may find it difficult to formulate track recommendations for students who perform inconsistently, as their achievement will not point directly to one specific level of secondary education. In addition, achievement inconsistency may occur more often or look different among students with different background characteristics, such as gender or socioeconomic status (SES), due to different achievement levels in specific subject domains among these students. In case of achievement inconsistency, boys and lower-SES students typically perform lower in the language domain, whereas girls and higher-SES students with achievement inconsistency more often perform lower in mathematics (e.g. Hakkarainen et al., 2013; Jacobs & Wolbers, 2018; Plante et al., 2013; Sirin, 2005; Uerz et al., 2004; van Leest et al., 2020). These group differences in achievement across different subject domains may cause differences in teachers' track recommendations. This is particularly true if teachers place greater emphasis on a particular subject domain when formulating these track recommendations (Driessen et al., 2008; Smeets et al., 2014). For example, if teachers base their track recommendations more strongly on reading comprehension than mathematics, this may disproportionally affect lower-SES students and boys, especially if these groups have larger achievement inconsistencies than other groups.

However, it is thus far unknown how achievement inconsistency affects teachers track recommendations. It may lead to 'careful' (i.e. lower) recommendations' based on students' achievement in their weakest domain, 'aggregated recommendations' based on the mean achievement of two domains, or alternatively, 'the benefit of the doubt-recommendation' based on students' achievement in their strongest domain. Moreover, achievement inconsistency may lead to differences between teachers in which track recommendations they formulate for such students, with some teachers opting for more careful recommendations and others for aggregated or higher recommendations. If teachers do indeed consider achievement inconsistency when formulating track recommendations, this could impact students' opportunities to attend higher secondary school tracks. For instance, if lower-SES students show more achievement inconsistency between subject domains than higher-SES students, and teachers tend to be more cautious in their track recommendations for

students with inconsistent achievement, this might result in reduced opportunities for lower-SES students to attend higher tracks.

To improve educational equality at the transition from primary to secondary education for all students, it is important that teachers formulate track recommendations that fit students' abilities and potential best (Tieben & Wolbers, 2010). Given the scarcity of prior research, it is not clear how achievement inconsistency across different subject domains affects students' track recommendation in general, or whether teachers consider achievement inconsistency differently for students with different background characteristics. Therefore, the aim of the present study was to examine the extent to which achievement inconsistency between language (i.e. reading comprehension) and mathematics within students occurs, how this is associated with students' SES and gender, and, in turn, with teachers' track recommendations.

(In)consistency of Students' Achievement and Track Recommendations

In some tracked educational systems, such as France, Germany, Luxembourg and the Netherlands, teachers have to provide a recommendation for the school track that students will attend in secondary education (Boone & Van Houtte, 2013; Glock et al., 2012; Korpershoek et al., 2016; Le Métais, 2003; Timmermans, Kuyper, et al., 2015). These track recommendations are formulated at the end of primary school, whereby teachers recommend a type of secondary education that they consider to match the student's potential ability level best (de Boer et al., 2010; Glock et al., 2013; Timmermans, Kuyper, et al., 2015). These track recommendations are important, since they determine students' allocation to a specific educational level in secondary school, and thereby the educational qualifications students can acquire (Glock et al., 2012; G. M. Strand, 2020). These track recommendations are based on teachers' expectations, i.e. teachers' inferences about students' potential behaviour or achievement (de Boer et al., 2010; Klapproth et al., 2012; Timmermans, Kuyper, et al., 2015). Although the different factors that teachers consider when formulating track recommendations may differ per educational system, depending on the extent to which teachers are able to include their own perceptions in the track recommendations, research in various countries has shown that in general track recommendations are primarily based on students' (standardised) achievement in reading comprehension and mathematics (Bos et al., 2004; Ditton & Krüsken, 2006; Feron et al., 2013; Geven et al., 2021; Klapproth et al., 2012; Südkamp et al., 2012; Timmermans, Kuyper, et al., 2015).

Although most students perform rather similar (or consistently) across both domains (i.e. there is a strong positive correlation between reading comprehension and mathematics), there are also students whose achievement across subject domains differs: their achievement across domains is inconsistent (e.g. high achievement in language and lower achievement in mathematics or vice versa; Luyten, 1998; Timmermans, de Wolf, et al., 2015; van Leest et al., 2020). If students perform rather consistent, students' achievement in both

subjects will likely indicate a similar level of secondary education (Böhmer et al., 2015), whereas inconsistent achievement will not point directly to one specific level of secondary education. Hence, it seems plausible that it is more difficult for teachers to formulate a track recommendation for these students.

Different scenarios seem possible when students' achievement is inconsistent. Teachers can (a) place most emphasis on the subject domain with the lowest achievement (a 'careful' recommendation), (b) place most emphasis on the subject domain with the highest achievement ('giving the benefit of the doubt'), or (c) aggregate achievement in different subject domains (an 'aggregated' recommendation). In addition, teachers may also weigh one subject domain (i.e. language or mathematics) more heavily than the other ('a predominant subject domain'). In the Netherlands, students' mathematics achievement seems most decisive for the level of education a teacher recommended, followed by language achievement (Driessen et al., 2008; Smeets et al., 2014).

Research on inconsistency of students' achievement across different subject domains is scarce. Previous research primarily focused on the consistency of students' cognitive versus socio-emotional characteristics (Böhmer et al., 2015; Glock et al., 2013), students' grades versus standardised test scores within subject domains (Glock et al., 2013), or teacher judgements versus standardised test scores (Südkamp et al., 2012; Timmermans, Kuyper, et al., 2015; van Leest et al., 2020; van Rooijen et al., 2016). To our knowledge, only one study by Glock et al. (2013) examined the relation between inconsistencies of achievement in different subject domains and track recommendations. Glock et al. (2013) conducted two different experimental procedures based on vignettes, yielding mixed results. In the first experiment, where teachers reviewed numerous student profiles simultaneously, there was no significant effect of inconsistency on track recommendations. However, in the second experiment, where teachers formulated a track recommendation for each student before moving to the next one, students with inconsistent achievement received lower track recommendations than students with consistent achievement. According to the researchers, these mixed results were likely due to the difference in experimental procedures between the two studies. Given their mixed findings and the use of vignettes, it remains unclear how inconsistencies in students' achievement affect track recommendations of actual students.

Impact of Students' Gender and SES

The effect of students' inconsistent achievement on track recommendations can be related to their background characteristics due to differences in achievement (Timmermans, Kuyper, et al., 2015; van Leest et al., 2020; Wang et al., 2018). In the present study, students' gender and SES were included. Prior research (Gentrup & Rjosk, 2018; Plante et al., 2013; Spinath et al., 2014; Uerz et al., 2004) indicated that boys had stronger mathematical abilities than girls, and girls had stronger language and reading abilities than boys. Therefore, it could be that achievement inconsistencies occur as often for boys as for girls, but in different directions with boys having higher mathematics than language

achievement and girls having higher language than mathematics achievement. If teachers rely more often on mathematics when formulating track recommendations (Driessen et al., 2008; Smeets et al., 2014), it could be that boys receive higher track recommendations than girls with the same average achievement. Yet, it is unknown whether the effect of inconsistency on track recommendations is different for boys and girls. In addition to the indirect effects of gender on track recommendations due to inconsistent achievement, there could potentially be differences between boys and girls in track recommendations despite their achievement levels. Previous research in various countries reported mixed findings with regard to the effect of students' gender on track recommendations. Some research (e.g. Dutch Inspectorate of Education, 2014; Jürges & Schneider, 2011; Timmermans et al., 2016) reported that boys received on average a lower track recommendation than girls, whereas other research (e.g. Boone & Van Houtte, 2013; Driessen, 2005; Klapproth et al., 2013; Krolak-Schwerdt et al., 2017; Timmermans et al., 2018; Van Rooijen et al., 2016) found (almost) no difference between boys and girls.

Prior research also showed differences in achievement between students with different SES backgrounds. The higher students' SES, the better they perform on reading comprehension and mathematics, and the higher the track recommendation they receive (Caro, Lenkeit, et al., 2009; van Leest et al., 2020). Lower-SES students specifically perform lower on language than on mathematics achievement compared to higher-SES students (Jacobs & Wolbers, 2018; Sirin, 2005; van Leest et al., 2020). Thus far, it is unclear whether the effect of inconsistency on track recommendations is different for students with different SES. In addition, it has been repeatedly found in various countries that students' SES has a small direct impact on track recommendations. That is, irrespective of students' prior achievement, track recommendations are, on average, slightly more positive for higher-SES students than for lower-SES students (e.g. Batruch et al., 2023; Boone & Van Houtte, 2013; Caro & Lehmann, 2009; Driessen et al., 2005, 2007; Feron et al., 2016; Klapproth et al., 2012; Korpershoek et al., 2016; Luyten & Bosker, 2004; Pit-ten Cate et al., 2016; Timmermans et al., 2013, 2016, 2018; Van Rooijen et al., 2017).

Present Study

Because track recommendations have a strong impact on students' future educational careers (de Boer et al., 2010; van Rooijen et al., 2017), it is important that teachers formulate the most appropriate track recommendations according to students' (potential) abilities (Tieben & Wolbers, 2010). However, it may be more difficult for teachers to formulate a track recommendation when a student's achievement between different subject domains is inconsistent. To our knowledge, there are no studies examining the effects of achievement inconsistency on track recommendations in a naturalistic setting.

The present study addressed two research questions (RQs) in the context of inconsistency in student achievement in the Dutch educational system. The first research question was: To what extent does achievement inconsistency between reading comprehension and mathematics occur (RQ1a) and to what extent is achievement inconsistency related

to students' gender and SES (RQ1b)? In addition, it was explored whether the direction of students' achievement inconsistency (that is, in which subject domain students showed the highest achievement: reading comprehension or mathematics) differed based on students' gender and SES (RQ1c). The second research question was to examine whether inconsistency predicted track recommendations beyond students' gender, SES, and overall achievement (RQ2a). Furthermore, it was examined whether teachers considered achievement inconsistency differently for students with different gender or SES when formulating track recommendations (RQ2b). As there may be variation between schools in how the track recommendations are formulated, it was also explored whether the effects of students' prior achievement and the achievement inconsistency on track recommendations differed between schools (RQ2c). In the absence of prior research on the topic of inconsistency, no detailed expectations were formulated regarding the relation between inconsistency, gender, SES, and track recommendations, except for the direction of the inconsistency. Based on prior research (Gentrup & Rjosk, 2018; Hakkarainen et al., 2013; Jacobs & Wolbers, 2018; Plante et al., 2013; Sirin, 2005; Uerz et al., 2004), it was expected that in case of achievement inconsistency, boys were more likely to show better achievement in math than in language, and vice versa for girls, while lower-SES students overall perform lower than higher-SES students, but specifically lower on language than on mathematics.

Method

Sample

The data used in the present study were part of a larger dataset on students' educational development across the transition from primary to secondary education, including data from an online student monitoring platform containing different kinds of information about students, such as students' educational achievement in primary school and background characteristics (van Leest et al., 2020). An organisation representing primary schools with access to this online monitoring platform downloaded and anonymised the data from schools who approved using their data.

Our sample consisted of 4,248 Grade 6 students from 101 primary schools in a large city in the Netherlands. Students were from two cohorts: students who were in Grade 6 of primary education in the academic year 2014-2015 (50.4% of the sample) and 2015-2016 (49.6% of the sample).[7]

7 In 2014, there was a policy reform of the Dutch educational system regarding the track recommendation procedures. The most important change of this reform was a changed time schedule, that resulted in not having results of the standardised school leavers' test available to teachers when formulating a track recommendation. This revised tracking recommendation procedure was followed in both cohorts from the present study.

Dutch Educational System. In the Netherlands, students attend primary school until the age of twelve (OECD, 2020; Smeets et al., 2014; Strello et al., 2021). Whereas primary education consists of basic education without tracking, secondary education is organised hierarchically in different ability tracks. In the final grade (sixth grade) of primary educations, students receive a track recommendation formulated by their primary school teacher for one of the six hierarchical secondary school tracks (Naayer et al., 2016; Smeets et al., 2014). Secondary schools are required to allocate students to the secondary school track determined by the track recommendation (Ministerie van Onderwijs Cultuur en Wetenschap, 2014). Therefore, students are (almost) always allocated to the secondary school track indicated by their track recommendation. Students also make a mandatory nationwide school leavers' test. However, teachers do not have access to the results of this test when formulating track recommendations due to the fact that the test is taken after teachers formulated their track recommendations (Korpershoek et al., 2016; Oomens et al., 2019). In Dutch secondary education, each secondary school track represents a different educational path, including different educational qualifications for tertiary education. Switching upwards and downwards between different secondary school tracks, i.e. intra-secondary transitions, is possible, but does not happen very often due to limited possibilities within or between schools (Jacob & Tieben, 2009; Lek & van de Schoot, 2019; LeTendre et al., 2003; OECD, 2016b; Schnepf, 2002; Tieben & Wolbers, 2010). Therefore, placement in the first year of secondary education, which is based on the track recommendation, is very decisive for students' future educational careers.

Measures

Track Recommendation. In the Netherlands, teachers formulate an initial track recommendation at the end of primary education before a school leavers' test is administered[8], but teachers do have access to standardised test scores throughout students' primary school career. The six secondary school tracks, from lowest to highest track, are: (1) practical training, (2) basic pre-vocational secondary education, (3) middle pre-vocational secondary education, (4) theoretical pre-vocational education, (5) senior general secondary education, and (6) pre-university education. Track recommendation was considered as a continuous variable (cf. Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993).

Prior Achievement. Students' most recent reading comprehension and mathematics scores on standardised tests of primary schools' monitoring and evaluation system were included as measures for prior achievement, as these scores are generally most predictive of track recommendations (PO-raad & VO-raad, 2014). Most tests were conducted in December

8 In the city in which the data was collected, the initial track recommendations do not allow for combined track recommendations of adjacent tracks.

or January of Grade 6, otherwise an earlier test, at least conducted halfway Grade 5, was included. The schools participating in this study, all used standardised tests developed by Cito (the Dutch National Institute for Educational Measurement). The reading comprehension test scores range from -87 to 147, and mathematics test scores range from 0 to 168 (Cito, 2016). Prior research (Feenstra et al., 2010; Janssen et al., 2010) provided support for high internal consistency (α >.80) and high validity of these tests. Because different but comparable test versions were used, the test scores were standardised for each test version to account for potential differences between test versions.

Achievement Inconsistency. An absolute discrepancy score between students' standardised achievement scores in reading comprehension and mathematics (see variable prior achievement for a description of both tests) was computed to indicate achievement inconsistency within a student. This score was calculated by subtracting students' standardised mathematics score from standardised reading comprehension score when their reading comprehension score was higher, and vice versa. Higher scores indicated a higher level of inconsistency between the two subject domains. Achievement inconsistency ranged from 0.00 to 4.54, with 1.00 meaning that there was 1 standard deviation difference between the achievement in both subject domains.

Direction of Inconsistency. Based on the variable achievement inconsistency, a variable to indicate the direction of inconsistency was created. Students were divided into five groups: (1) students with 1 to 2 *SD* higher achievement in mathematics, and (2) students with 1 to 2 *SD* higher achievement in reading comprehension, (3) students with 2 or more *SD* higher achievement in mathematics, (4) students with 2 or more *SD* higher achievement in reading comprehension, and (5) students with less than 1 *SD* difference between the two subject domains (i.e. the reference group).

Gender. A dichotomous dummy variable was created for students' gender; boys formed the reference group (50.1% of the total sample).

Socioeconomic Status (SES). Students' six-digit postal code was used as an approximation of students' families' SES, as this measure can be an useful marker of SES (e.g. Danesh et al., 1999). The SES variable was composed of three indicators, provided by Statistics Netherlands (CBS): (a) the most recent mean household income after tax, (b), the mean real estate value, and (c) the number of people who are unemployed or have social welfare benefits (van Leeuwen, 2019). Using principal component analysis (PCA), the indicators were recoded into a factor score. For Dutch cities, the six-digit postal code provides a valid indication of SES, shared by only 15 to 20 households (Deckers et al., 2016; Guhn et al., 2010; van Hattem et al., 2009). The six-digit postal codes contained missing data (39.5% of the total sample is complete). In these cases, SES was estimated based on the five-digit postal codes (97.1% of the missing values) and the four-digit postal codes (2.7% of the

missing values). For 0.1% of the total sample, the postal code was completely missing. The five-digit and four-digit postal code classifications were composed of the same indicators as the six-digit postal code classification. Students' SES is a continuous variable ranging from -2.66 to 3.65. A higher score on this variable indicated a higher SES.

Data Analyses

First, to examine the occurrence and direction of achievement inconsistency, the descriptive statistics were examined. In addition, to examine the association between (the direction of) achievement inconsistency and students' gender and SES, an independent samples t-test for gender, and a chi-squared test for SES were performed. Second, to examine whether the inconsistency of students' achievement predicted track recommendations beyond students' overall achievement, gender and SES, a two-level multilevel model was estimated in SPSS 27 with students (level 1) nested in schools (level 2) (Burstein, 1980; Hox et al., 2018). To investigate the distribution of variance at both levels, an empty model with school track recommendation as dependent variable was estimated (Model 0). In Model 1, students' prior achievement was added to investigate whether students' achievement was a predictor of track recommendation. Next, students' background characteristics gender and SES were added as fixed effects to examine whether students' gender and SES were predictors of track recommendation on top of prior achievement (Model 2). In Model 3a, achievement inconsistency was added as a continuous variable to examine whether students' achievement inconsistency predicted track recommendations. In addition, an additional analysis (Model 3b) was performed to examine whether the effects of inconsistency were related to the level and direction of inconsistency. Subsequently, interaction effects of students' background characteristics gender and SES with achievement inconsistency were included to investigate whether the effects of inconsistency were dependent on gender or SES (Model 4). Finally, a model in which random slopes were allowed for prior achievement and inconsistency was estimated to examine whether the relationships between achievement inconsistency, prior achievement, and track recommendations differed between schools (Model 5). Explained variance (R^2) was calculated for all models, including the random slopes model (cf. Snijders & Bosker, 2012). Effect sizes were based on standardised regression coefficients with 0.2, 0.5, and 0.8 as indicative of small, medium, and large effect sizes, respectively (Cohen, 1988). To facilitate the interpretation of the results, all continuous variables were standardised prior to being included in the analyses.

The dataset contained a low percentage of missing data, varying between 0.0% and 0.2% of the total sample (see Table 1 for the *N* of all variables). The Little's MCAR (Missing Completely at Random) test demonstrated the data can be considered as being missing completely at random (X^2 (7) = 10.09, p = .183). In other words, missing values were not systematically related to values of other variables in the dataset, meaning that there was no sign of attrition bias. Although the percentage of missing data was low, it was accounted for by the FIML (Full Information Maximum Likelihood) method (Schafer & Graham, 2002). The analyses were conducted for both cohorts together. [9]

Results

Descriptive statistics for all variables are reported in Table 1, and correlations between variables are reported in Table 2.

Table 1

	N	*M*	*SD*	Min.	Max.
Achievement inconsistency	4241	0.61	0.52	0.00	4.54
SES	4242	0.31	0.91	-2.66	3.65
Prior achievement					
Reading comprehension test	4245	64.36	20.60	-27	147
Mathematics test	4243	114.53	14.20	21	168
Track recommendation	4242	4.66	1.26	1	6

Descriptive Statistics of Track Recommendation, SES, Prior Achievement, and Achievement Inconsistency

Table 2

	1	2	3	4	5
1. Achievement inconsistency					
2. Gender (0 = boys)	.00				
3. SES	-.06***	-.03			
4. Reading comprehension test	.04**	.06***	.25***		
5. Mathematics test	-.09***	-.11***	.23***	.68***	
6. Track recommendation	-.10***	-.03	.29***	.77***	.81***
*p < .05; **p < .01; ***p < .001.					

Correlations Between Achievement Inconsistency, SES, Gender, Prior Achievement, and Track Recommendation

9 There were no significant differences between the two cohorts in the descriptive statistics of the variables (e.g. means, standard deviations, maximum values) (p values all > .05). Nevertheless, we also performed the multilevel analyses for both cohorts separately to check for differences between the cohorts, and no differences between the two cohorts were found. Therefore, only the combined analyses for both cohorts are reported.

Inconsistency of Students' Achievement (RQ1)

The first aim of the present study was to examine the extent to which achievement inconsistency between mathematics and language occurred across students (RQ1a), and the extent to which inconsistency was related to students' gender and SES (RQ1b). Additionally, it was explored whether the direction of students' achievement inconsistency was different for students based on their gender and SES (RQ1c).

Concerning RQ1a on the extent of achievement inconsistency, the results (see Table 1) indicated that students had a mean inconsistency score of 0.61, meaning that, overall, there was an inconsistency of 0.61 standard deviation between the two subject domains. There was a high significant positive correlation between students' reading comprehension and mathematics achievement, indicating that most students perform consistently across subject domains (see Table 2). For the majority of the students (81.4%) the difference between the two subject domains was less than one standard deviation. Hence, the other 18.6% of the students had one or more standard deviation of difference between the two subject domains. Of these students, 12.1% (i.e. 2.2% of the total sample) had a difference of two or more standard deviations between the two subject domains.

As can be seen in Table 2, there was a small significant positive correlation between students' reading comprehension and achievement inconsistency and a small significant negative correlation between students' mathematics achievement and achievement inconsistency, indicating that students with higher reading comprehension had higher achievement inconsistency (i.e. a larger discrepancy between reading comprehension and mathematics achievement). Of the students whose achievement differed one or more standard deviations, 52.0% performed higher in reading comprehension than in mathematics, and, consequently, 48.0% performed higher in mathematics than in reading comprehension. Of the students whose achievement differed two or more standard deviations, 71.6% performed higher in reading comprehension than in mathematics, and, consequently, 28.4% performed higher in mathematics than in reading comprehension.

Concerning RQ1b, no significant differences in achievement inconsistency between boys and girls were found ($t(4239) = -0.07$, $p = .946$), as can be seen in Table 2. As expected for RQ1c, boys performed significantly higher in mathematics ($t(4241) = 6.97$, $p < .001$) than girls, while girls performed significantly higher in reading comprehension ($t(4243) = -3.98$, $p < .001$) than boys. Furthermore, boys with achievement inconsistency (≥ 1 *SD*) more frequently had higher mathematics than reading comprehension achievement (66.0% of the boys), while girls with achievement inconsistency (≥ 1 *SD*) more frequently had higher reading comprehension than mathematics achievement (71.5% of the girls), X^2 (4, $N = 4241$) = 75.75, $p < .001$. Regarding SES, a small significant negative correlation between students' SES and the inconsistency of students' achievement was found (see Table 2), indicating that lower-SES students had higher achievement inconsistency compared to higher-SES students (RQ1a). Concerning RQ1b, there was no significant relation between the student's SES and the direction of the inconsistency (that is, whether students performed higher in reading comprehension or mathematics), X^2 (8, $N = 4235$) = 10.75, $p = .217$.

Inconsistency and Track Recommendation (RQ2)

The second aim of the present study was to examine the extent to which inconsistency of students' achievement between reading comprehension and mathematics was associated with their track recommendation (RQ2a), and the extent to which achievement inconsistency had a different effect on track recommendations based on students' gender and SES (RQ2b).

The results of the multilevel regression models are presented in Table 3. The results of Model 0 revealed that 29.2% of the variance in track recommendations was attributable to factors at school level, and the remaining 70.8% to factors at student level. By adding students' prior achievement to the model, Model 1 indicated that the variance in track recommendations was primarily explained by students' prior achievement ($R_{\text{Model.1}}^2$ = 75.9%). As expected, students' mathematics achievement most strongly predicted track recommendations. After accounting for prior achievement, students' gender and SES together explained approximately 0.6% of the variance in track recommendations ($R_{\text{Model.2}}^2$ = 76.4%). For gender, no significant effect on track recommendations was found. For SES, there was a small significant effect of SES on track recommendations, indicating that lower-SES students received lower track recommendations.

Concerning RQ2a, the results of Model 3a illustrated that, after accounting for prior achievement, gender, and SES, the inconsistency of students' achievement was negatively associated with track recommendations ($b^\star$ = -.07, p < .001). This finding suggests that teachers tended to give lower track recommendations when the achievement inconsistency between subject domains was larger. Yet, the effect size was small; on top of students' SES, gender and prior achievement, the achievement inconsistency explained only 0.1% of the variance in track recommendations ($R_{\text{Model.3a}}^2$ = 76.5%).

To further examine the effects of the level and direction of inconsistency in additional analyses, students were divided into five groups based on the degree of difference between the two subject domains: (1) within 1 *SD* difference, (2) 1 to 2 *SD* difference mathematics higher than reading comprehension, (3) 1 to 2 *SD* difference reading comprehension higher than mathematics, (4) 2 or more *SD* difference with mathematics higher than reading comprehension, and (5) 2 or more *SD* difference reading comprehension higher than mathematics (i.e. the reference group). The results of Model 3b revealed that, when controlling for students' prior achievement, achievement inconsistency primarily had an effect on students' track recommendations when students had a difference between 1 and 2 standard deviations between the two subject domains. Those students received lower track recommendations than students with low achievement inconsistency (i.e. < 1 *SD* difference between the two subject domains). When the achievement inconsistency was large (i.e. ≥ 2 *SD* difference), only students with higher mathematics than reading comprehension received lower track recommendations compared to students with low inconsistency.

In Model 4, the interactions of students' background characteristics and their achievement inconsistency were added to the multilevel model to examine whether the effect of inconsistency on track recommendations differed for students with different gender or SES (RQ2b). The interaction of inconsistency with students' gender was non-significant. Hence, boys and girls with the same level of discrepancies between their achievement received similar track recommendations. Additionally, there was a small significant interaction effect of students' SES and achievement inconsistency on track recommendations. That is, the effects of achievement inconsistency were slightly stronger for students with a higher SES. As can be seen in Figure 1, this was a small effect. For presentation purposes of the figure, students' SES was divided in three SES groups (lower-SES, middle-SES and higher-SES) and controlled for the variables prior achievement and gender. Even though the regression line was slightly steeper for the higher-SES group, the overall differences in track recommendations – which were mostly due to differences in overall achievement – were much larger in comparison. After accounting for students' gender, SES, prior achievement and achievement inconsistency, the interaction effects together explained only 0.3% of the variance in track recommendations ($R_{Model.4}^2$ = 76.8%).

Model 5 included random slopes for students' prior achievement and achievement inconsistency to examine whether the effects of achievement and inconsistency differed between schools (RQ2c). The results indicated significant random slopes for prior achievement in both subject domains. Additionally, a non-significant random slope for achievement inconsistency was found, indicating that the small negative effect of inconsistency on track recommendations was similar across schools. Including random slopes added 2.3% to the explained variance in track recommendations ($R_{Model.5}^2$ = 79.1%).

"Mijn leerlingen zitten op taalgebied vaak op een lager niveau doordat ze bijvoorbeeld thuis geen Nederlands spreken of minder Nederlands. Hierdoor is het soms moeilijker in te schatten op welk niveau je de leerling moet plaatsen. Spelling lukt meestal met goed oefenen wel, maar vooral woordenschat en begrijpend lezen is een probleem."

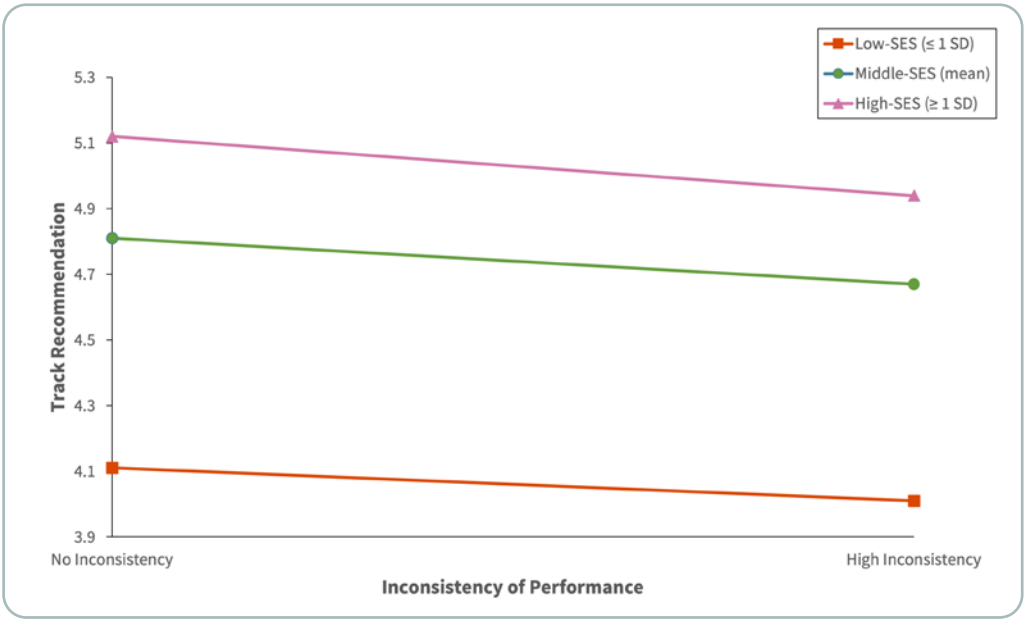
Table 3

	Model 0: Empty		Model 1: Prior achievement		Model 2: + SES + gender		Model 3a: + Achievement inconsistency		Model 3b: + Interaction effects		Model 4: + Interaction effects		Model 5 + Random slopes	
	*b**	SE_b	*b**	SE_b	*b**	SE_b	*b**	SE_b	*b**	SE_b	*b**	SE_b	*b**	SE_b
Fixed effects														
Intercept	.07	.05	.22***	.02	.23***	.02	.24***	.02	.26***	.02	.23***	.02	.23***	.02
Prior achievement														
Reading comprehension test			.34***	.01	.33***	.01	.35***	.01	.33***	.01	.35***	.01	.36***	.01
Mathematics test			.48***	.01	.48***	.01	.47***	.01	.48***	.01	.47***	.01	.47***	.01
Students' background characteristics														
Gender (girls)					-.02	.01	-.01	.01	-.01	.01	-.01	.01	-.02	.01
SES					.03***	.01	.03***	.01	.03***	.01	.03***	.01	.03***	.01
Achievement inconsistency							-.07***	.01			-.06***	.01	-.06***	.01
Interaction effects														
Achievement inconsistency * Gender											-.01	.01	-.02	.01
Achievement inconsistency * SES											-.03***	.01	-.02**	.01
Achievement inconsistency groups (ref group = achievement inconsistency < 1 SD)														
1 to 2 SD difference														
Mathematics > reading comprehension									-.15***	.03				
Reading comprehension > mathematics									-.15**	.03				
2 or more SD difference														
Mathematics > reading comprehension									-.37***	.09				
Reading comprehension > mathematics									-.12	.06				
Random effects														
Slope reading comprehension test													.01**	.00
Slope mathematics test													.01***	.00
Slope achievement inconsistency													.00	.00
Variance														
School level	.25***	.04	.03***	.01	.03***	.01	.03***	.01	.03***	.01	.03***	.01	.02***	.00
Student level	.60***	.01	.17***	.00	.17***	.00	.17***	.00	.17***	.00	.16***	.00	.15***	.00
R^2			.759		.764		.765		.767		.768		.791	
Model fit (-2LL)	10133.85		4714.58		4685.71		4585.84		4590.40		4558.69		4423.40	
Number of schools	101		101		101		101		101		101		101	
Number of students	4242		4235		4229		4229		4229		4229		4229	

*p < .05; **p < .01; ***p < .001.

Standardised Estimates of Multilevel Models Predicting Track Recommendations With Prior achievement, SES, Gender, and Achievement Inconsistency (Groups)

Figure 1



Multilevel Regression Lines for the Effect of Students' Achievement Inconsistency on Track Recommendations for Students With Different SES Backgrounds

Discussion

In tracked educational systems, track recommendations determine students' allocation in secondary education, and thereby, students' educational careers (Glock et al., 2012; G. M. Strand, 2020). Therefore, it is important that track recommendations are based on students' abilities and potential. It may be challenging for teachers to formulate track recommendations when students perform inconsistently across different subject domains, as their achievement will not directly indicate one particular level of secondary education. The aim of this study was to investigate the extent to which achievement inconsistency between reading comprehension and mathematics occurs, whether achievement inconsistency differs for students with different background characteristics, and how this is associated with track recommendations.

Overall, while most students performed rather consistently across the subject domains reading comprehension and mathematics, about 20% of the students performed inconsistently (i.e. ≥ 1 *SD* difference). Students' achievement inconsistency played only a minor role in teachers' track recommendations. When students performed inconsistently, track

recommendations tended to be slightly lower on average. Although this effect suggests that teachers tend to give 'careful recommendations' in case of achievement inconsistency, the effect was so small, it seems that teachers mostly give 'aggregated recommendations' instead. Thus, teachers based their track recommendation mainly on an aggregation over students' achievement in the subject domains reading comprehension and mathematics instead of placing emphasis on the subject domain with the lowest or highest achievement. The random slopes for this effect were not significant, indicating that this effect does not vary between schools. Hence, across different schools, teachers tend to give these aggregated, somewhat cautious, track recommendations.

Students' Achievement Inconsistency

While the overall effect of inconsistency on track recommendations was small, achievement inconsistency primarily seemed to have an effect on track recommendations, when the inconsistency itself was moderate (1 to 2 *SD*) to large (≥ 2 *SD*). Students with high achievement inconsistency whose reading comprehension achievement was lower than their mathematics achievement received lower track recommendations than students with small achievement inconsistency (< 1 *SD*). For students with large achievement inconsistency whose mathematics achievement was lower than their reading comprehension achievement, the effect of inconsistency on track recommendations just failed to reach significance (p = .057). This may be explained by the fact those students in general performed lower across both subject domains than other students and, consequently, already received lower track recommendations. In this case, there may be no additional effect of achievement inconsistency on track recommendation beyond students' prior achievement. Teachers thus seemed to give more careful recommendations when the difference in achievement between the two subject domains is high, especially when students reading comprehension achievement is lower than their mathematics achievement. This could be due to the fact that the Dutch educational system is a tracked system, where students are allocated to one level of secondary education, and they take all their courses at that level. When students' achievement is highly inconsistent, it will be extremely difficult for students to pass tests in their weaker subject domains. These findings suggest that, for these students, it might be better if they were able to take different courses at different levels, as, for example, happens in Sweden (Le Métais, 2003). That way, they can follow all their courses at a level that matches their abilities and prevents them from being underchallenged in their stronger subject domains.

Considering Students' Achievement Inconsistency in Track Recommendations

In line with previous Dutch research, students' mathematics achievement appeared to be most decisive for track recommendations (e.g. Driessen et al., 2008; Smeets et al., 2014). These findings showed support for the idea that there is one subject domain most decisive for formulating a track recommendation ('a predominant subject domain'). However,

these effects were small, and the results of the random slopes analyses also indicated that there was a difference between schools in the extent to which the subject domains reading comprehension and mathematics were considered when formulating track recommendations. Some schools mostly relied on students' mathematics achievement, while other schools seemed to place more emphasis on students' reading comprehension achievement. Consequently, two students with similar achievement levels could receive different track recommendations at different schools. Such differences between schools with regard to the (extent of) information and criteria they consider when formulating track recommendations may be considered undesirable, as students in different tracks will have very different educational opportunities, partly due to limited opportunities of switching between different school tracks (Schnepf, 2002; van Rooijen et al., 2017). Differences between schools may be due to school characteristics or schools' student population. Further research concerning differences between schools regarding how teachers formulate the track recommendations is needed to understand how these differences may be explained and how they impact students' future school careers. It would also be interesting to examine whether and how guidelines for formulating track recommendations are used by schools. Such guidelines are available to schools in the region our data originates from, but it is unclear how these are being used.

Differences in Track Recommendations Based on Students' Background

In addition, as argued, it may be more difficult for teachers to formulate equal track recommendations for students given the varying achievement levels across subjects among different student groups. Regarding gender, the results indicated that the degree of inconsistency was similar for boys and girls. Yet, as expected, the achievement inconsistency of boys and girls was different in nature. Aligning with previous research (Gentrup & Rjosk, 2018; Hakkarainen et al., 2013; Plante et al., 2013; Uerz et al., 2004), it was found that boys with achievement inconsistency mostly performed higher in mathematics than reading comprehension, while girls with achievement inconsistency mostly performed higher in reading comprehension than mathematics. Despite these differences, boys and girls received comparable track recommendations, and the effect of inconsistency on track recommendations was equally strong for boys and girls. Therefore, no gender effects seemed to be present when teachers formulate track recommendations. This is in line with some research (e.g. Boone & Van Houtte, 2013; Driessen, 2005; Klapproth et al., 2013; Krolak-Schwerdt et al., 2017; Timmermans et al., 2018; Van Rooijen et al., 2016), whereas some other research did find gender differences in track recommendations with girls receiving higher track recommendations than boys (e.g. Dutch Inspectorate of Education, 2014; Jürges & Schneider, 2011; Timmermans et al., 2016).

Concerning SES, lower-SES students performed, on average, somewhat more inconsistently than higher-SES students. Students' SES was not related to the direction of the inconsistency, indicating that there was no difference between higher- and lower-SES students

in whether they performed higher in reading comprehension or mathematics. In line with previous research (e.g. Batruch et al., 2023; Boone & Van Houtte, 2013; Caro & Lehmann, 2009; Driessen et al., 2005, 2007; Feron et al., 2016; Klapproth et al., 2012; Korpershoek et al., 2016; Luyten & Bosker, 2004; Pit-ten Cate et al., 2016; Timmermans et al., 2013, 2016, 2018; Van Rooijen et al., 2017), lower-SES students received lower track recommendations than higher-SES students. This was mostly, but not completely, due to lower achievement of lower-SES students. Achievement inconsistency seemed to have a slightly stronger impact for higher-SES students than for lower-SES students. That is, teachers seemed to formulate somewhat more careful track recommendations for higher-SES students with achievement inconsistency than for lower-SES students with achievement inconsistency. This could be due to the fact that, on average, track recommendations are already higher for higher-SES students than for lower-SES students. Generally, teachers need to choose between the higher secondary tracks for higher-SES students. Consequently, when higher-SES students have achievement inconsistency, teachers may perhaps not choose the highest track, but tend toward the second-highest track. Although it was a small effect, it suggests that students with similar achievement inconsistency, but different SES may receive different track recommendations. Further research concerning these differences in track recommendations is needed to understand how these differences have an impact on students' future school career.

Limitations and Future Research

In interpreting the results of the present study, a few limitations need to be considered. First, in the present study, SES was measured using students' six-digit postal code. While these six-digit postal codes are, on average, only shared by a small number of households and are therefore considered to be an accurate impression of the SES of those households (Deckers et al., 2016; Guhn et al., 2010; van Hattem et al., 2009), they are not a measure of each individual household. Besides that, the six-digit classification contained missing values. For these missing values, the five-digit, and to a very small extent the four-digit postal codes were used which are less precise classifications. Therefore, these results need to be interpreted with some caution.

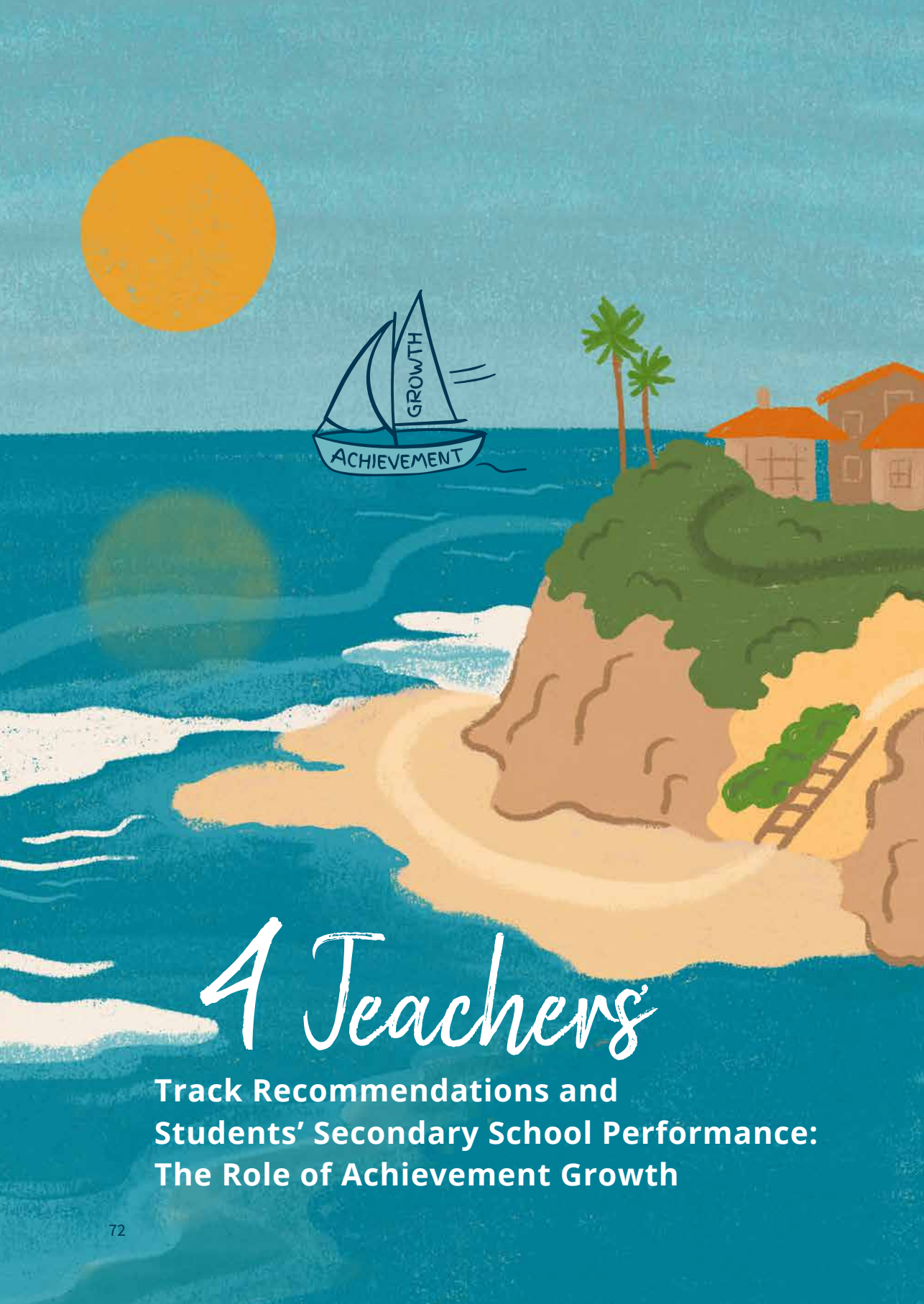
Second, the data was obtained from students from a large city in the Netherlands which might affect the generalisability of the results of the present study. Track recommendations can be formulated in different ways in different regions. Regions may, for example, differ in whether or not they allow combined track recommendations of two adjacent tracks. Moreover, results may also be different in other countries with different educational systems.

Third, the focus of the present study was not on the actual track placement of students, because these initial recommendations reflect primarily how teachers formulate a track recommendation, without interference of, for example, the results of a standardised school leavers' test. Therefore, the initial track recommendations primarily reflect teachers' inferences about students' potential behaviour or achievement.

For future research, it would be interesting to examine the extent to which achievement inconsistency has an impact on students' secondary school career. Although achievement inconsistency only had a minor effect on track recommendations, it is unknown how inconsistent achievement affects students' future educational success. It could be that some students will be hindered by their subject domain with the lowest achievement, suggesting that a more careful recommendation may be suitable. In addition, it may also matter in which subject domain students with inconsistent achievement show the lowest achievement. It might be that lower achievement in reading comprehension may have more harmful effects on achievement in other subject domains, as most subject typically rely heavily on comprehension of written texts, compared to lower achievement in mathematics. If so, careful recommendations may be more warranted in case of inconsistencies characterized by lower achievement in reading comprehension. However, prior research suggests that students who were placed in higher tracks than expected based on their achievement, usually were successful in that track (Inspectie van het Onderwijs, 2014). If that also applies in case of achievement inconsistency, then getting 'the benefit of the doubt', that is, a track recommendation based on the subject domain with the highest achievement, seems beneficial for students' future school careers. Furthermore, it would be interesting for future research to include other student characteristics that may be predictive of students' future school success, such as motivation, development, work habits or classroom behaviour (Feron et al., 2016; Klapproth et al., 2012; Oomens et al., 2019), as well to examine the additional information such characteristics provide for formulating track recommendations.

Conclusion

The present study highlights the interplay between students' achievement inconsistency, SES, and track recommendations. It contributes to the knowledge base on how teachers formulate track recommendations by studying the occurrence and effects of inconsistencies in achievement. The findings of the present study suggested that a tracked educational system, in which students follow all their courses at the same level, may not be appropriate for the rather substantial group of students whose achievement differs between subject domains (e.g. about 20% of the students showed a rather substantial inconsistency between subject domains). An educational system which allows for intrapersonal differences in abilities may potentially provide these students with a more suitable learning environment. Moreover, findings also indicated differences between schools in how track recommendations were formulated. Thereby, the findings also suggested a need for clearer guidelines on how to weigh different achievement indicators in students track recommendations to create equal opportunities for all students.



4 Teachers

Track Recommendations and Students' Secondary School Performance: The Role of Achievement Growth

Abstract

In tracked educational systems, teachers' track recommendations at the end of primary education are used to refer students to the secondary school track in which they have the best chance to realise their potential. Teachers mostly base their track recommendations on students' most recent achievement. However, achievement growth during primary education may also be important for track recommendations. Therefore, the aim of this study was to examine the potential significance of students' achievement growth for teachers' track recommendations. The sample consisted of 4,189 Grade 6 students from 102 Dutch primary schools. Data were analysed using multilevel Latent Growth Curve Analyses (LGCA). Results showed that achievement growth differs based on students' gender and SES. Teachers did not consider achievement growth in their track recommendations. As such, SES and gender differences in growth curves did not translate into different track recommendations. Additionally, students' SES, but not gender, showed a small direct effect on track recommendations. Achievement growth was not predictive of secondary school performance. Overall, these findings indicated that teachers did not consider students' achievement growth when formulating track recommendations, nor it is predictive of students' secondary school success.

Keywords:

equal educational opportunities, transition from primary to secondary education, school track recommendation, secondary school success, SES, gender, achievement growth, reading comprehension, mathematics, teacher judgement

This chapter is based on: van Leest, A., Hornstra, L., van Tartwijk, J., & van de Pol, J. (2024). *Teachers' track recommendations and students' secondary school performance:* The role of achievement growth. [Manuscript under review].

Acknowledgement of author contributions: All authors contributed to the design of the study. AvL collected the data and prepared the data for analyses; AvL analysed the data with help from LH; AvL drafted the manuscript; All authors contributed to critical revision of the manuscript; LH, JvdP, and JvT supervised the study.

Introduction

In tracked educational systems, teachers' track recommendations at the end of primary education include teachers' expectations about students' potential achievement to be able to allocate students to a secondary school track in which they have the best chances to realise their potential (Glock et al., 2012; van Leest et al., 2024). Although prior research has indicated that these track recommendations are primarily based on students' (most recent) achievement in primary education (Klapproth & Fischer, 2019; Timmermans, Kuyper, et al., 2015; van Leest et al., 2020), it may be important to consider achievement growth as well, as it may help teachers to predict at what rate students develop their achievement in the future.

Additionally, there may be differences in achievement growth related to students' background characteristics, such as gender and socioeconomic status (SES) (Klapproth & Fischer, 2019). For example, many lower-SES students start primary school with lower initial achievement than higher-SES students (Helbling et al., 2019; Zumbuehl & Dillingh, 2020). Some of these lower-SES students may catch up with their higher-SES peers during primary school, implying faster achievement growth. If teachers consider achievement growth in their track recommendations, differences in achievement growth between students with different background characteristics can affect track recommendations of different student groups. It is therefore important to understand how achievement growth differs across groups and how this, in turn, may be related to track recommendations. To our knowledge, research examining the extent to which teachers consider students' achievement growth in track recommendations, and how achievement growth relates to students' secondary school performance is scarce.

Students' Achievement Growth

In the present study, *achievement growth* is used to refer to the academic progress a student makes over a period of time, assessed by standardised test scores. Within a group of students, three main types of growth patterns can be distinguished: (a) the *compensatory effect*, which describes a pattern where initially lower-performing students show higher growth rates than initially higher-performing students (Aarnoutse & van Leeuwe, 2000; Salaschek et al., 2014); (2) the *linear effect*, which describes a pattern where students' growth rate does not differ based on students' initial achievement, that is, there is no difference in terms of growth rate between initially lower- and higher-performing students (Kuhfeld & Soland, 2021); and finally (3) the *Matthew effec*t, which describes a pattern of accumulated advantage where initially higher-performing students also have a higher growth rate, while linitially lower-performing students start with lower achievement, have a lower growth rate, and will end up at a lower achievement level than initially higher-performing students (Crosnoe et al., 2010; Helbling et al., 2019; Salaschek et al., 2014; Shaywitz et al., 1995). Empirical research indicated that all three patterns can be found among students, and that

these patterns appear to sometimes differ across school subjects (Caro, Lenkeit, et al., 2009; Curby et al., 2009; Salaschek et al., 2014). For example, for mathematics, the compensatory and Matthew effect patterns have been found among primary school students, and for (word) reading, the Matthew effect has been found.

Track Recommendations for Students With Different Achievement Growth

Teachers' track recommendations are primarily based on students' most recent achievement in primary school (Jürges & Schneider, 2011; Timmermans, Kuyper, et al., 2015; van Leest et al., 2020, 2024). However, students' most recent achievement levels may not necessarily give an optimal reflection of their potential educational performance in secondary education, because students may have had different achievement growth (Caro, Lenkeit, et al., 2009; Curby et al., 2009; Helbling et al., 2019). It may be difficult for teachers to include achievement growth in their track recommendations, as students' achievement growth, especially growth over a longer period of time (i.e. several years), may be more difficult to record, reflect, or interpret than students' achievement at one single time point (Curby et al., 2009). Moreover, achievement growth may not be linear and may fluctuate at different rates over a longer period (Kuhfeld & Soland, 2021). In these cases, students' achievement growth may not directly point to a particular secondary school track, and, therefore, teachers need to decide themselves which track is most optimal for a student.

When looking at students' achievement over time, three factors can be distinguished: (a) students' initial achievement level, (b) students' most recent achievement level, and (c) students' achievement growth rate. Thus far, scholars researching the effects of achievement growth on track recommendations typically considered students' initial achievement levels as well as their achievement growth (Caro, Lenkeit, et al., 2009; Crosnoe et al., 2010; McClelland et al., 2006). However, this may lead to an overestimation of the effect of achievement growth. For example, suppose there are two students with similar achievement levels at the end of primary education and comparable track recommendations but different initial achievement levels: one student had lower initial achievement but showed faster achievement growth, while the other had higher initial achievement but showed slower achievement growth. When taking into account students' initial achievement and their growth, this will probably result in different outcomes for both students. However, if students' achievement at the end of primary education is included as a predictor of track recommendation in addition to their achievement growth rather than their initial achievement level, it may show that teachers do not consider achievement growth in their track recommendations, but instead rely on students' final achievement. Hence, the unique value of achievement growth can more accurately be established when students' most recent achievement (rather than initial achievement) and achievement growth are considered. Therefore, in the present study, the role of students' most recent achievement in the final year of primary education (sixth grade) and their achievement growth in the three years

prior are considered as predictors of track recommendations. To our knowledge, this study is the first which examines the role of achievement growth and takes students' most recent achievement as starting point.

Moreover, research on the relation between students' achievement growth and track recommendations is scarce. Only two German studies are available that both considered achievement growth on top of students' initial achievement (Caro, Lenkeit, et al., 2009; Klapproth & Fischer, 2019). The study of Klapproth and Fischer (2019) used vignettes to investigate whether preservice teachers considered students' achievement growth over half a school year in Grade Point Average (GPA) when formulating track recommendations, whereas Caro, Lenkeit, et al. (2009) considered mathematics achievement growth from Grades 4 to 6, derived from a longitudinal study. Both studies showed that teachers took students' growth into account when formulating track recommendations. Students who improved their GPA or who developed their mathematic skills more rapidly were more likely to receive a higher track recommendation, even when there was accounted for mean (initial) achievement levels. Although Caro, Lenkeit, et al. (2009) showed that the impact of students' growth in mathematics on track recommendations was significant, the impact of students' initial mathematics achievement was much larger. It may be that when looking at students' most recent achievement, there is no longer an additional effect of achievement growth. Hence, to fully understand the role of achievement growth for teachers' track recommendations, students' most recent achievement levels need to be considered.

Additionally, Caro, Lenkeit, et al. (2009) only included mathematics achievement as predictor of track recommendations. Because previous research has shown that there are achievement differences among students across subject domains (Luyten, 1998; van Leest et al., 2024), as well as different growth patterns across different subject domains (Caro, Lenkeit, et al., 2009; Curby et al., 2009; Salaschek et al., 2014), and that teachers consider multiple subject domains when formulating their track recommendations (van Leest et al., 2020), it may be worthwhile to examine achievement growth in multiple school subjects.

Achievement Growth and Students' Gender and SES

Furthermore, there may be differences in achievement growth related to students' background characteristics, such as gender and socioeconomic status (SES) (Klapproth & Fischer, 2019). However, most studies did not address students' achievement growth over multiple years, and even if it was included, there were mixed findings. Research on gender differences in achievement growth also showed mixed results. For example, according to Entwisle et al. (1994) there were different growth curves for mathematics achievement in primary school between boys and girls with boys showing faster achievement growth than girls, while Caro, Lenkeit, et al. (2009) reported higher growth rates for girls. However, Leahey and Guo (2001) found that boys and girls showed comparable mathematics growth curves, with small gender differences in favour of boys only emerging at the end of secondary school. For reading comprehension, girls showed higher growth rates than boys (Moon & Hofferth, 2016; Sax, 2009). Prior research suggested that these differences in growth curves

could result in differences in track recommendations between boys and girls. For example, Klapproth and Fischer (2019) reported that boys had more chances of receiving a higher track recommendation than girls if their achievements improved.

According to prior research, students' SES is mostly indirectly related to track recommendations via its effect on achievement in reading comprehension and mathematics (Caro, Lenkeit, et al., 2009; van Leest et al., 2020) and growth (Caro, Lenkeit, et al., 2009). In general, research showed a widening achievement gap based on SES: lower-SES students developed their skills more slowly than higher-SES students (M. Becker et al., 2006; R. Becker & Schubert, 2006). On top of the already lower initial achievement for lower-SES students, this results in negative cumulative effects for lower-SES students (Helbling et al., 2019). However, some of these lower-SES students may catch up with their higher-SES peers during primary school. If teachers consider achievement growth in their track recommendations, differences in achievement growth can affect track recommendations of students with different backgrounds (Boone et al., 2018; Timmermans et al., 2018; van Leest et al., 2024). It is therefore important to understand how achievement growth differs across groups and how this, in turn, may be related to track recommendations.

Predicting Performance in Secondary Education

Research has shown that achievement in primary education is strongly related to achievement in secondary education (de Boer et al., 2010; Poncelet & Metis Associates, 2004; van Rooijen et al., 2016). Likewise, it could also be that achievement growth in primary education is predictive of continued growth in achievement in secondary education. That is, a student who acquires new knowledge or skills quickly may have faster achievement growth in both primary and secondary education. This could suggest that it is important to take students' achievement growth into account when formulating track recommendations. On the other hand, tracking may also affect students' achievement growth in secondary education. Moreover, being allocated to a particular secondary school track may limit the extent to which achievement growth in primary education is predictive of educational success in secondary education, not only because students may perform according to the track they are in (i.e. self-fulfilling prophecy effects; Brophy, 1983; De Boer et al., 2010; Jussim & Harber, 2005), but also because it is rather difficult to switch tracks, especially upwards (Brunello & Checchi, 2007; Feron et al., 2016; Gamoran, 2009).

Present Study

Altogether, research on the role of students' achievement growth in teachers' track recommendations is scarce. Given this scarcity, it remains unclear how students' growth in achievement during primary education affects track recommendations in general and how it affect track recommendations of students with different background characteristics. Furthermore, it is unknown to what extent students' achievement growth in primary education is predictive of their secondary school performance. To answer these questions, the shape of students' achievement growth curves during primary education was examined

(Research Question [RQ] 1a) and how these growth curves differed by students' gender and SES (RQ1b), as students from different backgrounds may have different growth patterns due to differences in achievement. If teachers take achievement growth into account when formulating track recommendations, differences in achievement growth between groups of students may result in differences in track recommendations between students with different backgrounds, even in case of similar achievement at the end of primary school. Therefore, it was examined to what extent students' achievement growth in primary education is reflected in teachers' track recommendations, on top of students' most recent achievement at the end of primary school (RQ2). Together, the answers to RQ1b and RQ2 determine if there are variations in growth curves between student groups and how this impacts track their recommendations. Finally, it was examined to what the extent students' achievement growth in primary education adds to the prediction of students' (achievement) level(s) in secondary education, on top of their most recent achievement in primary school. To answer this question, the role of students' track placement was taken into account (RQ3).

Method

Sample

We used an existing longitudinal dataset which was part of a larger dataset on students' educational development across the transition from primary to secondary education. This dataset contained data from an online student monitoring platform containing different kinds of information about students, such as students' educational achievement in primary school and background characteristics (van Leest et al., 2020). Data from schools who gave consent for using their data were anonymised by an institution representing the schools with access to this monitoring platform. The anonymous data were shared with the research team. The data were gathered in a large city in the Netherlands. More detailed information about this dataset can be found in Van Leest et al. (2020). The sample consisted of 4,189 Grade 6 students (50.1% girls) from 102 primary schools in a large city in the Netherlands. Two cohorts of students were included: students who were in Grade 6 in the academic years 2014-2015 (51.6 %) and 2015-2016 (48.4%)[10]. The dataset included their achievement in Grades 3 to 6 of primary education (the last four grades of primary education), and Grades 7 and 8 of secondary education (the first two grades of secondary education). The sample is representative of a large Dutch city in terms of SES with 16.3% of the students having a lower SES (< -1 *SD*), 65.3% having a middle SES, and 18.0% of having a high erSES (> +1 *SD*) background (Centraal Bureau voor de Statistiek, 2022; Leidelmeijer & Burema, 2022).

Table 1

Secondary school track		Grade 7[b]	Grade 8
Single track	Two adjacent tracks[11]		
(1) practical training		1	2
	(1) practical training / (2) basic pre-vocational	1.5	2.5
(2) basic pre-vocational		2	3
	(2) basic pre-vocational / (3) middle pre-vocational	2.5	3.5
(3) middle pre-vocational		3	4
	(3) middle pre-vocational / (4) theoretical pre-vocational	3.5	4.5
(4) theoretical pre-vocational		4	5
	(4) theoretical pre-vocational / (5) senior general	4.5	5.5
(5) senior general		5	6
	(5) senior general / (6) pre-university	5.5	6.5
(6) pre-university		6	7

Note. Grade 7 = first year of secondary education, i.e. students' track placement;
Grade 8 = second year of secondary education.
[a]Based on an adapted version (van Aarsen et al., 2013) of the "educational ladder" (Bosker et al., 1985)
[b]The track recommendation is coded as the same as Grade 7

The Educational Ladder[a]

10 In 2014, there was a policy reform of the Dutch educational system regarding the track recommendation procedures. The most important change of this reform was a changed time schedule, that resulted in not having results of the standardised school leavers' test available to teachers when formulating a track recommendation (Korpershoek et al., 2016; Oomens et al., 2019). This revised tracking recommendation procedure was followed in both cohorts from the present study.

11 In some Dutch regions, combined track recommendations or secondary school tracks of adjacent tracks are possible.

Measures

Track Recommendation. In the Netherlands, at the end of primary education (sixth grade), teachers formulate a recommendation to refer students to one single track of secondary education.[11] In the dataset we used, these track recommendations were recoded based on an adapted version (van Aarsen et al., 2013) of the "educational ladder" (Bosker et al., 1985), resulting in a scale from 1 to 6 (see Table 1). This scale corresponds to the six secondary school tracks, from lowest to highest track: (1) practical training, (2) basic pre-vocational secondary education, (3) middle pre-vocational secondary education, (4) theoretical pre-vocational education, (5) senior general secondary education, and (6) pre-university education. In this study, track recommendation was considered a continuous variable (cf. Norman, 2010; Sullivan & Artino, 2013).

Primary School Achievement. Standardised test scores in reading comprehension and mathematics from Grades 3 to 6 were included. They were derived from the Monitoring and Evaluation system developed by the Dutch institute for educational measurement (Cito). Schools typically administer the math tests twice a year (middle and end of the school year) and the reading comprehension test once a year (middle of the school year). When students had more than one test score in a single period, the last test score was used because of the possibility of retesting.[12] Students' initial achievement was the first test score on reading comprehension or mathematics in Grade 3. Students' most recent achievement was the test score on reading comprehension or mathematics in Grade 6. Prior research provided support for high validity, and high internal consistency ($\alpha > .80$) of these standardised tests (Feenstra et al., 2010; Janssen et al., 2010).

Track Placement. Track placement refers to the track where students were assigned to in their first year of secondary education, coded on the educational ladder (see Table 1). Track placement was considered as a continuous variable (cf. Norman, 2010; Sullivan & Artino, 2013).

Secondary School Performance. Secondary school performance consisted of reading comprehension and mathematics achievement, and students' educational level.
Reading Comprehension and Mathematics Test. Nationwide standardised reading comprehension and mathematics tests, developed by Cito, were used. These tests were adminis-

11 In the city in which the data was collected, a combination of two adjacent tracks for the track recommendations is not allowed.

12 Retesting is done when students perform much lower than expected for a specific grade (Dyslexie Centraal, 2023a, 2023b). Most of the time, students get a test from a lower grade to see how they perform on that test.

tered in seventh grade between April and June. The reliability of the reading comprehension and mathematics tests is generally very high (α > .92 or above) (Kuyper et al., 2014; Zijsling et al., 2017).

Level of Secondary Education. Students' educational level in eighth grade was reported by the schools. To account for grade retention or skipping a grade, this variable was coded on the educational ladder (see Table 1). Level of education was considered as a continuous variable (cf. Norman, 2010; Sullivan & Artino, 2013).

Background Characteristics. Students' gender and SES, measured in sixth grade (final year of primary education), were included as students' background characteristics.
Gender. A dichotomous variable was created for students' gender with boys forming the reference group.

Socioeconomic Status (SES). Students' six-digit postal code was used as an approximation of students' SES (e.g. Danesh et al., 1999). In Dutch cities, the six-digit postal code is shared by only 15 to 20 households (Deckers et al., 2016; Guhn et al., 2010; van Hattem et al., 2009). Three indicators, provided by Statistics Netherlands (CBS), were recoded into a factor score using principal component analysis (PCA) to compose the SES variable: (a) the mean real estate value, (b) the most recent mean household income after tax, and (c) the number of people who have social welfare benefits and/or are unemployed (van Leeuwen, 2019). When the six-digit postal code contained missing data (38.6% of the total sample is complete), SES was estimated based on the five-digit postal codes (85.7% of the missing values) and the four-digit postal codes (2.8% of the missing values), using the same indicators as the six-digit postal code classification. Students' SES was a continuous variable ranging from -2.42 to 3.83. In the analyses, students' SES was considered as a continuous variable (Hox et al., 2018; Robitzsch, 2020).

Data Analyses

The data were analysed using Mplus Version 8.4 Mac (Muthén & Muthén, 2017). Two-level complex latent growth curve analyses (LGCA) were estimated to model students' achievement development as a function of time (Stoel & Galindo-Garre, 2011) by including a hierarchical level of measurements (Level 1) within students (Level 2). Because most primary schools (59.3%) had only one Grade 6 class, there was no distinction between the class and the school level (cf., Timmermans et al., 2016, 2019). Therefore, only the school level was taken into account by a correction of the standard errors (i.e. using "type is complex" in Mplus).

With the LGCAs, first, it was examined which growth curves fitted the data best (RQ1). Thereto, an intercept (first achievement in Grade 3) and a slope (growth over four primary school years) for students' reading comprehension and mathematics were estimated for each student. In total, four reading comprehension test scores (one per year) and seven

mathematics test scores (two per year) were used to estimate students' growth curve models. Besides linear growth patterns, a quadratic growth term was also estimated to examine potential curvilinear growth patterns. Research has shown that a linear achievement growth model can be estimated with at least three data points, but the precision of the parameter estimates improves when including more data points (Raudenbush & Liu, 2000; Rogosa et al., 1982). Analyses were performed separately for each domain to avoid interaction effects. Next, students' gender and SES were added to investigate whether there were differences in growth curves among certain groups of students.

To answer RQ2 and RQ3, a series of models with the same predictors (i.e. students' prior achievement, achievement growth, and background characteristics), but different outcome variables (i.e. track recommendation for RQ2, and secondary school performance for RQ3) were estimated. The analyses were performed combined for both domains, to mimic teachers' real-life situations as closely as possible. Students' test scores in Grade 6 were used as intercepts to be able to examine whether growth uniquely explained variance in track recommendations (RQ2) or secondary school performance (RQ3) beyond students' most recent achievement. For both sets of outcome variables, first, an empty model with only the dependent variable was estimated (Model 0). Next, students' most recent achievement (Model 1) and achievement growth (Model 2) were added. Next, students' background characteristics gender and SES were added as fixed effects to examine whether students' gender and SES were predictive of the outcome variables beyond students' primary school achievement (Model 3). Covariances between students' background characteristics and the intercepts and slopes of reading comprehension and mathematics were included. For RQ3, the results of Models 2 and 3 indicate whether there were effects between secondary school tracks. However, the track students are allocated to largely determines their future achievement, due to the path dependency of the Dutch educational system. Therefore, for RQ3, additional models (Model 2b and Model 3b) were estimated which included track placement as a predictor of secondary school performance to investigate whether effects within a secondary school track existed.

Explained variance (R^2) was calculated for all models (cf. Snijders & Bosker, 2012). Effect sizes were based on the standardised regression coefficients (STDYX) with 0.2, 0.5, and 0.8 as indicative of small, medium, and large effect sizes, respectively (Cohen, 1988).

We used the Akaike Information Criterion (AIC; Akaike, 1987), Bayesian Information Criterion (BIC; Raftery, 1993), Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), Comparative Fit Index (CFI; Bentler, 1990), Tucker Lewis Index (TLI; Tucker & Lewis, 1973), and Standardised Root Mean Square Residual (SRMR; Hu & Bentler, 1995) to compare the model fit of the estimated models. A lower value on AIC or BIC indicates a better fit (Fabozzi et al., 2014; Kass & Raftery, 1995). A value of RMSEA between .05 and .08 is generally considered as a good fit, with a value close to zero as excellent fit (Awang, 2012; Byrne, 1994). For CFI and TLI, values between .95 and 1.00 indicate a good fit, with a value close to one as excellent fit (Hu & Bentler, 1999; Schumacker & Lomax, 2010). Finally, for SRMR a

value of zero indicates perfect fit, whereas a value less than .08, is generally considered a good fit (Byrne, 1994; Hu & Bentler, 1999).

The dataset contained missing data. For RQ3, only use a subset of the data (*N* = 333 students) could be used, because only for those students secondary school data was available (see Table 2 for the N of all variables). Moreover, some schools did not administer all standardised tests throughout primary education since schools were not obliged to administer the tests. Especially the test scores in Grades 3 and 6 were often missing (respectively about 54% and 33%). For the other tests, only 4 to 9% was missing.

Table 2

	N	*D*	*SD*	Min.	Max.	%
Track recommendation	4189	4.68	1.29	1	6	
Primary school achievement						
Reading comprehension test						
Grade 3	1910	27.60	14.65	-20	101	
Grade 4	3892	36.19	14.26	-19	121	
Grade 5	4040	51.31	16.26	-19	122	
Grade 6	2777	66.40	19.41	2	154	
Mathematics test						
Grade 3 – 1st test	1962	74.15	14.731	10	123	
Grade 3 – 2nd test	1928	81.92	14.162	0	132	
Grade 4 – 1st test	3979	88.55	13.843	15	150	
Grade 4 – 2nd test	3856	94.28	12.723	8	143	
Grade 5 – 1st test	4011	103.00	13.285	14	160	
Grade 5 – 2nd test	3802	109.29	13.97	22	164	
Grade 6 – 1st test	2881	116.69	12.03	21	168	
Background characteristics						
Gender – boys	2089					49.87%
Gender – girls	2100					50.13%
Socioeconomic status (SES)	4178	0.40	0.91	-2.42	3.83	
Secondary school performance						
Reading comprehension test	316	185.10	47.27	70	323	
Mathematics test	171	160.82	25.71	110	240	
Track placement	331	5.21	1.01	2	6	
Level of education – Grade 8	324	6.21	1.02	3	7	

Descriptive Statistics of Track Recommendation, Primary School Achievement, Student Background Characteristics, and Secondary School Performance

Results

The descriptive statistics of the variables are presented in Table 2. The correlations between students' achievement in primary and secondary education, track recommendation, level of secondary education, gender, and SES are presented in Table 3. Overall, positive correlations were found with small to high effect sizes.

Table 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Track																	
2. Read PE Grade 3	.61***																
3. Read PE Grade 4	.69***	.71***															
4. Read PE Grade 5	.75***	.66***	.71***														
5. Read PE Grade 6	.75***	.61***	.66***	.68***													
6. Math PE Grade 3 – 1st	.71***	.52***	.54***	.55***	.50***												
7. Math PE Grade 3 – 2nd	.72***	.50***	.53***	.57***	.50***	.83***											
8. Math PE Grade 4 – 1st	.73***	.47***	.58***	.58***	.51***	.80***	.82***										
9. Math PE Grade 4 – 2nd	.76***	.48***	.58***	.60***	.53***	.79***	.81***	.86***									
10. Math PE Grade 5 – 1st	.78***	.47***	.55***	.62***	.54***	.77***	.79***	.84***	.86***								
11. Math PE Grade 5 – 2nd	.79***	.47***	.53***	.60***	.57***	.74***	.78***	.81***	.84***	.87***							
12. Math PE Grade 6 – 1st	.77***	.40***	.49***	.55***	.63***	.70***	.71***	.74***	.76***	.79***	.83***						
13. Gender	-.01	.16***	.07***	.06***	.07***	-.18***	-.19***	-.17***	-.16***	-.17***	-.14***	-.11***					
14. SES	.30***	.17***	.21***	.24***	.19***	.21***	.19***	.21***	.24***	.20***	.23***	.17***	-.01				
15. Read SE Grade 7	.42***	.27***	.35***	.35***	.34***	.24**	.20*	.23***	.23***	.28***	.28***	.26***	.14*	.07			
16. Math SE Grade 7	.76***	.56***	.57***	.62***	.64***	.69***	.71***	.66***	.66***	.77***	.78***	.70***	.06	.32***	.74***		
17. Track placement	.95***	.52***	.65***	.68***	.69***	.62***	.67***	.66***	.68***	.71***	.75***	.69***	-.02	.33***	.51***	.79***	
18. Level SE Grade 8	.91***	.53***	.63***	.66***	.68***	.61***	.65***	.63***	.66***	.69***	.74***	.68***	-.01	.34***	.53***	.79***	.95***

Note. Read = Reading comprehension; SES = socioeconomic status; PE = primary education; SE = secondary education.
*p < .05; **p < .01; ***p < .001.

Correlations Between Track Recommendation, Primary School Achievement, Background Characteristics, and Secondary School Performance

Students' Achievement Growth (RQ1)

First, the shape of students' achievement growth curves with students initial achievement in Grade 3 as the starting point was analysed (RQ1a). The results of model fitting are summarised in Tables 4 (reading comprehension) and 5 (mathematics). Overall, the results indicated that a linear growth curve fitted the data better compared to a curvilinear growth curve. Yet, for reading comprehension, the model with all four achievement measurements still did not fit the data well. The modification indices indicated that the first reading comprehension measurement did not fit the linear curve. This could be caused by greater variation in reading comprehension when students are younger, for example, due to differences in technical reading skills (Aarnoutse, 2017; Schijf, 2009; Verhoeven & van Leeuwe, 2008). Therefore, in further analyses, the factor loading of the first reading comprehension measurement was freely estimated, which resulted in good model fit. The final models for reading comprehension (Table 4, Model 1b) and mathematics (Table 5, Model 1a) fitted the data well.

The variance components of the initial achievement levels and the growth factors were significant, indicating that there was significant variation between students regarding their slopes and intercepts in reading comprehension and mathematics (as illustrated in Figures 1 and 2). The figures show observed data from a small sample of students as well as the estimated linear growth curve of the entire group of students. Both figures indicated that the achievement growth of many students fluctuates, is somewhat unsteady and non-linear.

"Bij kinderen die veel groei laten zien, is het niet altijd gegarandeerd dat het doorzet in het vo. Bij een aantal wel, maar niet bij allemaal. We zijn daarom voorzichtig met adviseren. We krijgen rapporten terug van het vo en we zien dan vaak dat bij leerlingen bij wie we twijfelden en die we naar een hoger niveau hebben laten gaan, bijna altijd weer terug zakken naar een lager niveau."

Table 4

	Model 1a: Linear growth curve				Model 1b: Curvilinear growth curve				Model 1c: Linear growth curve with Grade 3 test score freely estimated			
			95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects												
Intercepts												
Reading comprehension intercept	24.43***	0.61	23.25	25.62	27.71***	0.60	26.53	28.89	20.80***	0.61	19.61	21.98
Reading comprehension slope	13.34***	0.25	12.88	13.85	7.12***	0.48	6.19	8.08	15.21***	0.21	14.81	15.61
Reading comprehension quadratic slope					2.04***	0.15	1.74	2.33				
Reading comprehension intercept WITH												
Reading comprehension slope	15.32***	2.49	10.45	20.19	-15.64	15.10	-45.24	13.95	8.75***	2.32	4.21	13.28
Reading comprehension quadratic slope					7.70	3.97	-0.09	15.48				
Model fit												
AIC	100,107.29				99,648.32				99,511.89			
BIC	100,164.15				99,730.44				99,575.06			
RMSEA	0.147				0.119				0.023			
CFI	0.99				1.00				1.00			
TLI	0.99				0.99				1.00			
SRMR	0.085				0.032				0.035			
Number of schools	4,093				4,093				4,093			
Number of students	100				100				100			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Unstandardised Estimates of Different Growth Curve Models for Students' Reading Comprehension Achievement in Primary Education

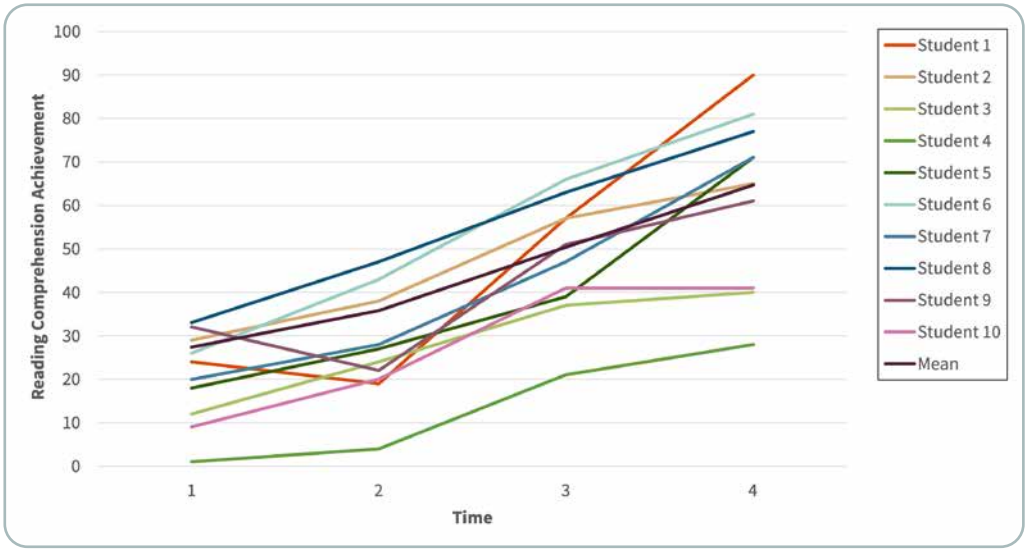
Table 5

	Model 1a: Linear growth curve				Model 1b: Curvilinear growth curve			
			95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects								
Intercepts								
Mathematics intercept	74.87***	0.62	73.67	76.08	74.60***	0.62	73.39	75.82
Mathematics slope	13.94***	0.62	73.67	14.21	14.33***	0.41	13.52	15.14
Mathematics quadratic slope					-0.12	0.12	-0.36	0.12
Mathematics intercept WITH								
Mathematics slope	-12.11***	1.27	-14.61	-9.62	-23.08***	5.81	-34.47	-11.69
Mathematics quadratic slope					3.30*	1.58	0.21	6.40
Model fit								
AIC	154,590.15				154,525.49			
BIC	154,665.99				154,626.61			
RMSEA	0.054				0.059			
CFI	0.99				0.99			
TLI	0.99				0.99			
SRMR	0.086				0.077			
Number of schools	4,106				4,106			
Number of students	101				101			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

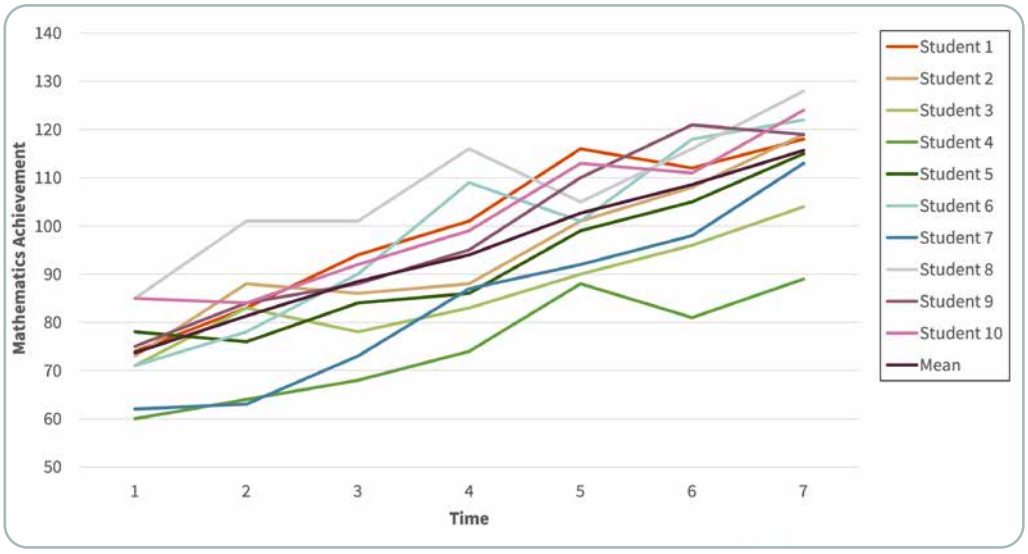
Unstandardised Estimates of Different Growth Curve Models for Students' Mathematics Achievement in Primary Education

Figure 1



Students' Reading Comprehension Growth in Primary School
(Sample Data From 10 Students and the Estimated Growth Curve)

Figure 2



Students' Mathematics Growth in Primary School
(Sample Data From 10 Students and the Estimated Growth Curve)

The results of the growth curve models (Model 1 of Tables 6 and 7) indicated that higher initial reading comprehension achievement was associated with faster achievement growth (suggesting a 'Matthew effect'), while lower initial mathematics achievement was associated with faster achievement growth ('compensatory effect').

By adding students' background characteristics gender and SES to the growth model, it was investigated whether students' initial achievement and their growth rates differed based on students' gender and SES (RQ1b). For reading comprehension, the results of Model 2 (Table 6) indicated a small significant positive effect of gender on the intercept (b^* = .15, p < .001) and a small negative effect of gender on the slope (b^* = -.12, p < .001), indicating that girls had higher initial reading comprehension achievement than boys, but boys had faster reading comprehension growth than girls. For mathematics achievement, it was the exact opposite. The results of Model 2 (s 7) indicated that boys had higher initial mathematics achievement than girls (b^* = -.19, p < .001), but girls had faster mathematics growth than boys (b^* = .14, p < .001). For SES, the results (Model 2 of Tables 6 and 7) indicated a small significant positive effect of SES on both intercepts ($b_{Reading}{}^*$ = .23, p < .001; $b_{Math}{}^*$ = .23, p < .001), indicating that higher-SES students had higher initial reading comprehension and mathematics achievement. Furthermore, for reading comprehension, it was found that higher-SES students also showed faster achievement growth ($b_{Reading}{}^*$ = .24, p < .001; $b_{Math}{}^*$ = -.02, p = .549). The models with gender and SES had a (slighty) better fit than the growth curve models without predictors.

Table 6

	Model 1: Growth curve				Model 2: + Gender + SES			
			95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects								
Intercepts								
Reading comprehension intercept	20.80***	0.61	19.61	21.98	17.82***	0.62	16.61	19.04
Reading comprehension slope	15.21***	0.21	14.81	21.98	15.26***	0.20	14.87	15.65
Reading comprehension intercept WITH								
Reading comprehension slope	8.75***	2.32	4.21	13.28	7.94***	2.13	3.76	12.11
Gender (girls)								
Reading comprehension intercept					3.43***	0.44	2.57	4.29
Reading comprehension slope					-0.63***	0.17	-0.96	-0.30
SES								
Reading comprehension intercept					2.99***	0.48	2.06	3.93
Reading comprehension slope					0.70***	0.13	0.44	0.96
Model fit								
AIC	99,511.89				99,044.03			
BIC	99,575.06				99,132.45			
RMSEA	0.023				0.023			
CFI	1.00				1.00			
TLI	1.00				1.00			
SRMR	0.035				0.030			
Number of schools	4,093				4,087			
Number of students	100				100			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Unstandardised Estimates of Growth Curve Models for Students' Reading Comprehension Achievement in Primary Education (RQ1)

Table 7

	Model 1: Growth curve				Model 2: + Gender + SES			
			95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects								
Intercepts								
Mathematics intercept	74.87***	0.62	73.67	76.08	76.11***	0.70	74.73	77.49
Mathematics slope	13.94***	0.14	13.67	14.21	13.60***	0.13	13.34	13.86
Mathematics intercept WITH								
Mathematics slope	-12.11***	1.27	-14.61	-9.62	-10.92***	1.29	-13.45	-8.38
Gender (girls)								
Mathematics intercept					-5.21***	0.51	-6.21	-4.21
Mathematics slope					0.71***	0.12	0.48	0.94
SES								
Mathematics intercept					3.43***	0.53	2.39	4.47
Mathematics slope					-0.06	0.10	-0.26	0.14
Model fit								
AIC	154,590.15				153,998.33			
BIC	154,665.99				154,099.42			
RMSEA	0.054				0.048			
CFI	0.99				0.99			
TLI	0.99				0.99			
SRMR	0.086				0.071			
Number of schools	4,106				4,100			
Number of students	101				101			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Unstandardised Estimates of Growth Curve Models for Students' Mathematics Achievement in Primary Education (RQ1)

Predictive Value of Achievement Growth for Track Recommendations (RQ2)

To examine whether and to what extent students' achievement growth was reflected in teachers' track recommendations on top of students' most recent achievement, students' most recent achievement was included as point of reference, and, as a result, the slope parameters must be interpreted inversely (RQ2). The results are presented in Table 8. Model 1 revealed that students' most recent reading comprehension and mathematics (i.e. the intercepts) were significantly positively related to students' track recommendations ($b_{Reading}$* = .63, p < .001; b_{Math}* = .35, p < .001). This indicates that students with higher recent reading comprehension and mathematics achievement received higher track recommendations. Model 2 showed that achievement growth (i.e. the slopes) in reading comprehension and mathematics was not significantly related to track recommendations ($b_{Reading}$* = .70, p = .252; b_{Math}* = -.39, p = .284), indicating that teachers do not consider students' growth in reading comprehension and mathematics in their track recommendations on top of students' most recent achievement. Whereas Model 1 showed a bad model fit, Model 2 showed a good model fit. Together these achievement variables explained 82.6% of the variance in track recommendations.

Students' background characteristics were added to the model to investigate whether students' gender and SES were predictive of students' track recommendations after accounting for most recent achievement and achievement growth. Results of Model 3 indicated no significant effect of gender or SES on track recommendations (b_{Gender}* = --.09, p = .235; b_{SES}* = .08, p = .131). This model had a good fit. All together these variables explained 83.4% of the variance in track recommendations.

Table 8

	Model 0: Empty				Model 1: Intercepts				Model 2: + Slopes				Model 3: + Gender + SES			
			95% CI				95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects																
Intercept	4.68***	0.07	4.55	4.81	-1.69***	0.24	-2.16	-1.21	-6.01*	3.07	-12.02	0.00	-5.13*	2.15	-9.34	-0.92
Reading comprehension intercept					0.06***	0.00	0.06	0.07	0.06***	0.01	0.04	0.07	0.06***	0.01	0.04	0.07
Reading comprehension slope									0.31	0.26	-0.20	0.82	0.28	0.22	-0.14	0.71
Mathematics intercept					0.04***	0.00	0.03	0.04	0.08**	0.03	0.01	0.14	0.07**	0.02	0.02	0.11
Mathematics slope									-0.20	0.19	-0.57	0.17	-0.19	0.16	-0.51	0.13
Gender (girls)													-0.23	0.19	-0.60	0.15
SES													0.11	0.07	-0.03	0.25
Variance																
R^2						.852				.826				.834		
Model fit																
AIC		14,026.82				295,850.92				258,541.31				274,819.17		
BIC		14,039.50				295,984.06				258,744.20				275,117.16		
RMSEA		0.000				0.272				0.047				0.044		
CFI		0.00				0.90				1.00				0.99		
TLI		1.00				0.90				1.00				0.99		
SRMR		0.000				0.784				0.054				0.047		
Number of schools		4,189				4,189				4,189				4,189		
Number of students		102				102				102				102		

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.

*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Unstandardised Estimates of Growth Curve Models 0 to 3 Predicting Track Recommendations With Reading Comprehension and Mathematics Achievement Growth (RQ2)

Predictive Value of Achievement Growth for Secondary School Performance (RQ3)

The third aim of the present study was to investigate the extent to which students' achievement growth predicted students' achievement levels in secondary education on top of students' most recent achievement in primary education (RQ3). The results are presented in Model 2 of Tables 9 (reading comprehension), 10 (mathematics) and 11 (level of education). In Model 3, it was investigated whether students' gender and SES predicted students' secondary school performance, on top of their achievement (growth) in primary education. For both models additional analyses were conducted including track placement to examine whether the effects still exist within the tracked system, as students may perform according to the track they are in. All models showed a good fit, except for Model 1 which only included the intercepts of the growth curve models.

For reading comprehension, Model 1 (Table 9) revealed that students' most recent achievement in primary education was significantly positively related to students' reading comprehension achievement in secondary school (b^* = .27, p = .002). Students with higher reading comprehension achievement at the end of primary education had higher reading comprehension achievement in secondary education. No significant effects for achievement growth were found (b^* = -.12, p = .272), indicating that students' achievement growth in reading comprehension during primary education did not predict their reading comprehension achievement in secondary education (Model 2a). However, the effect of students' most recent reading comprehension achievement became non-significant when students' track placement was added to the model (Model 2b; b^* = .03, p = .751). Within a track, students did not seem to differ in their primary school reading comprehension achievement. Model 3a showed a significant (direct) effect of students' gender on reading comprehension achievement on top of students' achievement (growth) in primary education (b^* = .13, p = .020). Girls performed higher on reading comprehension than boys. For SES, no significant effect was found on top of students' achievement (growth) in primary education (b^* = -.05, p = .407). However, the effect of SES became significant when students' track placement was added to the model (Model 3b; (b^* = -.12, p = .039). Within a track, lower-SES students performed lower on reading comprehension than higher-SES students. For gender, the results did not change when track placement was added. Within a track, students' reading comprehension achievement at the end of primary education did not predict students' reading comprehension achievement in secondary school.

For mathematics, Model 1 (10) revealed that students' most recent mathematics achievement in primary education was significantly positively related to mathematics achievement in secondary school (b^* = .53, p < .001): higher mathematics achievement at the end of primary education resulted in higher mathematics achievement in secondary school. No significant effects for achievement growth were found (b^* = -.04, p = .579), indi-

cating that students' achievement growth in mathematics during primary education did not predict their mathematics achievement in secondary education (Model 2a). The findings of Model 3a showed a significant effect of gender on mathematics achievement (b^* = .20, p < .001), indicating that girls performed higher in secondary school mathematics than boys, after accounting for mathematics achievement (growth) in primary education. If boys and girls enter secondary education with similar levels of achievement in mathematics, girls outperform boys in mathematics at secondary education. No significant (direct) effects of students' SES on their mathematics achievement in secondary school were found on top of students' achievement (growth) in primary education (b^* = .01, p = .828). Adding students' track placement to the models did not change the results (see Models 2b and 3b).

For students' obtained level of education in Year 2 (Grade 8), Model 1 (Table 11) revealed that students' most recent reading comprehension and mathematics achievement in primary education were significantly positively related to students' level of education in secondary school ($b_{Reading}{}^*$ = .47, p < .001; $b_{Math}{}^*$ = .40, p < .001). Students with higher reading comprehension and mathematics achievement at the end of primary education, attended a higher level of secondary education. No significant effects for achievement growth were found ($b_{Reading}{}^*$ = .46, p = .315; $b_{Math}{}^*$ = -.28, p = .339), indicating that achievement growth did not predict students' level of secondary education (Model 2a). Furthermore, while no gender effects were found (see Model 3a; b^* = -.02, p = .978), students' SES was significantly positively related to students' level of secondary education (b^* = .11, p = .005), on top of their most recent achievement in primary school. Higher-SES students attained a higher secondary school level in Year 2 than lower-SES students, even when they entered secondary school with similar achievement (growth) in primary education. This was a small-sized effect. Adding students' track placement to the models turned all effects non-significant ($b_{Reading}{}^*$ = .02, p = .635; $b_{Math}{}^*$ = .06, p = .404; $b_{SES}{}^*$ = .02, p = .404), indicating that within the track, students did not differ in their reading comprehension, mathematics achievement in primary education and SES (see Models 2b and 3b).

Table 9

	Model 0: Empty				Model 1: Intercept				Model 2a: + Slope			
			95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects												
Intercept	185.10***	4.03	177.21	192.99	130.52***	16.52	134.68	184.90	96.06***	20.27	56.33	135.80
Reading comprehension intercept					1.02**	0.31	0.41	0.41	0.79**	0.29	0.22	1.36
Reading comprehension slope									-2.03	1.86	-5.68	1.61
Track placement												
Gender (girls)												
SES												
Variance												
R^2					.700				.125			
Model fit												
AIC	5,908.35				292,760.11				256,350.60			
BIC	5,931.20				292,949.75				256,622.42			
RMSEA	0.540				0.255				0.043			
CFI	0.00				0.72				0.99			
TLI	0.00				0.71				0.99			
SRMR	0.499				0.731				0.058			
Number of schools	333				4,111				4,111			
Number of students	79				101				101			

	Model 2b: + Slope + track placement				Model 3a: + Gender + SES				Model 3b: + Gender + SES + track placement			
			95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects												
Intercept	63.43***	17.30	29.52	97.35	87.68***	22.39	43.79	131.56	48.63**	17.58	14.16	83.09
Reading comprehension intercept	0.10	0.32	-0.52	0.72	1.01**	0.33	0.36	1.65	0.20	0.34	-0.46	0.86
Reading comprehension slope	1.54	2.38	-3.12	6.21	-1.30	1.98	-5.18	2.59	1.45	2.22	-2.90	5.80
Track placement	26.68***	6.29	14.36	39.01					27.32***	5.56	16.43	38.21
Gender (girls)					12.95*	5.70	1.77	24.13	13.04*	5.34	2.57	23.50
SES					-2.95	3.58	-9.95	4.06	-6.60*	3.26	-12.99	-0.21
Variance												
R^2	.285				.160				.321			
Model fit												
AIC	256,486.06				272,703.96				272,844.12			
BIC	256,814.77				273,097.05				273,306.95			
RMSEA	0.042				0.040				0.039			
CFI	0.99				0.99				0.99			
TLI	0.99				0.99				0.09			
SRMR	0.058				0.050				0.049			
Number of schools	4,111				4,189				4,189			
Number of students	101				102				102			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Unstandardised Estimates of Growth Curve Models 0 to 3 Predicting Secondary School Reading Comprehension Achievement With Primary School Achievement, Achievement Growth, Track Placement, and Background Characteristics Gender, and SES (RQ3)

Table 10

	Model 0: Empty				Model 1: Intercept				Model 2a: + Slope			
			95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects												
Intercept	160.82***	2.28	156.35	165.29	39.41**	16.07	7.93	70.90	-40.59	21.75	-83.21	2.04
Mathematics intercept					1.42***	0.16	0.50	1.63	1.76***	0.13	1.51	2.01
Mathematics slope									-0.49	0.90	-2.26	1.27
Track placement												
Gender (girls)												
SES												
Variance												
R^2					.286				.445			
Model fit												
AIC	5,908.35				292,760.11				256,350.60			
BIC	5,931.20				292,949.75				256,622.42			
RMSEA	0.540				0.255				0.043			
CFI	0.00				0.72				0.99			
TLI	0.00				0.71				0.99			
SRMR	0.499				0.731				0.058			
Number of schools	333				4,111				4,111			
Number of students	79				101				101			

Table 10 continued

	Model 2b: + Slope + track placement				Model 3a: + Gender + SES				Model 3b: + Gender + SES + track placement			
			95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects												
Intercept	-8.90	27.77	-63.33	45.52	-50.08*	21.78	-92.76	-7.39	-28.66	24.38	-76.44	19.12
Mathematics intercept	1.08***	0.24	0.61	1.56	1.82***	0.14	1.56	2.08	1.23***	0.23	0.78	1.67
Mathematics slope	0.81	1.14	-1.42	3.04	-0.17	0.93	-1.99	1.66	0.74	1.07	-1.36	2.84
Track placement	13.32***	2.66	8.12	18.52					12.67***	2.31	8.14	17.20
Gender (girls)					13.19***	3.45	6.43	19.95	11.94***	3.41	5.25	18.63
SES					0.42	1.95	-3.40	4.25	-2.87	1.89	-6.57	0.84
Variance												
R^2	.548				.503				.591			
Model fit												
AIC	256,486.06				272,703.96				272,844.12			
BIC	256,814.77				273,097.05				273,306.95			
RMSEA	0.042				0.040				0.039			
CFI	0.99				0.99				0.99			
TLI	0.99				0.99				0.09			
SRMR	0.058				0.050				0.049			
Number of schools	4,111				4,189				4,189			
Number of students	101				102				102			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Unstandardised Estimates of Growth Curve Models 0 to 3 Predicting Secondary School Mathematics Achievement With Primary School Achievement, Achievement Growth, Track Placement, and Background Characteristics Gender, and SES (RQ3)

Table 11

	Model 0: Empty				Model 1: Intercept				Model 2a: + Slope			
			95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects												
Intercept	6.21***	0.08	6.06	6.36	0.88	0.51	-0.13	1.88	-2.81	2.20	0.19	0.45
Reading comprehension intercept					0.04***	0.01	0.03	0.05	0.03***	0.01	0.02	0.04
Reading comprehension slope					0.03***	0.01	0.02	0.05	0.18	0.17	-0.16	0.51
Mathematics intercept									0.07**	0.02	0.02	0.11
Mathematics slope									-0.12	0.13	-0.36	0.13
Track placement												
Gender (girls)												
SES												
Variance												
R^2					.665				.724			
Model fit												
AIC	5,908.35				292,760.11				256,350.60			
BIC	5,931.20				292,949.75				256,622.42			
RMSEA	0.540				0.255				0.043			
CFI	0.00				0.72				0.99			
TLI	0.00				0.71				0.99			
SRMR	0.499				0.731				0.058			
Number of schools	333				4,111				4,111			
Number of students	79				101				101			

Table 11 continued

	Model 2b: + Slope + track placement				Model 3a: + Gender + SES				Model 3b: + Gender + SES + track placement			
			95% CI				95% CI				95% CI	
	b	SE_b	LL	UL	*b*	SE_b	LL	UL	*b*	SE_b	LL	UL
Fixed effects												
Intercept	0.91	0.67	-0.41	2.23	-1.72	1.53	-4.72	1.28	0.87	0.61	-0.32	2.07
Reading comprehension intercept	0.00	0.00	0.00	0.01	0.03***	0.01	0.02	0.05	0.00	0.00	-0.01	0.01
Reading comprehension slope	-0.01	0.06	-0.12	0.10	0.11	0.14	-0.17	0.38	-0.02	0.06	-0.13	0.10
Mathematics intercept	0.00	0.01	-0.01	0.02	0.05**	0.02	0.02	0.08	0.01	0.01	-0.01	0.02
Mathematics slope	0.01	0.04	-0.08	0.09	-0.07	0.10	-0.28	0.13	0.01	0.05	-0.08	0.10
Track placement	0.86***	0.05	0.76	0.97					0.86***	0.05	0.76	0.96
Gender (girls)					0.00	0.13	-0.26	0.25	0.06	0.05	-0.05	0.16
SES					0.13***	0.05	0.04	0.23	0.03	0.03	-0.03	0.08
Variance												
R^2	.924				.747				.925			
Model fit												
AIC	256,486.06				272,703.96				272,844.12			
BIC	256,814.77				273,097.05				273,306.95			
RMSEA	0.042				0.040				0.039			
CFI	0.99				0.99				0.99			
TLI	0.99				0.99				0.09			
SRMR	0.058				0.050				0.049			
Number of schools	4,111				4,189				4,189			
Number of students	101				102				102			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardised root mean square residual.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Unstandardised Estimates of Growth Curve Models 0 to 3 Predicting Secondary School Level With Primary School Achievement, Achievement Growth, Track Placement, and Background Characteristics Gender, and SES (RQ3)

Discussion

Previous research has shown that track recommendations are primarily based on students' (most recent) achievement in primary education (van Leest et al., 2020). Using longitudinal data following students across the transition from primary to secondary education, allowed

us to explore the extent to which teachers also consider students' achievement growth in reading comprehension and mathematics achievement in their track recommendations and to what extent achievement growth is predictive of secondary school performance. Overall, the findings suggested that students' achievement (growth) in primary education differed by their gender and socioeconomic status (SES). Teachers primarily considered students' most recent achievement in primary education, but not students' achievement growth when formulating track recommendations. Hence, gender and SES differences in growth curves did not impact teachers' track recommendations. Lastly, students' most recent achievement but not their achievement growth was predictive of students' secondary school performance.

Students' Achievement Growth in Primary Education (RQ1)

Students' mean growth curve of both reading comprehension and mathematics development was linear. However, there was considerable variation between individual students regarding their growth curve. In line with previous studies, which found 'Matthew effects' in the reading domain (Curby et al., 2009; Salaschek et al., 2014), the findings of the present study showed that higher initial reading comprehension achievement was associated with faster achievement growth, indicating that achievement gaps in reading comprehension become larger over time (i.e. a widening achievement gap). Additionally, in line with these studies showing 'compensatory effects' in mathematics (Aarnoutse & van Leeuwe, 2000; Salaschek et al., 2014), the findings of the present study showed that lower initial mathematics achievement was associated with faster achievement growth. This suggests that achievement gaps in mathematics become smaller over time (i.e. a narrowing achievement gap).

For gender, the findings indicated a narrowing achievement gap for reading comprehension and mathematics. Initially, girls had higher reading comprehension achievement than boys, but boys showed faster achievement growth in reading comprehension, while the opposite was true for mathematics. For mathematics, the findings of the present study were in line with those of (Caro, Lenkeit, et al., 2009). Regarding SES, there was only a widening achievement gap in reading comprehension. Higher-SES students had higher reading comprehension achievement and faster achievement growth than lower-SES students, which was in line with prior research (Helbling et al., 2019).

Considering Students' Growth Curves in Track Recommendations (RQ2)

Although students showed different achievement growth, students' achievement growth was not considered in track recommendations by the teachers in the sample of the present study. This was contrary to the results of previous research (Caro, Lenkeit, et al., 2009; Klapproth & Fischer, 2019). As such, gender and SES differences in growth curves were not reflected in teachers' track recommendations. The difference between the findings of the present study and that of Caro, Lenkeit, et al. (2009), and Klapproth and Fischer (2019) may

be explained by the point of reference used. These studies used students' initial achievement in Grade 4 as point of reference, while in the present study students' achievement at the end of primary school was used. Taking the initial achievement as reference point may falsely suggest that teachers consider achievement growth in their track recommendations. As a result, the effect of students' achievement growth in track recommendations may be overestimated. However, by taking students' most recent achievement in primary education as reference point, the findings of the present study were able to show that teachers do not consider achievement growth in their track recommendations, but primarily rely on students' most recent achievement.

Although students' achievement growth in primary education may help teachers to determine students' potential, it may in fact be difficult for teachers to use students' growth curves to indicate one appropriate level of secondary education as there is considerable variation in growth curves between individual students, as well as non-linear fluctuations within students' achievement during primary education. The Dutch Inspectorate of Education (2014) also suggested that guidelines regarding track recommendations may be more difficult to apply when there is no clear picture of the student, and therefore may not always be applied in such cases. The local guidelines regarding the formulation of track recommendations in the city our data originates from (POVO, 2015; Smeets et al., 2014) recommend teachers to primarily rely on students' achievement in the two final grades of primary education rather than relying on achievement growth. However, the Dutch government recommended, right after the time of the data collection of the present study, teachers to consider students' growth in their track recommendations as it may predict students' educational trajectory in secondary education (Inspectie van het Onderwijs, 2018d). The findings suggest that, at the time of our data collection, teachers did not yet follow this national recommendation. In recent years, the Dutch government has increasingly argued for giving students the benefit of the doubt in cases where the track recommendation is not clear (Inspectie van het Onderwijs, 2019; Onderwijsraad, 2019, 2021). It would therefore be interesting to see if achievement growth may have been considered more strongly in more recent track recommendations.

Furthermore, when examining the answers on RQ1 and RQ2 together, it can be determined whether there are variations in growth curves between groups of students and if teachers evaluate them differently in their track recommendations. Although growth curves in reading comprehension and mathematics differed between groups of students, teachers did not consider achievement growth in their track recommendations on top of students' most recent achievement.

The Predictive Value of Students' Growth Curves in Secondary Education (RQ3)

In line with previous research (de Boer et al., 2010; Poncelet & Metis Associates, 2004; van Rooijen et al., 2016), students' most recent achievement at the end of primary school was found to be predictive of students' secondary school performance. Students with higher

reading comprehension or mathematics achievement at the end of primary education, had higher secondary school performance, in terms of test scores as well as their level of education, regardless of their previous growth curve. Moreover, higher-SES students attained a higher secondary school level in Year 2 than lower-SES students, even when they entered secondary school with similar achievement and achievement growth in primary education.

It is argued that postponing the allocation decision helps to reduce the widening achievement gap between students in lower and higher tracks (Brunello & Checchi, 2007; Hanushek & Wössmann, 2006; Horn, 2013; Miller, 2018; van Elk et al., 2009), and between lower- and higher-SES students (Bauer & Riphahn, 2006; Meghir & Palme, 2005; OECD, 2008, 2020; Pekkala Kerr et al., 2013), as it gives these students more time to develop themselves. This study's results indicate that policymakers and educators should consider the potential impact of taking achievement growth into account. While this may improve track recommendations to better match students' potential, it could also have unintended consequences for equity of opportunity between different groups. That is, the findings of the present study suggest that the achievement gap between lower- and higher-SES students widens with age. Hence, if this gap continues to widen, postponing the track placement decision could cause even greater differences in track allocation, unless effective interventions are implemented that reduce socioeconomic achievement gaps.

Furthermore, the findings of the present study indicated that achievement growth was not predictive of secondary school performance. If achievement growth would have been predictive of secondary school performance, this would suggest that teachers should consider it in their track recommendations. However, the opposite is not necessarily true. The finding that achievement growth in primary education was not predictive of secondary school performance, does not necessarily imply that teachers should not consider achievement growth in their track recommendations, as this finding was possibly caused by tracking itself. That is, attending a certain track may have an impact on students' achievement within that track: students who are placed in a track below their potential, may only put in the work necessary to succeed in that track, while students who are in a track which exceeds their most recent achievement in primary education, may work extremely hard and receive additional support to succeed at that level. Hence, the track recommendation then becomes a self-fulfilling prophecy (cf. Rosenthal & Jacobson, 1968) in which case track placement prevents students with strong achievement growth to further realize their potential.

Because teachers did not consider achievement growth in their track recommendations, gender and SES differences in achievement growth did not translate into differences in track recommendations. This can actually be considered beneficial for lower-SES students as they their growth rates in achievement were lower than those of their higher-SES counterparts. The same is the case for boys: girls had a higher mathematics growth rate than boys, but boys had a higher reading comprehension growth rate than girls. Combined with the fact that Dutch teachers weigh mathematics achievement more heavily than reading

comprehension (i.e. 'a predominant subject domain') (Driessen et al., 2008; Smeets et al., 2014; van Leest et al., 2024), weighing achievement growth would negatively impact boys' track recommendations.

Finally, to account for the track students were placed in at the start of secondary education, students' track placement was included as a predictor in the analyses of RQ3. The results of these models showed that, compared to the models without track placement, some of the significant effects of reading comprehension achievement in primary school on achievement in secondary school were no longer significant after track placement was added as a predictor. This indicates that, as expected, students in higher tracks, generally performed better in reading comprehension in primary school than students in lower tracks. However, for students in a similar track, it was not the case that achievement differences in reading comprehension in primary education predicted who performed better in secondary education. This is in line with research showing that students will generally perform according to the track they are in (i.e. due to self-fulfilling prophecy effects, cf. Rosenthal & Jacobson, 1968), regardless of their prior performance. The effects of mathematics and gender on secondary school performance remained significant after including track placement. Hence, differences in mathematics achievement in primary education predicted both achievement differences between students in different tracks, but also achievement differences between students within the same track. Moreover, boys and girls generally performed differently in mathematics in secondary education, even when they were in a similar track.

In addition, some effects became significant after students' track placement was added. For example, SES became significant for reading comprehension in secondary school, indicating that within a track, students with different SES backgrounds performed differently on reading comprehension, but not between tracks. This finding supports the idea of achievement overlapping between tracks: there are students in higher tracks that perform similarly as students in lower tracks.

Limitations and Future Research

In interpreting the results of the present study, a number of limitations are worth noting. First, in the present study, SES was based on students' six-digit postal code. While these six-digit postal codes are, on average, only shared by 15 to 20 households and are considered to be an accurate impression of the SES of those households (Deckers et al., 2016; Guhn et al., 2010; van Hattem et al., 2009), it is not a measure of each individual household.

Second, the specific context of this research, that is, track recommendations formulated in a large city in the Netherlands, might affect the generalisability of the results of the present study. Track recommendations can be formulated in different ways in different regions. Other regions may, for example, allow combined track recommendations. Moreover, results are likely to be different in other countries with different educational systems.

Third, secondary school performance in the present study was measured only once in the first year(s) of secondary school. Although students' growth in primary education did not contribute to students' track recommendations or secondary school performance, the present study did not include students' achievement growth after several years of secondary school. It would have been interesting to include achievement measures or students' track after several years, as most switching between levels of secondary education or grade repetition in secondary education happens in later years (van Vuuren & van der Wiel, 2015; Veenstra, 1999). Moreover, the present study only included one measure of students' performance in secondary school. Therefore, it is not clear whether students continue their growth curve from primary school in secondary school or not. For future research, it could be interesting to include multiple assessments of students' achievement in secondary school to investigate how achievement growth in primary school is associated with achievement growth in secondary school and how track recommendations and track placement may affect students' growth curves. It could be particularly interesting to focus on lower secondary education, since an increasing number of Dutch secondary schools introduced a form of heterogeneous lower secondary education, where students from different tracks are taught together. This may reduce the impact of tracking in the first years of secondary education, because lower-performing students tend to perform better in heterogeneous classes (Grift et al., 2010; van de Werfhorst, 2021). This is believed to be due to factors, such as a higher level of instruction, the influence of higher-performing peers, and less harm to students' self-esteem (De Fraine et al., 2003; Oakes, 2005; Venkatakrishnan & Wiliam, 2003).

Furthermore, it would be interesting for future research to examine more in-depth, for instance with qualitative research (interviews, case studies), how teachers perceive different achievement aspects of students' achievement growth (initial achievement, growth curves, most recent achievement). It could be that teachers might find it difficult to consider growth curves, as growth often fluctuates. More in-depth research could reveal the reasons why teachers do not include achievement growth in their track recommendations.

In addition, since there are no clear guidelines on how teachers should consider students' achievement growth in their track recommendations (Inspectie van het Onderwijs, 2014). Therefore, there could be differences between teachers in the extent to which they include achievement growth in their track recommendations. If some teachers weigh a high level of growth negatively (because a student initially had lower achievement) and others positively (because they consider students achievement growth to reflect a student's potential), these effects may cancel each other out, resulting in an overall non-significant effect. Future research could therefore also include random effects to test for differences between teachers, schools or even school boards, as school boards typically formulate the guidelines and procedures regarding the formulation of track recommendations (e.g. Timmermans et al., 2023).

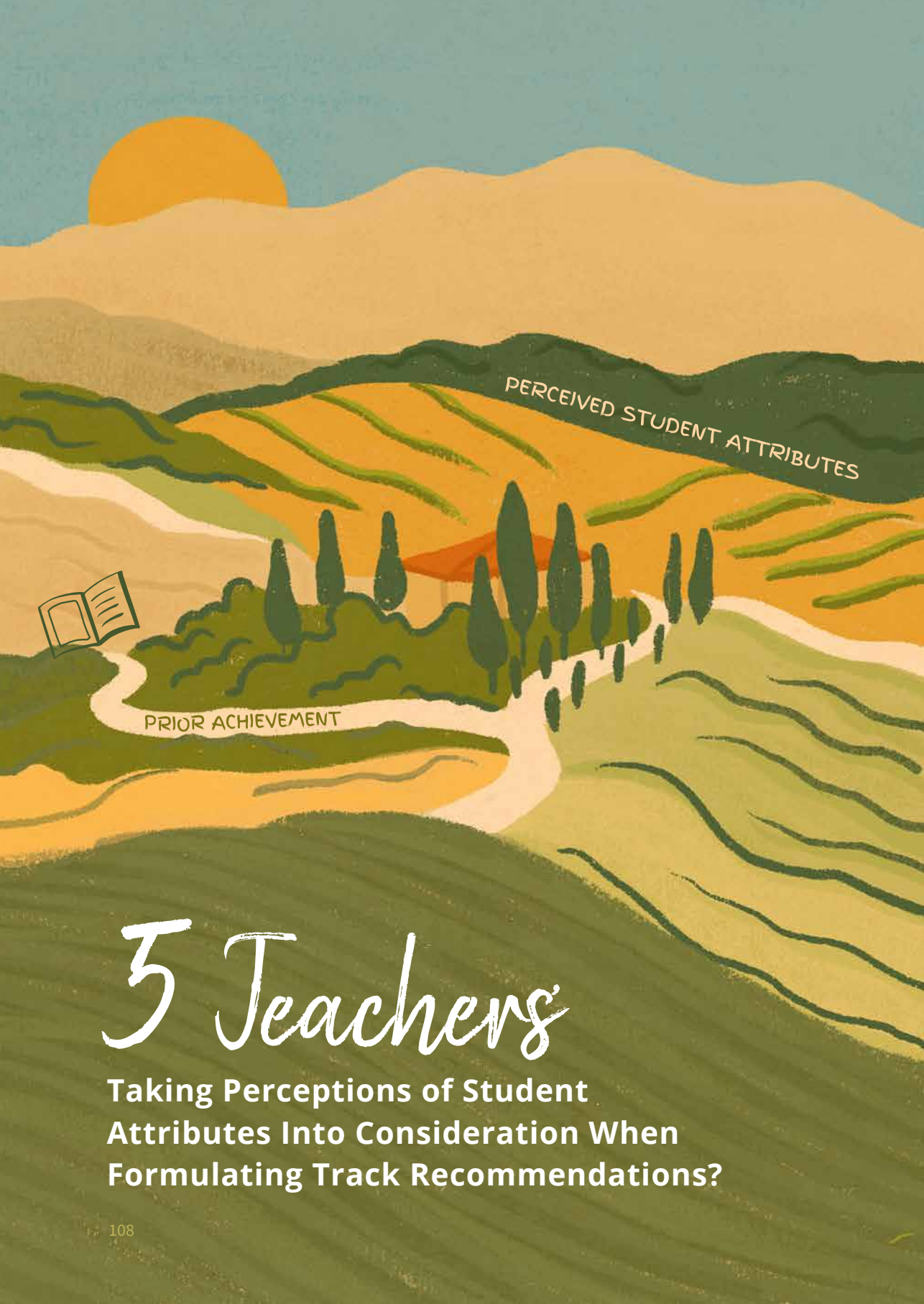
Conclusion

The present study extends the knowledge base on the role of students' achievement growth in primary education in teachers' track recommendations and secondary school achievement. The findings of the present study indicated that students' achievement growth in primary education is not reflected in their teachers' track recommendations nor in their secondary school performance. As such, SES and gender differences in growth curves did not translate into different track recommendations. SES and gender differences in track recommendations were mostly caused by differences in students' achievement at the end of primary school. Additionally, lower-SES students received slightly lower track recommendations after accounting for prior achievement, suggesting a small SES bias in teachers' track recommendations.

"We kijken naar de Cito-gegevens 6, 7, en 8 en met name begrijpend lezen en rekenen, omdat de middelbare scholen daar ook vooral om vragen. Het kan zijn dat dat fluctueert en daarom vind ik het belangrijk om naar de lijn te kijken. Het kan zijn dat er een stijgende lijn in zit. Dan geeft mij dat hoop voor de toekomst. Zijn er veel pieken en dalen, dan kijken we naar de methodengebonden toetsen."

"Een leerling was een laatbloeier: hij begon met een 4 voor begrijpend lezen in groep 5, toen naar 3 in groep 6, toen 2, naar 1. Hij doet nu vwo. Dan zie je die lijn die gaat omhoog: die gaat het gewoon redden."

"We kijken dan ook naar de ontwikkeling in prestaties, vooral als we twijfels hebben."



5 Teachers

Taking Perceptions of Student Attributes Into Consideration When Formulating Track Recommendations?

Abstract

In some tracked educational systems, track recommendations are formulated by primary school teachers to determine the secondary school level students will be allocated to. While teachers mostly base their track recommendations on students' prior achievement, the extent to which teachers also consider perceived student attributes, such as students' perceived work habits or parental involvement, and the extent to which these perceived student attributes are predictive of secondary school performance is unclear. Therefore, the aim of this study was to examine the extent to which teachers consider their perceptions of student attributes in their track recommendations and the extent to which these attributes are predictive of their secondary school performance. Participants were 17,953 Grade 6 students from 1,105 Dutch primary school teachers and 4,150 Grade 9 students from 1,289 Dutch secondary school classes. Data were analysed using multilevel models. Results showed that teacher perceived student attributes played only a minor role in track recommendations and secondary school performance. Yet, the extent to which these attributes were considered by teachers differed based on students' background and differed between teachers. For secondary school performance, teacher perceived student attributes had limited predictive value. Overall, these findings indicated that teachers may need to be careful with taking perceived student attributes into account when formulating track recommendations since their limited predictive value for students' performance in secondary education.

Keywords:

equal educational opportunities, transition from primary to secondary education, school track recommendation, secondary school success, SES, gender, migration background, perceived student attributes, work habit, classroom behaviour, parental involvement, reading comprehension, mathematics, teacher judgement

This chapter is based on: van Leest, A., Hornstra, L., van Tartwijk, J., & van de Pol, J. (2024). Teachers taking perceptions of student attributes into consideration when formulating track recommendations? British Educational Research Journal, 50(3), 1127-1171. https://doi.org/10.1002/berj.3957

Acknowledgement of author contributions: All authors contributed to the design of the study. AvL retrieved the data from DANS and prepared the data for analyses; AvL analysed the data with help from LH; AvL drafted the manuscript; All authors contributed to critical revision of the manuscript; LH, JvdP, and JvT supervised the study.

Introduction

In tracked educational systems, secondary education consists of hierarchical school tracks (Contini & Scagni, 2011; Glock et al., 2012; Korpershoek et al., 2016). In some countries with tracked educational systems, such as Germany, Belgium, Luxembourg, and the Netherlands, primary school teachers formulate track recommendations to determine the allocation of students to these tracks (Glock et al., 2012; Korpershoek et al., 2016; Timmermans, Kuyper, et al., 2015). These track recommendations are extremely important in students' educational career and beyond as they largely determine the level of secondary education students will be allocated to, and thereby the educational qualifications students will acquire (Boone & Demanet, 2020; Contini & Scagni, 2011; Klapproth et al., 2012; Reed et al., 2015; Rodrigues et al., 2018; G. M. Strand, 2020; Timmermans, Kuyper, et al., 2015).

Previous research indicates that teachers primarily rely on students' achievement in primary education when formulating these track recommendations (Driessen, 2005; Feron et al., 2013; Luyten & Bosker, 2004; Timmermans, Kuyper, et al., 2015; van Leest et al., 2020, 2024), but also to a small extent on students' background characteristics, such as their gender, SES, or migration background (Batruch et al., 2023; Boone et al., 2018; Klapproth et al., 2013; Korpershoek et al., 2016; Krolak-Schwerdt et al., 2017; Timmermans et al., 2013; van Leest et al., 2020, 2024). The extent to which teachers also consider other student attributes in their track recommendations, such as their perceptions of students' work habits or parental involvement, and how predictive these perceived attributes are of secondary school performance is, however, unclear. While there is strong consensus that students' background characteristics should not influence teachers' track recommendations (Inspectie van het Onderwijs, 2018d; Luyten, 2004; OECD, 2016b), there is no such consensus regarding whether teachers may want to consider their perception of student attributes, such as perceived work habits, teacher-student relationship, and parental involvement, in their track recommendations. More insights into the role of perceived student attributions concerning teachers' track recommendations and students' future school performance in secondary school will help to inform the debate on whether or not it is desirable for teachers to consider these attributes in their recommendations.

The present study aimed to provide more insight in the role of teacher perceptions of student attributes on top of prior achievement in the formulation of track recommendations. More specifically, the aim was to examine the extent to which teachers include their perceptions of student attributes in their track recommendations, if teachers do that differently based on students' gender, SES, or migration background, and if there are between-teacher differences in the extent to which they include these perceived attributes in their track recommendations. Moreover, using a longitudinal design spanning across the transition from primary to secondary education, it was investigated the extent to which teacher perceived student attributes are predictive of students' performance in secondary education.

Impact of Perceived Student Attributes on Track Recommendations

Track recommendations are formulated in the final grade of primary education (Grade 6) to indicate a secondary school level that matches students' potential performance in secondary education best (Boone & Van Houtte, 2013; de Boer et al., 2010; Klapproth et al., 2012; Le Métais, 2003). Teachers' track recommendations are based on their expectations (Boone & Van Houtte, 2013; Glock et al., 2012; Le Métais, 2003). While prior achievement typically explains about 80% of the variance in track recommendations (Timmermans, Kuyper, et al., 2015; van Leest et al., 2020), teachers may also consider other aspects when formulating track recommendations such as their perceptions of various other student attributes (Driessen, 2005; Luyten & Bosker, 2004; Sneyers et al., 2018; Timmermans et al., 2016, 2019). Instead of simultaneously analysing the role of multiple teachers' perceived student attributes, most prior research focused on one specific student attribute (Alvidrez & Weinstein, 1999; Driessen, 2006; Rubie-Davies, 2010; Timmermans et al., 2019). In the present study, multiple perceived student attributes are included that were indicated by prior research as factors that teachers possibly consider when formulating track recommendations, in addition to factors as students' achievement and background characteristics (Inspectie van het Onderwijs, 2018d). Therefore, the perceived student attributes that were included in the present study cover a range of factors that are related to students' school behaviour and involvement, namely students' work habits, classroom behaviour, underachievement, popularity, teacher-student relationship, and parental involvement. Given that the aim of the study was to examine how teachers weigh these attributes, the focus of the present study is on teacher perceptions of these attributes.

Teachers have been found to include their perceptions of several student attributes in their track recommendations, on top of students' achievement, explaining approximately between 1% and 10% of the variance in track recommendations (Driessen, 2005, 2006; Driessen et al., 2005, 2008; Luyten & Bosker, 2004; Sneyers et al., 2018; Timmermans et al., 2016, 2019). That is, perceived work habits, underachievement, and parental involvement have been found to be considered by teachers in their track recommendations (Driessen, 2005; Driessen et al., 2008; Luyten & Bosker, 2004; Sneyers et al., 2018; Timmermans et al., 2016, 2019). Students who – according to their teacher – underachieved received lower track recommendations than students who did not underachieve, and students who were perceived to have strong work habits or highly involved parents received higher track recommendations than students who, according to their teachers, had poorer work habits or less involved. Furthermore, some studies found no significant relationship between teacher perceived classroom behaviour and track recommendation (Luyten & Bosker, 2004), whereas other studies found higher track recommendations for students with negative perceived classroom behaviour (Driessen et al., 2005, 2006; Timmermans et al., 2016).

Mixed results were also found for the teacher-student relationship (TSR) (Driessen et al., 2008; Luyten & Bosker, 2004; Sneyers et al., 2018; Timmermans et al., 2016, 2019). Regarding the TSR, some studies include this concept as one unidimensional construct,

while others distinguish three relationship dimensions based on the three-dimensional model of (Pianta et al., 1995): dependency, conflict, and closeness. Dependency refers to the degree of students' (inappropriate) possessiveness and overdependence in the relationship with the teacher, conflict refers to the degree of discordance, unpleasantness, and unpredictability of the relationship with the student, and closeness refers to the degree of security, emotional support, warmth, and openness in the relationship with the student. Two studies that conceptualized the TSR as a unidimensional construct found no significant association with track recommendation (Sneyers et al., 2018; Timmermans et al., 2016). When including the three different dimensions of this concept, Timmermans et al. (2019) found lower track recommendations for students who were perceived to be more dependent on their teacher. However, perceived conflict and closeness were not found to be associated with track recommendations. Furthermore, perceived popularity was also not found to be associated with track recommendations (Timmermans et al., 2016). Altogether, these findings indicated that teachers considered some of these student attributes when formulating track recommendations.

Differences Between Students With Different Background Characteristics and Differences Between Teachers

Prior research has shown that students with similar achievement levels sometimes receive different track recommendations based on their background characteristics, such as gender (e.g. De Boer et al., 2010; Driessen et al., 2008; Klapproth et al., 2013; Korpershoek et al., 2016; Krolak-Schwerdt et al., 2017; Luyten & Bosker, 2004; Timmermans et al., 2013, 2016; Van Leest et al., in press; Van Rooijen et al., 2017), socioeconomic status (SES) (e.g, Batruch et al., 2023; Boone & Van Houtte, 2013; Korpershoek et al., 2016; Timmermans et al., 2013, 2018; Van Leest et al., 2020, in press; Van Rooijen et al., 2017), and migration background (e.g. Barg, 2013; Batruch et al., 2023; Boone et al., 2018; Driessen et al., 2005, 2007; Glock & Krolak-Schwerdt, 2013; Klapproth et al., 2012; Ledoux et al., 2011; Riley & Ungerleider, 2008; Roeleveld et al., 2011). Furthermore, there is also some research indicating that teachers perceive student attributes, such as perceived classroom behaviour or parental involvement, more positively or negatively based on students' background characteristics (Bakker et al., 2007; Bennett et al., 1993; Riley & Ungerleider, 2012; Sui-Chu & Willms, 1996). Most likely, stereotypes related to student attributes, such as gender, SES or migration background, play a role here (Glock et al., 2013; Glock & Krolak-Schwerdt, 2013; Krolak-Schwerdt et al., 2013; Peterson et al., 2016; Strand, 2013; Van den Bergh et al., 2010). When students' characteristics are – according to the teacher – in line with the stereotype of the group the student belongs to, the teacher may be more likely to rely on the activated stereotype when formulating perceptions about that student (Glock, 2016a, 2016b; Klapproth et al., 2018; Strand, 2011, 2012). Consistent with this line of reasoning, it may be that teachers also weigh perceived student attributes differently in track recommendations for student with different background characteristics. For example, in case of

a lower-SES student, teachers may formulate a more careful – i.e. lower – track recommendation when they perceive this student to have poor work habits, while these issues may be overlooked or given less weight for higher-SES students. Therefore, in the present study it was examined whether perceived student attributes are considered differently by teachers based on students' gender, SES, and migration background.

The impact of teacher perceptions of student attributes on track recommendations may not only differ for different groups of students, but also between teachers (Timmermans et al., 2016, 2019). Prior research in Flanders, Belgium indicated that teachers seem to include different data sources and criteria when formulating track recommendations (Vanlommel & Schildkamp, 2019). Although there are less standardised student data available to teachers when formulating track recommendations in Flanders than in the Netherlands, we do expect that for the subjective part of the track recommendations, Dutch teachers act and behave in more or less the same way as their Flemish colleagues. Teachers thus have the opportunity – at least to some extent – to determine themselves which factors or student attributes they include when formulating track recommendations, which may cause teachers to vary in the extent to which they include different perceived attributes. In addition, more objective measures of student attributes are not structurally used by schools or available to teachers when formulating track recommendations (van den Berg et al., 2022). Even if these data are available to teachers, teachers mostly use these data intuitively (Vanlommel & Schildkamp, 2019). Hence, teachers may not only decide which student attributes to include, they also often assess or interpret these attributes themselves, resulting in subjective perceptions of these attributes (Praetorius et al., 2017; Vanlommel & Schildkamp, 2019; Zhu & Urhahne, 2020). If teachers make different decisions in the extent to which they consider their perceptions of student attributes in track recommendations, track recommendations can differ between students with the same perceived student attributes but different teachers.

Research has shown that there are small differences between teachers in how they weigh different student attributes in their track recommendations (Timmermans et al., 2019). Perceived classroom behaviour and conflictual teacher-student relationship and were sometimes weighed positively, and sometimes weighed negatively by different teachers, while perceived dependency, for example, was weighed negatively by all teachers in their track recommendations. For perceived work habits, a positive relationship with track recommendations was found for all teachers, although some teachers weigh this attribute more strongly in their recommendations than others (Timmermans et al., 2016). Although this research already indicates that there are small differences between teachers, more research on other attributes is needed to contribute to further insights regarding the differences in track recommendations. Therefore, in the present study, it is investigated whether there are differences between teachers in considering multiple perceived student attributes in their track recommendations.

Predictive Value of Perceived Student Attributes on Secondary School Performance

Although teachers have been found to consider some perceived student attributes in their track recommendations (Driessen, 2005; Driessen et al., 2008; Luyten & Bosker, 2004; Sneyers et al., 2018; Timmermans et al., 2016, 2019), the extent to which these perceived student attributes are actually predictive of students' secondary school performance is not yet known. In the present study, the focus was on students' achievement in third year of Dutch secondary education (Grade 9).[13] Ninth grade is an important grade (benchmark) in students' school career, because research has shown that the effect of the differences between the track recommendation formulated by the teacher and the track recommendation that is expected based on students' achievement on secondary school success is partially diminished after two years of secondary education (de Boer, 2009; de Jong & Steenbeeke, 2020).

Prior research in which student attributes were examined from the perspective of students themselves (or their parents with regard to parental involvement) suggested that these attributes are indeed predictive of students' future school performance (Bakker et al., 2007; Boonk et al., 2018; Borg, 2015; Roorda et al., 2017; Simpkins et al., 2020; van Rooijen et al., 2017), but it is unclear whether that is also the case for teacher perceptions of these attributes. Teacher perceptions of student attributes may be less predictive of students' future achievement. That is, as mentioned before, these teacher perceptions are subjective, meaning that teachers may not always accurately assess these attributes (Praetorius et al., 2017; Vanlommel & Schildkamp, 2019; Zhu & Urhahne, 2020). In addition, student attributes are not stable over time (Praetorius et al., 2017; Zhu & Urhahne, 2020). Hence, students' attributes in primary school may have limited predictive value for their performance in secondary school as students may have developed, for example, different work habits or classroom behaviours. This may particularly occur after the transition to secondary education as students have transitioned to a new educational context with different requirements and different teachers.

In all, research on the relation between teacher perceived student attributes in primary education and students' performance in secondary education is limited, amongst others since longitudinal research across the transition from primary to secondary education is scarce. There is, however, research among younger students indicating that teachers' perceptions of primary school students' parental involvement or teacher-student relationship were predictive of students' achievement in later grades in primary education (Hamre & Pianta, 2001; Izzo et al., 1999). However, the question remains whether teacher perceived

13 The Dutch secondary school system consists of six tracks that differ in educational qualifications students can acquire. The length of the track differs between four and six years of education: the lower educational tracks consist of four years, while the higher educational tracks consist of five or six years.

student attributes in primary education are also predictive of students' performance after their transition to secondary education. It is important to address this gap as this could help to inform the debate whether primary school teachers may or may not want to consider their perceptions of student attributes in their track recommendations. Therefore, in the present study, the aim was to examine the predictive value of these student attributes as perceived by primary school teachers for students' performance in secondary education.

Present Study

There is no consensus as to whether teachers may want to be careful with taking into account their perceptions of student attributes, such as perceived work habits or parental involvement, when formulating track recommendations. Therefore, the following research questions are adressed: To what extent do teachers include perceived student attributes in track recommendations, on top of students' prior achievement and background characteristics (RQ1a)? In the present study, teacher perceptions of students' work habits, classroom behaviour, underachievement, popularity, teacher-student relationship (attributes dependency, conflict, and closeness), and parental involvement were included as perceived student attributes. Because it is not clear whether teacher perceived student attributes are weighed to the same extent for students with different backgrounds and if teachers include these perceived attributions to the same extent in track recommendations, the extent to which teachers consider perceived student attributes differently for students with different backgrounds (RQ1b), and the extent to which there are differences between teachers in considering perceived student attributes in track recommendations (RQ1c) were also examined. Furthermore, the longitudinal dataset used in the present study gives us the unique opportunity to investigate the predictive value of these perceived student attributes in the context of the transition from primary education to secondary education. Therefore, the second research question of the present study is: To what extent are teacher perceived student attributes in primary school predictive of students' secondary school performance, on top of students' prior achievement (RQ2)?

Method

Sample

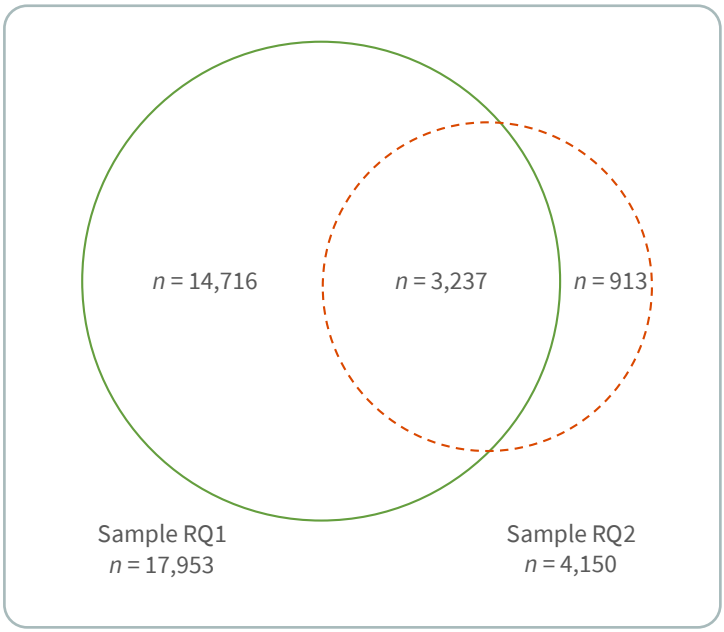
The data used in this research were derived from the Dutch COOL$^{5\text{-}18}$ cohort study. This longitudinal dataset includes multiple cohorts and contains data on students' educational career in primary and secondary education, including test scores and survey data on, amongst others, work habits and classroom behaviour. Data collection started in the academic year 2007-2008, with data collection waves every three years. Students and parents gave consent for participating in the study. The COOL$^{5\text{-}18}$ data are open access

and were retrieved from Data Archiving and Networked Services (DANS), a Dutch national repository for research data.

The Dutch primary schools of the COOL$^{5-18}$ data were selected based on four aspects, namely the socio-ethnic composition of a school's student population, type of school, province in which the school is located, and degree of urbanization of the school's location. To check the representativeness of the COOL$^{5-18}$ primary school data, the selection of primary schools of the COOL$^{5-18}$ data was compared to the national school record provided by the Dutch Ministry of Education on these four aspects. No significant differences ($p < .01$) between the COOL$^{5-18}$ dataset and the national cohorts were found (Driessen et al., 2009). For the secondary school data, the starting point was to include students who participated in a previous COOL$^{5-18}$ data collection. These data were supplemented by adding the classmates of these so-called target students. To check the representativeness of the COOL$^{5-18}$ secondary school data, a comparison was made between the distribution of COOL$^{5-18}$ students and the national distribution provided by the Dutch Education Executive Agency (i.e. Dienst Uitvoering Onderwijs, DUO) on secondary school level, province in which the school is located, degree of urbanization of the school's location, students' age, and gender. Small significant differences were found for the first two factors (Timmermans & Zijsling, 2014; Zijsling et al., 2017), with the COOL$^{5-18}$ dataset containing more students in the higher tracks, and an unequal distribution of schools across provinces. More detailed information about the data (collection) can be found on the COOL$^{5-18}$ website (COOL$^{5-18}$, n.d.), and in several technical reports (Driessen et al., 2009, 2012; Kuyper et al., 2014; Timmermans et al., 2017; Timmermans & Zijsling, 2014; Zijsling et al., 2014, 2017).

To include data across the transition from primary to secondary education, the aim was to include cohorts for which data were available from Grades 6 (last grade of primary education) and 9 (third grade of secondary education). The selected datasets included students who were either in sixth grade in the academic years 2007-2008 (Cohort 1) or 2010-2011 (Cohort 2). Together, the datasets included 18,866 students. For both research questions, a different subset of the data was used. For RQ1, data consisted of 17,953 students (9,839 from Cohort 1 and 8,114 from Cohort 2) from 674 primary schools of 1,105 primary school teachers. There were no significant differences (all $p > .05$) between the subsample used in the present study and the full COOL$^{5-18}$ cohort data, and between the two cohorts used in the present study. Because not all secondary schools (and thus students) participated in the data collection waves, the primary school data could only be partially linked to secondary school data. Students who only participated in secondary school were excluded from the present research. Therefore, for RQ2, the dataset consisted of 4,150 students (2,587 from Cohort 1 and 1,563 from Cohort 2) from 1,289 secondary school classes from 741 primary school classes. 913 students in secondary education had a missing on the variable track recommendation and were therefore excluded from the analyses of RQ1. See Figure 1 and Table 1 for a representation of both samples.

Figure 1



Data Samples of RQ1 and RQ2

Table 1

	Cohort 1	Cohort 2	Total
Students only in data sample RQ1	7,585	7,131	14,716
Students only in data sample RQ2	333	580	913
Students in data sample RQ1 and RQ2	2,254	983	3,237
Total	10,172	8,694	18,866

Data Samples of RQ1 and RQ2 per Cohort

Measures

The variables used in the present study were assessed and measured in the exact same way for all cohorts.

Track Recommendation. At the time of data collection, teachers in the Netherlands formulated a track recommendation in the final grade of primary education (Grade 6) after a school leavers' test was administered[14]. The results of this test were available when teachers formulated track recommendations and could be taken into account by teachers when formulating their track recommendations. Teachers were asked to indicate a single track or two adjacent tracks. In the present study, teachers' recommendations were recoded based on an adapted version (van Aarsen et al., 2013) of the "educational ladder" (Bosker et al., 1985) which was developed to map students' level of secondary education (see Table 2). The track recommendation is coded similarly as the coding for students' tracks in Grade 7 (the first year of secondary education in this educational ladder). The track recommendations were coded on a scale from 1 to 6, corresponding to the six secondary school tracks, from lowest to highest track: (1) practical training, (2) basic pre-vocational secondary education, (3) middle pre-vocational secondary education, (4) theoretical pre-vocational secondary education, (5) senior general secondary education, and (6) pre-university education. A combination of adjacent tracks is coded as the average of the two tracks. Track recommendation was considered as a continuous variable (cf. Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993).

14 Due to a policy reform in the Netherlands, the track recommendation procedure changed (Ministerie van Onderwijs Cultuur en Wetenschap, 2014). Before the policy reform, the school leavers' test was administered before teachers formulated track recommendations (Luyten & Bosker, 2004), while after the policy reform, the results of this test were no longer available when formulating track recommendations (Oomens et al., 2019).

Table 2

Secondary school track		Grade 7[b]	Grade 8	Grade 9
Single track	Two adjacent tracks			
(1) practical training		1	2	3
	(1) practical training / (2) basic pre-vocational	1.5	2.5	3.5
(2) basic pre-vocational		2	3	4
	(2) basic pre-vocational / (3) middle pre-vocational	2.5	3.5	4.5
(3) middle pre-vocational		3	4	5
	(3) middle pre-vocational / (4) theoretical pre-vocational	3.5	4.5	5.5
(4) theoretical pre-vocational		4	5	6
	(4) theoretical pre-vocational / (5) senior general	4.5	5.5	6.5
(5) senior general		5	6	7
	(5) senior general / (6) pre-university	5.5	6.5	7.5
(6) pre-university		6	7	8

Note. Grade 7 = first year of secondary education; Grade 8 = second year of secondary education; Grade 9 = third year of secondary education.

[a]Based on an adapted version (van Aarsen et al., 2013) of the "educational ladder" (Bosker et al., 1985)

[b]The track recommendation is coded the same as Grade 7

The Educational Ladder[a]

Teacher Perceived Student Attributes. Teachers' perceptions of students' work habits, classroom behaviour, underachievement, popularity, three attributes of teacher-student relationship, and parental involvement in sixth grade of primary education were included as perceived student attributes. Teachers' perceptions of these attributes were assessed by a student profile questionnaire teachers filled out about each student. Each factor contained several items measured on a 5-point Likert type scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The number of items, example items and reliability of each scale is presented in Table 3. The questions about teachers' perceptions of student attributes (work habits, classroom behaviour, underachievement, and popularity) were derived from the Three Minute Student Profile used for the PRIMA cohort studies (the longitudinal cohort study before COOL[5-18]) (Driessen et al., 2006). The teacher-student relationship was divided in three attributes based on the three-dimensional model of (Pianta et al., 1995):

dependency, conflict, and closeness. The teacher-student relationship questionnaire is a shortened version of the Student-Teacher Relationship Questionnaire (Koomen et al., 2007), that is based on the Student-Teacher Relationship Scale (Pianta, 2001).

Table 3

Scale	Number of items	Example items	Reliability (α)[a]
Work habits	4	This student works accurately.	.81
Classroom behaviour	4	This student behaves according to the class rules.	.81
Underachievement	3	This student is able to perform better.	.85
Popularity	3	This student is popular among classmates.	.86
TSR: Dependency	5	This student asks for help from me even when it is not necessary.	.90
TSR: Conflict	5	This student feels that I am treating him/her unfairly.	.93
TSR: Closeness	5	I share an affectionate, warm relationship with this student.	.87
Parental involvement	3	The parents of this student are actively involved in school.	.89

Note. TSR = teacher-student relationship.
[a] (Driessen et al., 2009, 2012).

Scale Item Reliability of Teacher Perceived Student Attributes

Primary School Achievement. When formulating track recommendations, Dutch primary school teachers can rely on different types of achievement measurements, such as students' standardised test results on reading comprehension, mathematics, as well as the results of the school leavers' test in sixth grade (final grade of primary education). These three measures were included as measures of students' (prior) primary school achievement.

Reading Comprehension and Mathematics Test. The reading comprehension and mathematics test are standardised tests, derived from the Monitoring and Evaluation system developed by Cito, the Dutch National Institute for Educational Measurement. These tests were administered during winter or early spring, halfway sixth grade. The reading comprehension test consists of multiple-choice items to measure students' ability to understand texts, with test scores range from -87 to 147. The mathematics test consists of multiple-choice and open items on ten subtopics (e.g. numbers, main arithmetic, fractions, ratios, proportions, and geometry), with test scores ranging from 0 to 168 (Cito,

2016; Driessen et al., 2009, 2012). Prior research (Feenstra et al., 2010; Janssen et al., 2010) provided support for high internal consistency (α > .80) and high validity of these tests.

School Leavers' Test. The school leavers' test (SLT), also known as the End of Primary Education Test, is a nationwide high-stakes test, administered in early spring of the final grade of primary education (Grade 6). Nowadays more school leavers' tests are available, but at the time of data collection only the test developed by Cito was available. The test used in the present research consists of 200 multiple-choice items, divided over three subtests: (a) 100 items on Dutch language (writing, decoding, reading comprehension, spelling/grammar), (b) 60 items on mathematics (arithmetic, algebra, geometry), and (c) 40 items on study skills (map-reading, interpreting study texts, information sources, graphs, diagrams, and tables). The results on these different subject domains are converted by Cito into a single test score ranging from 501 to 550 (van Boxtel et al., 2011). The SLT is calibrated each year to guarantee that students' average test scores are comparable across years. Previous research (College voor Toetsen en Examens, 2015; van Boxtel et al., 2011) reported that these tests have a high validity and high internal consistency (α > .95).

Secondary School Performance. Students' test results on reading comprehension and mathematics test, and their level of education in ninth grade (third grade of secondary education) were included as students' secondary school performance.

Reading Comprehension and Mathematics Test. Reading comprehension and mathematics test were assessed with standardised tests, developed by Cito for the COOL$^{5\text{-}18}$ project (Timmermans & Zijsling, 2014; Zijsling et al., 2014, 2017). The same tests are used for the different cohorts included in the present research. The reading comprehension test consists of 78 multiple-choice items to measure students' ability to understand texts, and the mathematics test consists of 60 multiple-choice items on four subtopics (i.e. numbers, measurement and geometry, relations, and ratios) (Kuyper et al., 2014; Zijsling et al., 2014, 2017). These tests were administered in Grade 9 between February and June using a structurally incomplete design. The items of the tests were divided into four modules based on content and difficulty. Each student made two modules that seemed suitable based on their secondary school level. The design ensured that a certain group of students always had a number of items in common with another group of students, the so-called anchor modules (Zijsling et al., 2014, 2017). Regarding the reading comprehension test, the items can be analysed using one underlying measurement scale because of the anchor modules (cf. Eggen, 2004; Kolen & Brennan, 1995). Based on description in the technical reports (Zijsling et al., 2014, 2017), the mathematics test scores were standardised per cohort to account for potential differences. The reliability of the reading comprehension and mathematics tests of these cohorts is respectively α = .92 and α = .94 (Timmermans & Zijsling, 2014; Zijsling et al., 2014, 2017).

Level of Secondary Education. Level of education in secondary education was recoded based on the adapted version of the educational ladder (the same scale as for the coding of track recommendations is used, see Table 2) to include the track as well as the grade

(school year) students attend, and to take into account grade repetition. Level of education was coded on a scale from 1 (first year practical training) to 8 (third year pre-university education), with a combination of adjacent tracks coded as the average of the two tracks. Level of education was considered as a continuous variable (cf. Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993).

Background Characteristics. Students' gender, SES, and migration background were measured in sixth grade (last year of primary education). This background information was retrieved from the school administrations.

Gender. A dichotomous dummy variable was created for students' gender; boys formed the reference group.

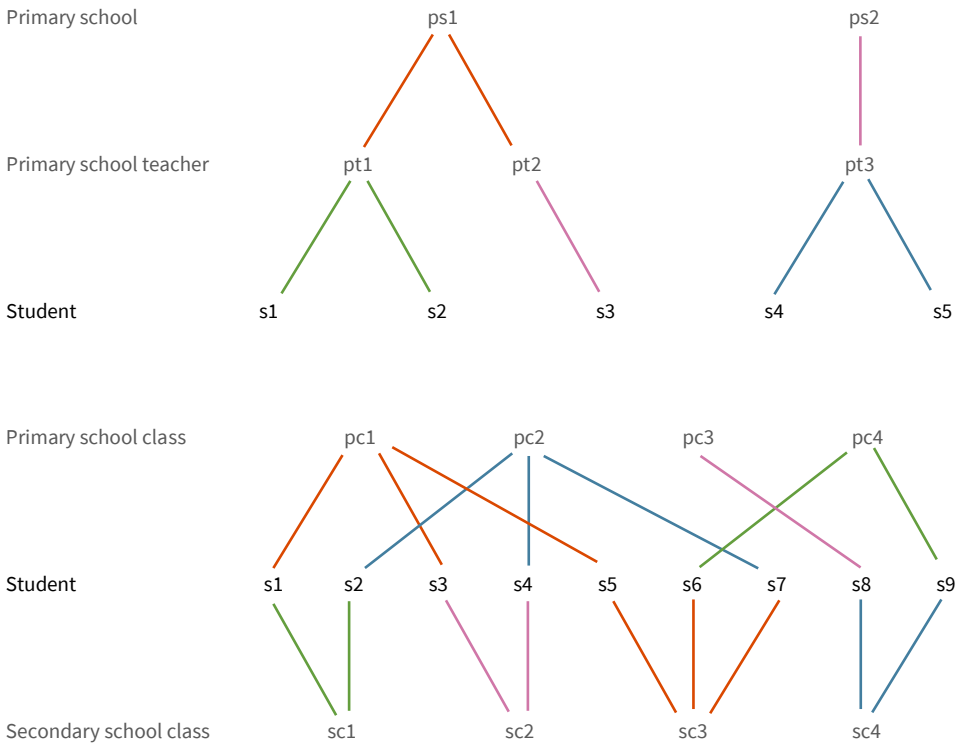
Socioeconomic Status (SES). Students' SES was based on the highest completed educational level of the student's parents or primary caregivers. SES was coded with four categories: (1) primary education, (2) lower vocational education, (3) senior secondary education, and (4) higher education or university. For the analyses, SES was considered as a continuous variable (Hox et al., 2018; Robitzsch, 2020).

Migration Background. A dichotomous dummy variable was created for students' migration background based on the birth countries of the student's parents. If there was a difference between both parents regarding their birth country, mothers' (or the female primary caregiver) birth country was used as indicator of students' migration background (Driessen et al., 2009, 2012). Students without a migration background formed the reference group.

Data Analyses

The data were analysed using Mplus Version 8.4 Mac (Muthén & Muthén, 2017). For RQ1, a two-level complex multilevel model was estimated (Burstein, 1980; Hox et al., 2018) using the robust maximum likelihood (MLR) estimator. Students (level 1) were nested in primary school teachers (level 2). Because most primary schools (62.08%) had only one Grade 6 teacher, there was no distinction between the teacher and the school level (cf. Timmermans et al., 2016, 2019). Therefore, the school level was only taken into account by a correction of the standard errors by using "type is complex" in Mplus. For RQ2, a cross-classified regression model using the Bayesian estimator was estimated. The data had a cross-classified nature as students from the same primary school went to different secondary schools, and students in the same secondary school were from different primary schools (see Figure 2 for a representation of data structures of both samples.). Hence, students (level 1) were cross-classified by primary school and secondary school (level 2). If students are in the same school cluster, either for primary or secondary school, the assumption of independence is violated. To account for possible influences of the data structure, it is important that cross-classification is incorporated in the analyses (Hox et al., 2018; O'Connell & McCoach, 2008; Snijders & Bosker, 2012).

Figure 2



Note. On top, the two-level complex multilevel data structure of the RQ1 sample. Students are nested in teachers, which are nested in primary schools. Below, the cross-classified data structure of the RQ2 sample. Students are clustered in both primary school classes and secondary school classes.

Data Structures of Samples RQ1 and RQ2

For both research questions, a series of models with the same predictors (i.e. students' prior achievement, teachers' perceptions of student attributes and students' background characteristics), but different outcome variables (i.e. track recommendation for RQ1, and secondary school performance for RQ2) were estimated. First, to examine the distribution of variance at both levels, an empty model with only the dependent variables was estimated (Model 0). Next, the three indicators of students' primary school achieve-

ment were added as fixed effects to examine whether students' achievement in primary education was a predictor of the outcome variables (Model 1). For the second research question, track recommendation was added to the model as a predictor. In Model 2, teacher perceived student attributes were added as fixed effects to examine whether these perceived attributes predicted the outcome variables beyond students' primary school achievement. Students' perceived work habits, classroom behaviour, underachievement, popularity, three attributes of the teacher-student relationship, and parental involvement were included as perceived student attributes. Next, students' background characteristics gender, SES, and migration background were added as fixed effects to examine whether they predicted the outcome variables beyond students' primary school achievement, and perceived student attributes (Model 3). For RQ2, an extra model (Model 3a) was estimated. Contrary to Model 3, Model 3a did not include students' track recommendation to be able to compare both models and shed light on the extent to which teacher perceptions of student attributes impact students' secondary school performance through their initial impact on track recommendations. Models 4 and 5 were only estimated for RQ1. Model 4 included the moderation of students' background characteristics gender, SES, and migration background on the perceived student attributes to examine whether teachers weighed the perceived student attributes differently for students with different backgrounds. Finally, a random-slopes model was estimated to examine whether the extent to which teachers relied on perceived student attributes differed between primary school teachers (Model 5). Explained variance (R2) was calculated for all models (cf. Snijders & Bosker, 2012). Effect sizes were based on standardised regression coefficients (STDYX) with 0.2, 0.5, and 0.8 as indicative of small, medium, and large effect sizes, respectively (Cohen, 1988). To facilitate the interpretation of the results, all continuous variables were standardised prior to being included in the analyses.

For RQ1, the percentage of missing data varied between 0.00% and 9.00%, except for the school leavers' test (20.84% missing data) (see Table 4 for the N of all variables). For RQ2, the percentage of missing data varied between 0% and 12%, except for the reading comprehension test in secondary education, track recommendation, and the school leavers' test (respectively 15.86%, 22.00%, and 34.17% missing data). Missing values on the SLT were due to the fact that administering the SLT was not mandatory (Driessen et al., 2009, 2012), and missing values on other variables were due to the fact that this information was just not available for those students. Students with missing values were not removed from the analyses. Instead, the missing data were accounted for by FIML (Full Information Maximum Likelihood) estimation (Schafer & Graham, 2002). The FIML method is based on the assumption that missing values are Missing at Random (MAR), and that the missing values can be predicted from the available data.

For RQ1, the Akaike Information Criterion (AIC; that is an estimator of prediction error; Akaike, 1987), and Bayesian Information Criterion (BIC; that is a combined model fit measure; Kass & Raftery, 1995; Raftery, 1993, 1995; Schwarz, 1978) were used to compare

the model fit of the models estimated. A lower value on AIC or BIC indicates a better fit. A difference of less than 2 points between AIC values of the models is considered as a better fit (Fabozzi et al., 2014; Olivares & Forero, 2010). A value of BIC between 2 and 6 indicates a good model fit, and a value between 6 and 10 is considered as a strong model fit (Fabozzi et al., 2014; Kass & Raftery, 1995; Olivares & Forero, 2010). In addition, the Standardised Root Mean Square Residual (SRMR; that is the average difference between the observed correlation and the model implied correlation between and SRMR within given the multilevel structure of the data; Hu & Bentler, 1995) was added. Because the SRMR is an absolute measure of fit, a value of zero indicates perfect fit. A value less than .08 is generally considered a good fit (Byrne, 1994; Hu & Bentler, 1999). For RQ2, not all fit statistics that were used for RQ1 were available due to a different type of analysis used. Therefore, for RQ2, the Posterior Predictive P-value (PPP; that is a general discrepancy measure; Meng, 1994), DIC (Deviance Information Criterion; that is a combined measure of model fit; Hamaker et al., 2011; Spiegelhalter et al., 2002), and pD (estimated number of parameters) were added for complexity of the model (Spiegelhalter et al., 2002). A PPP of .05 is considered as an excellent fit, and larger than .05, but less than .10 indicates a good fit (Cain & Zhang, 2019; Gelman, 2013; Hjort et al., 2006). A lower value on DIC indicates a better fit (Cain & Zhang, 2019; S.-Y. Lee & Song, 2012; Spiegelhalter et al., 2002; van der Linde, 2004), while a larger pD is considered as a better (or easier) model fit (Peugh & Feldon, 2020).

Since somewhat different datasets were used for both research questions, and the sample for RQ2 is not a total subsample of RQ1, an attrition analysis was conducted to see if the datasets were comparable. Results showed that there were some differences between both datasets, albeit the differences were mostly small-sized (see Table 4).

Results

The descriptive statistics of the variables are presented in Table 4. The correlations between track recommendations, achievement in primary and secondary education, level of secondary education in ninth grade, teachers' perceptions of student attributes, and student background characteristics are presented in Table 5. Overall, small to moderate significant positive correlations were reported between perceived student attributes and track recommendations or secondary school performance (*r* between .07 and .44), and negative correlations were found for perceived underachievement, and the perceived dependency and conflict attributes of teacher-student relationship (*r* between -.33 and -.14). Students' gender, SES, and migration background were weakly to moderately related to other variables (*r* is between -.20 and .39).

Table 4

	Sample for RQ1						Sample for RQ2					
	N	M	SD	Min.	Max.	%	N	M	SD	Min.	Max.	%
Track recommendation	17,953	4.14[a]	1.44	1	6		3,237	4.43[a]	1.33	1	6	
Primary school achievement												
Reading comprehension test	16,663	54.93[a]	17.06	-65	147		3,866	58.01[a]	16.74	5	147	
Mathematics test	16,343	114.29[a]	11.42	9	169		3,820	116.49[a]	10.26	37	160	
School leavers' test	14,211	533.75[a]	10.00	501	550		2,732	535.48[a]	9.62	503	550	
Perceived student attributes												
Work habits	16,856	3.72[a]	0.93	1	5		3,659	3.85[a]	0.87	1	5	
Classroom behaviour	17,036	4.01[a]	0.85	1	5		3,687	4.10[a]	0.81	1	5	
Underachievement	16,946	2.68[a]	0.94	1	5		3,663	2.60[a]	0.92	1	5	
Popularity	16,831	3.89[b]	0.82	1	5		3,651	3.95[b]	0.78	1	5	
Teacher-student relationship												
Dependency	17,047	2.29[a]	0.94	1	5		3,691	2.22[a]	0.89	1	5	
Conflict	17,044	1.93	0.96	1	5		3,691	1.89	0.90	1	5	
Closeness	17,027	3.84	0.68	1	5		3,689	3.84	0.65	1	5	
Parental involvement	16,991	3.81[a]	0.93	1	5		3,671	3.90[a]	0.87	1	5	
Background characteristics												
Gender – Boys	8,807					49.71%	2,006					48.34%
Gender – Girls	8,910					50.29%	2,144					51.66%
Socioeconomic status	77,304	2.94[a]	0.91	1	4		4,005	3.06[a]	0.86	1	4	
Migration background – No	13,680					77.65%	3,421					83.64%
Migration background – Yes	3,937					22.35%	669					16.36%
Secondary school performance												
Reading comprehension test							3,492	0.43	0.45	-2.16	2.19	
Mathematics test							3,743	0.25	0.37	-0.94	1.68	
Level of education							4,150	6.46	1.27	3	8	

Note. [a] = within a row, means with superscript differ significantly ($p < .001$) between the samples of RQ1 and RQ2; [b] = within a row, means with superscript differ significantly ($p < .05$) between the samples of RQ1 and RQ2. All differences are small in effect size (Cohen's $d \leq .21$).

Descriptive Statistics of Track Recommendation, Secondary School Performance, Primary School Achievement, Perceived Student Attributes and Student Background Characteristics

Table 5

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Track Recommendation																	
2. Reading comprehension PE	.70***																
3. Mathematics PE	.73***	.56***															
4. School leavers' test	.88***	.74***	.76***														
5. Work habits	.40***	.38***	.33***	.40***													
6. Classroom behaviour	.18***	.19***	.13***	.19***	.49***												
7. Underachievement	-.19***	-.24***	-.18***	-.24***	-.44***	-.29***											
8. Popularity	.16***	.15***	.12***	.15***	.30***	.31***	-.19***										
9. TSR: Dependency	-.31***	-.27***	-.26***	-.31***	-.34***	-.30***	.23***	-.26***									
10. TSR: Conflict	-.20***	-.20***	-.15***	-.21***	-.48***	-.68***	.32***	-.35***	.45***								
11. TSR: Closeness	.07***	.10***	.04***	.07***	.26***	.27***	-.20***	.31***	-.02*	-.33***							
12. Parental involvement	.40***	.35***	.29***	.39***	.32***	.26***	-.21***	.24***	-.18***	-.26***	.25***						
13. Gender (0 = boys)	-.02**	.07***	-.15***	-.03**	.26***	.18***	-.15***	.06***	-.01	-.19***	.18***	.03***					
14. SES	.38***	.34***	.27***	.37***	.14***	.12***	-.09***	.07***	-.14***	-.12***	.06***	.39***	-.01				
15. Migration background	-.18***	-.20***	-.13***	-.20***	-.03***	-.08***	.08***	-.03***	.06***	.08***	-.08***	-.29***	.00	-.40***			
16. Reading comprehension SE	.64***	.63***	.51***	.67***	.34***	.22***	-.23***	.09***	-.27***	-.20***	.05*	.30***	.10***	.35***	-.18***		
17. Mathematics SE	.68***	.54***	.67***	.70***	.30***	.16***	-.20***	.07**	-.27***	-.14***	.02	.29***	-.11***	.34***	-.18***	.65***	
18. Level of Education	.84***	.70***	.67***	.82***	.44***	.23***	-.29***	.17***	-.33***	-.24***	.09***	.41***	.07***	.44***	-.22***	.72***	.72***

Note. SES = socioeconomic status; PE = primary education; SE = secondary education; TSR = teacher-student relationship.

*$p < .05$; **$p < .01$; ***$p < .001$.

Correlations Between Track Recommendation, Level of Education, Primary School Achievement, Secondary School Performance, Perceived Student Attributes and Background Characteristics

Predictive Value of Teacher Perceived Student Attributes for Track Recommendation (RQ1)

The first aim of the present study was to examine the extent to which teachers consider student attributes on top of students' prior achievement when formulating track recommendations. The results of the multilevel regression models are presented in Table 6 (Models 0 to 3) and Table 7 (Models 4 and 5).[15] The results of the empty model without predictors (Model 0) revealed that 13.51% of the variance in track recommendations was attributable to factors at the teacher level, and the remaining 86.49% to factors at the student level and error. The results of Model 1 (Table 6) indicated that students with higher primary school achievement, i.e. higher reading comprehension, mathematics, and school leavers' test achievement, received higher track recommendations. Together these achievement variables explained 78.05% of the variance in track recommendations.

Model 2 (Table 6), in which teacher perceived student attributes were added, revealed that teacher perceived work habits, classroom behaviour, underachievement, popularity, the dependency attribute of the teacher-student relationship, and parental involvement were significantly positively or negatively, depending on the specific perceived student attribute, related to track recommendations beyond the effects of students' primary school achievement, while teacher perceived conflict and closeness attributes of the teacher-student relationships were not. Students who – according to their teacher – had stronger work habits, displayed more negative classroom behaviour, underachieved, were more popular, had a less dependent relation with the teacher, or had more involved parents, received slightly higher track recommendations. The effect sizes were all very small (as indicated by $b^* < .07$). Together, teachers' perceptions of student attributes explained only 0.70% of the total variance in track recommendations beyond prior achievement.

Model 3 (Table 6), in which students' gender, SES, and migration background were added, revealed a significant positive effect of gender and SES on track recommendations, indicating that after taking into account students' primary school achievement and teacher perceived student attributes, girls and higher-SES students received higher track recommendations than boys and lower-SES students. For students' migration background, no significant relation with track recommendations was found. Altogether, students' background characteristics explained only 0.30% of the total variance in track recommendations.

15 Some small differences were found between the two cohorts. These differences are reported in the notes under Table 6 and Table 7. The separate results for each cohort are available upon request from the first author.

Table 6

Standardised Estimates of Multilevel Models 0 to 3 Predicting Track Recommendations With Prior Achievement, Perceived Student Attributes, and Background Characteristics Gender, SES and Migration Background

		Model 0: Empty				Model 1: Prior achievement				Model 2: + Perceived student attributes				Model 3: + Gender + SES + migration background			
				95% CI				95% CI				95% CI				95% CI	
		*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL
Fixed effects	Intercept	-.01	.02	-.04	.03	.02*	.01	.00	.04	.02*	.01	.00	.04	.01	.01	-.02	.02
Prior achievement	Reading comprehension test					.17***	.01	.14	.19	.15***	.01	.13	.18	.15***	.01	.12	.17
	Mathematics test					.22***	.02	.18	.25	.21***	.02	.18	.24	.22***	.02	.18	.25
	School leavers' test					.62***	.02	.57	.66	.58***	.02	.53	.62	.57***	.02	.52	.61
Perceived student attributes	Work habits									.07***	.01	.06	.08	.07***	.01	.06	.08
	Classroom behaviour									-.02**	.01	-.03	-.01	-.02***	.01	-.03	-.01
	Underachievement									.05***	.01	.04	.06	.05***	.01	.04	.06
	Popularity									.01*[C]	.00	.00	.02	.01**[F]	.00	.00	.02
	TSR: Dependency									-.05***	.01	-.06	-.04	-.05***	.01	-.06	-.04
	TSR: Conflict									.00	.01	-.01	.02	.00	.01	-.01	.02
	TSR: Closeness									.00	.01	-.01	.01	.00	.01	-.01	.01
	Parental involvement									.07***	.01	.06	.08	.06***	.01	.05	.07
Background characteristics	Gender (girls)													.03**[F]	.01	.01	.04
	SES													.05***	.01	.04	.06
	Migration background (yes)													.01	.01	-.01	.03
Variance	Class level	.14***	.01	.11	.16	.04***	.00	.04	.05	.04***	.00	.04	.05	.04***	.00	.03	.04
	Student level	.87***	.01	.84	.90	.18***	.01	.17	.19	.17***	.01	.16	.18	.17***	.01	.16	.18
	R^2						.781				.787				.790		
Model Fit	-2LL		49,808.08				160,708.04				545,281.86				639,090.82		
	AIC		49,814.08				160,732.04				545,353.86				639,180.83		
	BIC		49,837.47				160,825.58				545,634.50				639,531.62		
	SRMR within		.00				.41				.28				.25		
	SRMR between		.00				.01				.02				.02		
Number of	Schools		674				674				674				674		
	Classes		1,105				1,105				1,105				1,105		
	Students		17,953				17,953				17,953				17,953		

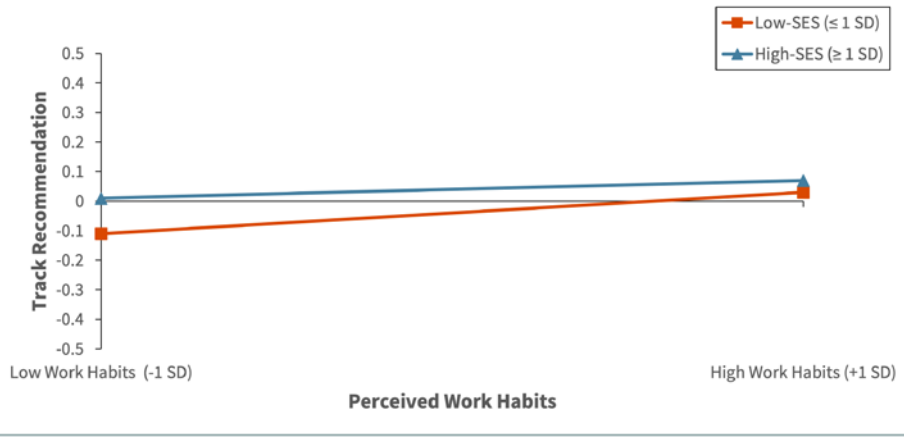
Note. CI = confidence interval; LL = lower limit; UL = upper limit; TSR = teacher-student relationship; SES = socioeconomic status; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; SRMR = Standardised Root Mean Square Residual.

[C] = variable is significant in the overall analyses, but not in Cohort 1 and Cohort 2 separately; [F] = variable is significant in the overall analyses and Cohort 2, but not in Cohort 1.

*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

By adding interactions between students' background characteristics, and teacher perceived student attributes to the multilevel model, it was investigated whether student attributes were weighed differently for students with different background characteristics. The results of Model 4 (Table 7) indicated some significant interactions effects for SES and migration background. For gender, no significant interaction effects were found. For SES, it was found that perceived work habits were a slightly stronger predictor for track recommendations for lower-SES students than for higher-SES students (see Figure 3). Lower-SES students with poorer work habits received lower track recommendations than higher-SES students with equally poor work habits, keeping all other variables constant. Nevertheless, it is important to note that the overall differences in track recommendations between lower- and higher-SES students (which are mostly due to achievement differences) are much larger in comparison. For migration background, it was found that when formulating track recommendations, teachers considered perceived closeness to a stronger extent for students with a migrations background than without (see Figure 4). Hence, students with a migration background who were perceived to have a close relationship with the teacher received higher track recommendations than students without a migration background with the same level of perceived closeness, keeping all other variables constant. Also in this case, it is important to note that the overall differences between students with and without a migration background in track recommendations are actually much larger due to achievement differences. Together the moderation effects of students' background characteristics explained an additional 1.09% of the total variance in track recommendations.

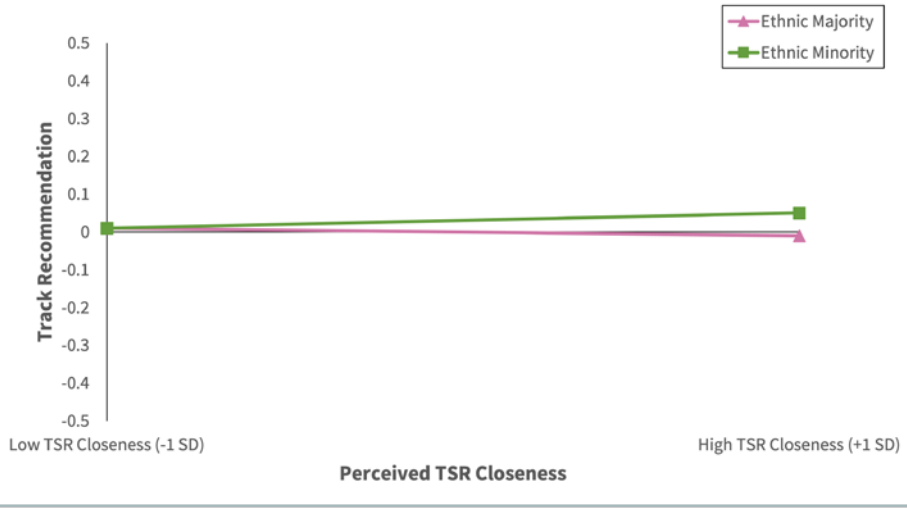
Figure 3



Note. SES = socioeconomic status.

Multilevel Regression Lines for the Effect of Teacher Perceived Student Work Habits on Track Recommendations for Students With Different SES Backgrounds

Figure 4



Note. TSR = teacher-student relationship.

Multilevel Regression Lines for the Effect of Teacher Perceived Closeness Attribute of the Teacher-Student Relationship on Track Recommendations for Students With Different Migration Backgrounds

In addition, when formulating track recommendations, the extent to which teachers consider students' perceived student attributes may differ. The results of Model 5 (Table 7) indicated significant random slopes for perceived work habits and parental involvement, but not for the other variables. Hence, teachers significantly differed from one another in the extent to which they weighed their perceptions of students' work habits and parental involvement. These differences are displayed in Figure 5. For this figure, the correlations between perceived student attributes and track recommendation were calculated for each primary school teacher, while controlling for student achievement and background variables. For sake of trustworthiness of the correlations, only classes with N > 20 students were included. In Figure 5, these correlations are displayed in boxplots to show the variations in the estimates across teachers. On average, teachers weighed students' work habits and parental involvement positively. However, as Figure 5 shows, the distribution ranged from negative to positive, indicating that some teachers weighed these perceived student attributes negatively in their track recommendations, while other teachers weighed them positively. Hence, some teachers tended to formulate higher track recommendations for students whom they perceived as having strong work habits or highly involved parents, whereas other teachers formulated lower track recommendations for these students.

Despite these differences, the inclusion of these random slopes explained only 1.29% to the total variance in track recommendations.

Overall, when looking at the model fit, the between model has an acceptable fit according to the SRMR, but the 'comparative fit indices' AIC and BIC show that the successive models always have a poorer fit (that is true for Models 0 to 3). This confirms the finding that the added variables do not improve the models and explain little additional variance.

Table 7

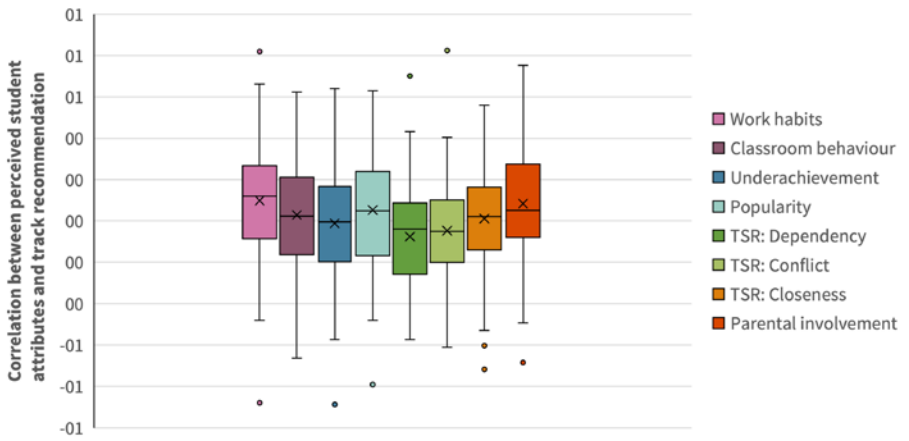
	Model 4: + Moderation of gender + SES + migration background				Model 5: + Random slopes of prior achievement + perceived student attributes			
			95% CI				95% CI	
	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL
Fixed effects								
Intercept	.00	.01	-.02	.03	.00	.01	-.02	.03
Prior achievement								
Reading comprehension test	.10***	.01	.09	.12	.11***	.01	.09	.12
Mathematics test	.16***	.01	.14	.18	.16***	.01	.14	.18
School leavers' test	.63***	.02	.60	.66	.63***	.01	.60	.66
Perceived student attributes								
Work habits	.05***	.01	.03	.07	.05***	.01	.03	.07
Classroom behaviour	-.02*[c]	.01	-.04	.00	-.02	.01	-.04	.00
Underachievement	.05***	.01	.03	.06	.05***	.01	.03	.06
Popularity	.01[b]	.01	.00	.03	.02*[f]	.01	.00	.03
Teacher-student relationship (TSR)								
Dependency	-.03***	.01	-.04	-.01	-.03***	.01	-.04	-.02
Conflict	-.01	.01	-.03	.01	-.01	.01	-.03	.01
Closeness	-.01	.01	-.03	.00	-.02	.01	-.03	.00
Parental involvement	.04***	.01	.03	.06	.04***	.01	.03	.06
Background characteristics								
Gender (girls)	.04***[f]	.01	.02	.06	.04***	.01	.03	.06
SES	.04***	.01	.03	.05	.04***	.01	.03	.05
Migration background (yes)	.03a	.01	.00	.05	.03[a]	.01	.00	.05
Moderation of gender								
Gender x Work habits	.00	.01	-.02	.02	.00	.01	-.02	.03
Gender x Classroom behaviour	.01	.01	-.02	.03	.01	.01	-.02	.03
Gender x Underachievement	.00	.01	-.02	.02	.00	.01	-.02	.02
Gender x Popularity	.00	.01	-.02	.02	.00	.01	-.02	.01
Gender x TSR: Dependency	-.01	.01	-.03	.01	-.01	.01	-.03	.01
Gender x TSR: Conflict	.02	.01	-.01	.04	.01	.01	-.01	.04
Gender x TSR: Closeness	-.01	.01	-.02	.01	.00	.01	-.02	.02
Gender x Parental involvement	-.01	.01	-.02	.01	-.01	.01	-.02	.01

Moderation of SES								
SES x Work habits	-.02***	.01	-.04	-.01	-.02***	.01	-.04	-.01
SES x Classroom behaviour	.00	.01	-.02	.01	.00	.01	-.02	.01
SES x Underachievement	-.01	.01	-.02	.00	-.01	.01	-.02	.00
SES x Popularity	.00	.01	-.01	.01	.00	.01	-.01	.01
SES x TSR: Dependency	.01	.01	-.01	.02	.01	.01	.00	.02
SES x TSR: Conflict	-.01	.01	-.02	.00	-.01	.01	-.03	.00
SES x TSR: Closeness	.01	.01	-.01	.01	.01	.01	.00	.02
SES x Parental involvement	.01	.01	.01	.01	.01	.01	-.01	.02
Moderation of migration background								
Migration background x Work habits	.01	.02	-.02	.05	.01	.02	-.02	.04
Migration background x Classroom behaviour	.02	.02	.00	.05	.03	.02	.00	.06
Migration background x Underachievement	.00	.01	-.03	.02	.00	.01	-.03	.02
Migration background x Popularity	.00	.01	-.02	.03	.00	.01	-.03	.02
Migration background x TSR: Dependency	.01	.01	-.02	.03	.01	.01	-.02	.04
Migration background x TSR: Conflict	.02	.02	-.01	.06	.03	.02	-.01	.06
Migration background x TSR: Closeness	.03*[c]	.01	.00	.06	.03*[f]	.01	.00	.06
Migration background x Parental involvement	.02[a]	.01	-.01	.04	.02[a]	.01	.00	.05
Random slopes								
Work habits					.00*	.00	.00	.01
Classroom behaviour					.00b	.00	.00	.00
Underachievement					.00b	.00	.00	.00
Popularity					.00b	.00	.00	.00
TSR: Dependency					.00	.00	.00	.00
TSR: Conflict					.00	.00	.00	.00
TSR: Closeness					.00	.00	.00	.00
Parental involvement					.00**[e]	.00	.00	.01
Variance								
Class level		.00	.03		.03***	.00	.03	.04
Student level		.01	.15		.15***	.01	.14	.16
R2	.801				.814			
Model fit								
-2LL	12,383.37				12,289.75			
AIC	12,465.37				12,387.75			
BIC	12,765.19				12,746.08			
Number of schools	503				503			
Number of classes	768				768			
Number of students	1,1078				1,1078			

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; TSR = teacher-student relationship; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.
[a] = variable is significant in Cohort 1, but not in Cohort 2; [b] = variable is significant in Cohort 2, but not in Cohort 1; [c] = variable is significant in the overall analyses, but not in Cohort 1 and Cohort 2 separately; [d] = variable is significant in Cohort 1 and Cohort 2, but not in overall analyses; e = variable is significant in the overall analyses and Cohort 1, but not in Cohort 2; [f] = variable is significant in the overall analyses and Cohort 2, but not in Cohort 1.
*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Standardised Estimates of Multilevel Models 4 and 5 Predicting Track Recommendations With (Random Slopes of) Prior Achievement, Perceived Student Attributes, and (Moderation of) Background Characteristics Gender, SES and Migration Background

Figure 5



Note. TSR = teacher-student relationship.

Differences Among Teachers in the Correlations Between Teacher Perceptions of Student Attributes and Track Recommendation

This figure shows boxplots of the correlations between perceived student attributes and track recommendations for different primary school teachers while controlling for student achievement and student background variables, thereby showing the variations between teachers in how they weigh the different student attributes in their track recommendations.

Predictive Value of Teacher Perceived Student Attributes for Secondary School Performance (RQ2)

Our second aim was to investigate the predictive value of students' prior achievement, track recommendations, teacher perceptions of student attributes, and background characteristics for students' secondary school performance. The results of the multilevel models are presented in Tables 8 to 10. The results of the empty model without predictors (Model 0) revealed that 6.08% of the variance in reading comprehension achievement was situated at primary school class level, and 60.42% to factors at secondary school class level. For the variance in mathematics achievement, it was respectively 6.15%, and 57.76%, and for the variance in level of education, it was it was respectively 3.73%, and 91.81%.

In Model 1 (of Tables 8 to 10) students' achievement in primary education was added to the model. Findings revealed that students who performed better in primary education, i.e. higher reading comprehension, mathematics, and school leavers' test achievement, and

who received a higher track recommendation, had higher secondary school performance, i.e. had higher reading comprehension and mathematics achievement in secondary education, and attended a higher level of secondary education, even after accounting for differences in track recommendations. The variance in students' secondary school performance was primarily explained by students' primary school achievement and track recommendations ($R_{Reading}^2$ = 50.19%; $R_{Mathematics}^2$ = 55.32%; $R_{LevelofEducation}^2$ = 40.67%).

Teacher perceptions of student attributes were added in Model 2 (of Tables 8 to 10). Students who were perceived as having better work habits performed better in reading comprehension in secondary education and attended a higher level of education. Students whose classroom behaviour was perceived as positive by their primary school teacher, performed better in reading comprehension. Furthermore, students who were perceived by their primary school teacher as more popular among fellow students or as more dependent of their teacher performed lower in mathematics. Finally, students whose parents' involvement was perceived as high by the primary school teacher, performed better in reading comprehension and mathematics in secondary education and attended a higher level of education. All effect sizes were small (as indicated by * < .05). On top of students' primary school achievement and track recommendations, teachers' perceptions of student attributes explained almost no additional variance of the variance in secondary school performance ($\Delta R_{Reading}^2$ = 0.29%; $\Delta R_{Mathematics}^2$= 0.20%; $\Delta R_{LevelofEducation}^2$ = 1.82%).

Model 3 (of Tables 8 to 10), to which students' background characteristics gender, SES, and migration background were added, showed that, on top of students' prior achievement and teacher perceived student attributes, students' background characteristics explained respectively 6.66%, 1.17%, and 2.46% of the variance in reading comprehension, mathematics, and level of education. Girls and higher-SES students performed better on all three achievement indicators in secondary education than their male and lower-SES peers. Additionally, students without a migration background had a higher educational level in Grade 9 than students with a migration background, but only when the track recommendation was taken into account. If Model 3 is compared with Model 3a (of Table 10), the findings showed that in the full model without the variable track recommendation (Model 3a), students' migration background was not significantly related to students' level of secondary education, while in the model with track recommendation (Model 3), students' migration background was significantly related to students' level of secondary education. Hence, students with and without a migration background with similar prior achievement and track recommendations in Grade 6, do not obtain the same educational level in Grade 9.

Another difference between Model 3 versus Model 3a of Table 8 is the finding that perceived work habits and closeness were respectively positively and negatively related to students' secondary school reading comprehension in the model without track recommendations (Model 3a), but not associated with secondary school achievement in the model with track recommendations (Model 3). These findings indicated that without controlling for students' track recommendation, some perceived attributes did have a small effect on

students' secondary school performance. However, for students in similar tracks, these attributes do not seem predictive of secondary school performance.

Overall, when looking at the model fit, the fit indices PPP, DIC and pD show that the successive models always have a poorer fit. This indicates that the added variables do not improve the models and explain little additional variance.

"De ene 2 is niet de andere 2. Hoe je een leerling beleeft in de klas: stukje algemene ontwikkeling, aanwezigheid, motivatie, interesse, samenwerken. Kijken naar de vaardigheden: iemand die in zichzelf is, niet durft, dat is een heel ander kind dan een die initiatief neemt, communicatief vaardig is, een grote algemene ontwikkeling heeft, en naar je toekomt om te zeggen dat hij dingen nog niet begrijpt, moeilijk vindt en vraagt of je extra oefenstof hebt: die wil ervoor gaan. Het is niet erg als een kind dat niet doet, maar het heeft wel invloed op het advies: het kan soms net het verschil zijn tussen havo en vwo."

Table 8

Standardised Estimates of Multilevel Model 0 to 3 Predicting Secondary School Reading Comprehension Achievement With Primary School Achievement, Track Recommendation, Perceived Student Attributes, and Background Characteristics Gender, SES, and Migration Background

		Model 0: Empty				Model 1: Primary school achievement				Model 2: + Perceived student attributes				Model 3: + Background characteristics				Model 3a: Without track recommendation			
				95% CI				95% CI				95% CI				95% CI				95% CI	
		*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL
Fixed effects	Intercept	-.04*	.03	-.10	.00	-.02	.02	-.06	.02	-.02	.02	-.05	.02	-.10***	.03	-.14	-.05	-.10***	.02	-.14	-.05
Primary school achievement	Reading comprehension					.34***	.02	.31	.38	.33***	.02	.29	.36	.31***	.02	.27	.34	.34***	.02	.30	.37
	Mathematics					.07***	.02	.03	.10	.06***	.02	.03	.09	.09***	.02	.06	.12	.12***	.02	.08	.15
	School leavers' test					.14***	.03	.08	.19	.12***	.03	.07	.17	.12***	.03	.07	.17	.24***	.02	.20	.28
	Track recommendation					.24***	.03	.20	.29	.23***	.02	.19	.27	.21***	.03	.16	.26				
Perceived student attributes	Work habits									.04**[C]	.02	.01	.08	.03	.02	.00	.07	.04***	.01	.02	.07
	Classroom behaviour									.05**	.02	.02	.08	.05**[E]	.02	.02	.08	.05***[C]	.02	.01	.07
	Underachievement									-.02	.01	-.04	.01	-.02	.01	-.05	.01	-.01	.01	-.04	.02
	Popularity									-.02[A]	.01	-.05	.01	-.02	.01	-.05	.01	-.02	.02	-.05	.01
	TSR: Dependency									-.02	.02	-.05	.01	-.02	.02	-.05	.00	-.03	.02	-.06	.00
	TSR: Conflict									-.01	.02	-.05	.02	-.01	.02	-.04	.03	.00	.02	.47	-.04
	TSR: Closeness									-.02	.02	-.05	.01	-.03[A]	.02	-.05	.00	-.03*	.01	-.06	.00
	Parental involvement									.04***	.01	.02	.07	.03*[E]	.02	.00	.06	.04***	.01	.01	.07
Background characteristics	Gender (girls)													.16***	.03	.11	.21	.16***	.03	.10	.20
	SES													.09***	.02	.06	.11	.10***	.01	.07	.12
	Migration background (yes)													.00	.04	-.09	.07	.00	.50	-.06	.07
Variance	Secondary school class level	.63***	.04	.56	.71	.11***	.01	.09	.14	.12***	.01	.09	.14	.10***	.02	.08	.14	.10***	.01	.08	.12
	Primary school class level	.06***	.01	.05	.08	.06***	.01	.04	.07	.06***	.01	.04	.08	.06***	.01	.04	.08	.06***	.01	.04	.07
	Student level	.35***	.01	.33	.37	.35***	.01	.33	.37	.34***	.01	.32	.36	.34***	.01	.32	.36	.35***	.01	.33	.37
	R^2						.502				.505				.518				.518		
Model fit	PPP		.000				.000				.000				.000				.000		
	DIC		16,165.14				53,489.14				136,830.35				157,573.66				148,523.78		
	p^D		3,771.61				2,945.65				3,014.55				3,005.43				2,990.06		
Number of	Secondary school classes		1,289				1,289				1,289				1,289				1,289		
	Primary school classes		741				741				741				741				741		
	Students		4,150				4,150				4,150				4,150				4,150		

Note. CI = confidence interval; LL = lower limit; UL = upper limit; TSR = teacher-student relationship; SES = socioeconomic status; PPP = posterior predictive p-value; DIC = deviance information criterion; p^D = estimated number of parameters.

[A] = variable is significant in Cohort 1, but not in Cohort 2; [C] = variable is significant in the overall analyses, but not in Cohort 1 and Cohort 2 separately; [E] = variable is significant in the overall analyses and Cohort 1, but not in Cohort 2.

*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Table 9

		Model 0: Empty				Model 1: Primary school achievement				Model 2: + Perceived student attributes				Model 3: + Background characteristics				Model 3a: Without track recommendation			
				95% CI				95% CI				95% CI				95% CI				95% CI	
		*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL
Fixed effects	Intercept	-.07*	.03	-.13	-.01	-.04*	.02	-.07	.00	-.03*	.02	-.07	.00	.06***	.02	.02	.10	.06***	.02	.03	.10
Primary school achievement	Reading comprehension					.08***	.02	.05	.11	.07***	.02	.04	.11	.09***	.02	.06	.12	.11***	.02	.08	.14
	Mathematics					.41***	.02	.38	.44	.40***	.02	.38	.43	.37***	.02	.34	.41	.41***	.02	.37	.44
	School leavers' test					.11***	.02	.07	.15	.11***	.02	.06	.14	.10***	.02	.06	.15	.22***	.02	.19	.26
	Track recommendation					.25***	.02	.21	.29	.23***	.02	.19	.27	.22***	.02	.18	.26				
Perceived student attributes	Work habits									.02	.02	-.01	.05	.05***[E]	.02	.02	.08	.06***[E]	.02	.03	.09
	Classroom behaviour									.02	.02	-.02	.04	.02	.02	-.01	.05	.02	.02	-.01	.05
	Underachievement									.01	.01	-.02	.03	.00	.02	-.03	.03	.00	.01	-.03	.03
	Popularity									-.05***	.01	-.07	-.03	-.05***	.01	-.08	-.02	-.05***	.01	-.08	-.03
	TSR: Dependency									-.05***	.01	-.08	-.03	-.05***[E]	.01	-.07	-.03	-.06***[E]	.01	-.09	-.04
	TSR: Conflict									.01	.02	-.02	.04	.01	.02	-.02	.04	.01	.02	-.02	.05
	TSR: Closeness									-.01	.01	-.03	.02	.01	.01	-.02	.03	.00	.01	-.03	.03
	Parental involvement									.06***	.01	.03	.08	.04**	.01	.01	.07	.05***	.01	.02	.07
Background characteristics	Gender (girls)													-.18***	.02	-.22	-.14	-.18***	.02	-.22	-.14
	SES													.06***[E]	.01	.04	.08	.08***[E]	.01	.05	.10
	Migration background (yes)													-.01	.04	-.08	.06	-.01	.03	-.09	.06
Variance	Secondary school class level	.59***	.03	.53	.66	.09***	.01	.07	.09	.09***	.01	.06	.11	.09***	.01	.07	.11	.08***	.01	.07	.11
	Primary school class level	.06***	.01	.05	.09	.06***	.01	.04	.06	.06***	.01	.04	.07	.05***	.01	.04	.06	.05***	.01	.03	.07
	Student level	.37***	.01	.35	.39	.32***	.01	.30	.30	.31***	.01	.30	.33	.31***	.01	.29	.33	.32***	.01	.29	.34
	R^2						.553				.555				.567				.562		
Model fit	PPP			.000			.000				.000				.000				.000		
	DIC			16,165.14			53,489.14				136,830.35				157,573.66				148,523.78		
	p^D			3,771.61			2,945.65				3,014.55				3,005.43				2,990.06		
Number of	Secondary school classes			1,289			1,289				1,289				1,289				1,289		
	Primary school classes			741			741				741				741				741		
	Students			4,150			4,150				4,150				4,150				4,150		

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; TSR: teacher-student relationship; PPP = posterior predictive p-value; DIC = deviance information criterion; p^D = estimated number of parameters.

[E] = variable is significant in the overall analyses and Cohort 1, but not in Cohort 2.

*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Standardised Estimates of Multilevel Model 0 to 3 Predicting Secondary School Mathematics Achievement With Primary School Achievement, Track Recommendation, Perceived Student Attributes, and Background Characteristics Gender, SES, and Migration Background

Table 10

Standardised Estimates of Multilevel Model 0 to 3 Predicting Secondary School Level of Education With Primary School Achievement, Track Recommendation, Perceived Student Attributes, and Background Characteristics Gender, SES, and Migration Background

		Model 0: Empty				Model 1: Primary school achievement				Model 2: + Perceived student attributes				Model 3: + Background characteristics				Model 3a: Without track recommendation			
				95% CI				95% CI				95% CI				95% CI				95% CI	
		*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL	*b**	SE_b	LL	UL
Fixed effects	Intercept	-.09***	.03	-.15	-.03	-.06*	.02	-.11	-.02	-.06**	.02	-.10	-.02	-.09***	.02	-.14	-.04	-.09***	.02	-.13	-.04
Primary school achievement	Reading comprehension					.09***	.01	.07	.10	.08***	.01	.06	.09	.08***	.01	.06	.09	.09***	.01	.07	.10
	Mathematics					.08***	.01	.06	.09	.08***	.01	.06	.09	.09***	.01	.07	.10	.10***	.01	.09	.11
	School leavers' test					.03*	.01	.01	.05	.03**	.01	.01	.05	.03***	.01	.01	.05	.10***	.01	.07	.11
	Track recommendation					.15***	.01	.12	.17	.13***	.01	.12	.15	.14***	.01	.12	.16				
Perceived student attributes	Work habits									.03***	.01	.01	.04	.02***[C]	.01	.01	.04	.03***[C]	.01	.01	.04
	Classroom behaviour									.01	.01	.00	.02	.01	.01	.00	.02	.01	.01	-.01	.02
	Underachievement									.00	.01	-.01	.01	.00	.01	-.01	.01	.00	.01	-.01	.02
	Popularity									.01	.01	.00	.02	.01	.01	.00	.02	.01	.01	-.01	.02
	TSR: Dependency									-.01	.01	-.02	.01	-.01*[E]	.01	-.02	.00	-.02***	.01	-.03	-.01
	TSR: Conflict									-.01	.01	-.03	.00	-.01	.01	-.02	.01	.00	.01	-.02	.01
	TSR: Closeness									-.01	.01	-.02	.00	-.01*[C]	.01	-.02	.00	-.01*[C]	.01	-.03	.00
	Parental involvement									.04***	.01	.03	.05	.03***[E]	.01	.02	.05	.04***[e]	.01	.02	.05
Background characteristics	Gender (girls)													.04***[F]	.01	.02	.06	.04***[F]	.01	.02	.06
	SES													.04***[C]	.01	.03	.05	.04***[C]	.01	.03	.05
	Migration background (yes)													.05**[C]	.02	.01	.07	.03	.02	.00	.07
Variance	Secondary school class level	1.01***	.04	.93	1.09	.57***	.03	.52	.63	.55***	.03	.51	.61	.53***	.03	.47	.58	.58***	.03	.52	.63
	Primary school class level	.04***	.00	.04	.05	.03***	.00	.02	.03	.03***	.00	.02	.04	.03***	.00	.02	.04	.03***	.00	.03	.04
	Student level	.05***	.00	.05	.05	.05***	.00	.05	.05	.05***	.00	.05	.05	.05***	.00	.05	.05	.05***	.00	.05	.05
	R^2						.407				.425				.449				.398		
Model fit	PPP		.000				.000				.000				.000				.000		
	DIC		16,165.14				53,489.14				136,830.35				157,573.66				148,523.78		
	p^D		3,771.61				2,945.65				3,014.55				3,005.43				2,990.06		
Number of	Secondary school classes		1,289				1,289				1,289				1,289				1,289		
	Primary school classes		741				741				741				741				741		
	Students		4,150				4,150				4,150				4,150				4,150		

Note. CI = confidence interval; LL = lower limit; UL = upper limit; SES = socioeconomic status; TSR: teacher-student relationship; PPP = posterior predictive p-value; DIC = deviance information criterion; p^D = estimated number of parameters.

[C] = variable is significant in the overall analyses, but not in Cohort 1 and Cohort 2 separately; [E] = variable is significant in the overall analyses and Cohort 1, but not in Cohort 2; [F] = variable is significant in the overall analyses and Cohort 2, but not in Cohort 1.

*p < .05, two-tailed; **p < .01, two-tailed; ***p < .001, two-tailed.

Discussion

When teachers formulate track recommendations, the main goal is to indicate what future educational level is most appropriate for a student. While primary school teachers mostly base their track recommendations on students' prior achievement (Driessen, 2005; Feron et al., 2013; Luyten & Bosker, 2004; Timmermans, Kuyper, et al., 2015; van Leest et al., 2020, 2024), it is not clear to which extent teachers take into account their perceptions of student attributes, such as perceived work habits or parental involvement. Using longitudinal data following students across the transition from primary to secondary education allowed us to examine the extent to which teacher perceptions of student attributes were predictive of student performance in secondary education. Thereby, the present study aimed to examine two research questions: (1) to what extent do teachers include perceived student attributes in addition to students' prior achievement, and (2) to what extent are these perceived student attributes predictive of students' secondary school performance? More insights on this will help to inform the debate on whether it is desirable for teachers to consider these perceived attributes in their recommendations.

Overall, the findings suggested that teachers included their perceptions of student attributes only to a limited extent in track recommendations, and teacher perceived student attributes were also only to a very limited extent predictive of students' secondary school performance. In addition, some teachers weighed some of these perceived attributes more heavily for some groups of students than for other groups of students, and teachers were also found to differ to a small extent from one another in the extent to which they weigh these perceived student attributes in their track recommendations. Together, these results suggest that primary school teachers are careful with taking perceived student attributes into account when formulating track recommendations.

In line with prior research, teacher perceptions of student attributes were found to have limited added value for the prediction of track recommendations, beyond students' prior achievement (Driessen, 2005; Driessen et al., 2008; Timmermans et al., 2016, 2019). Together, teachers' perceptions of student attributes explained only 0.70% of the total variance in track recommendations beyond prior achievement. Moreover, to our knowledge this study was the first to consider the predictive value of these attributes for students' performance in secondary school. The findings revealed that teacher perceptions of student attributes also had limited predictive value for secondary school performance, only explaining 0.29%, 0.20% and 1.82% of the additional variance beyond prior achievement for respectively reading comprehension, mathematics achievement, and level of education. More specifically, teacher perceptions of underachievement, work habits, classroom behaviour, popularity, dependency, and parental involvement all had small effects on track recommendations and these factors, except for perceived underachievement, also had a small effect on students' performance in secondary school. Track recommendations as well as students' performance in secondary education were primarily based on students' achievement in primary education. When students performed better in primary education, they

received higher track recommendations, performed better in secondary education, and attended a higher level of secondary education. The finding that teachers hardly consider these perceived attributes in their track recommendation may be considered a favourable finding given the very low predictive value of teacher perceptions of student attributes for students' performance in secondary education.

The findings furthermore suggest that it is possible that teacher perceptions of student attributes may – to a limited extent – influence students' secondary school performance indirectly via track recommendations. Without controlling for track recommendations, perceived work habits and closeness had a small effect on students' secondary school reading comprehension, whereas this relation was not present when track recommendations were included in the model (Model 3 versus Model 3a). Even more, teachers' perceptions of students' attributes as well as their track recommendations may indirectly affect secondary school performance in terms of self-fulfilling prophecy effects, i.e. it seems plausible that students will perform according to the level of education they are allocated to or the student attributes attributed to them, even if they have more potential (Brophy, 1983; de Boer et al., 2010; Jussim & Harber, 2005). For example, if a student receives a lower track recommendation because of his/her weak work habits in primary school, that student will possibly not be challenged to improve his/her work habits, while that student may be challenged to do so when given the opportunity to attend a higher secondary school track.

The track recommendation procedure in the Netherlands changed in 2014 due to a policy reform, and this may have implications for the role of perceived student attributes in the formulation of track recommendations. The data in the present study are from before the policy reform. First, after the policy reform, a changed time schedule resulted in the fact that teachers did not have the results of the standardised school leavers' test available to them when formulating a track recommendation. It may be that teachers weighed other information, such as perceived student attributes, in a different way or to a different extent than before the policy change. Second, since the policy change, teachers are allowed to reconsider their initial track recommendation (and adjust them upwards) if the results on the school leavers' test are higher than the initial track recommendation (Korpershoek et al., 2016; Oomens et al., 2019). Although in most situations where students qualified for upward corrections the initial track recommendation was not corrected (Ministerie van Onderwijs Cultuur en Wetenschap, 2019; Oomens et al., 2019; van Look et al., 2018), it is possible that when teachers consider upward corrections, teacher perceptions of student attributes may be considered more strongly. It would be interesting for future research to examine the role of perceived student attributes after this policy reform.

When looking at the direct effects of students' background characteristics on track recommendations on top of students' achievement, the effects were small. Together, the background characteristics only accounted for 0.30% of the variance in track recommendations, and effects were mostly found for gender and SES. This aligns with prior research which also indicated that there were small effects of students' gender and SES on track recommendations (Barg, 2013; Boone et al., 2018; Boone & Van Houtte, 2013; De Boer et

al., 2010; Driessen et al., 2008; Luyten & Bosker, 2004; Timmermans et al., 2013, 2016, 2018; Van Leest et al., 2020, in press), and prior research which indicated that students' migration background did not affect track recommendations (Boone & Van Houtte, 2013; de Boer et al., 2010; Driessen, 2005, 2011a, 2011b; Driessen et al., 2007; Ledoux et al., 2011; Roeleveld, Driessen, et al., 2011). Hence, teachers seem to consider students' background characteristics for track recommendations only to a limited extent, and even when they do, teachers consider students' migration background to a lesser extent than students' gender or SES.

Going beyond previous studies which only looked at direct effects of students' background characteristics on track recommendations, it was also examined whether background characteristics may impact teacher recommendations through different weighing of student attributes. The findings indicated that there were small, but significant differences between groups of students in the extent to which teachers considered different student characteristics. For example, lower-SES students with poorer work habits received lower track recommendations than higher-SES students with similar achievement and similar poor work habits. Only one (out of eight possible) interaction effects of SES and student attributes was significant, and only one interaction effect with migration background was significant. It is not clear if these significant interaction effects include perceived student attributes that are perceived to be stereotypical for the demographic (sub)group students belong to. Because it could be that teachers activate (and as a result rely on) these stereotypical ideas of how students of a certain group behave (i.e. based on social cognition literature (Glock, 2016a, 2016b; Klapproth et al., 2018; Strand, 2011, 2012), it would be interesting to examine in future research if this is also the case when formulating track recommendations. While, overall, these findings indicated that teachers weighed perceived student attributes mostly similar for students with different background characteristics, it is still possible that these attributes have a different predictive value for different groups with regard to their secondary school performance (for example, if lower-SES students benefit more from strong work habits). Future research could examine the extent to which the relationship between student attributes and student secondary school performance varies by students' background characteristics.

The present study also showed differences between teachers regarding the role of student attributes in their track recommendations. In line with previous research (Timmermans et al., 2016, 2019), small differences were found between primary school teachers in the extent to which perceived work habits and parental involvement were considered by the teacher when formulating track recommendations. Overall, students who were perceived to have stronger work habits or more involved parents received higher track recommendations. However, some teachers weighed these perceived student attributes negatively in their track recommendations. Teachers have the opportunity – at least to some extent – to determine themselves which factors or student attributes they include when formulating track recommendations, causing teachers to vary in the extent to which they include different perceived student attributes. This is in line with research suggesting that teachers seem to include different data sources and criteria when formulating track

recommendations (Vanlommel & Schildkamp, 2019). In addition, teachers may not only for a large part decide what factors or student attributes to include, they also assess these attributes themselves, with their own frame of reference, causing different teachers to perceive these attributes differently (Praetorius et al., 2017; Zhu & Urhahne, 2020). If teachers differ from one another, students with similar educational potential may end up in different educational tracks. Further research is needed to gain a better understanding of this matter.

Although the effects of perceived student attributes on track recommendations were, on average, small, the effects may be somewhat stronger among certain groups of students such as lower-SES students. Students for whom these student attributes already have a larger impact on their track recommendation due to their background characteristics, and who also attend a class where teachers include these attributes more strongly in their track recommendation, may experience a cumulative effect of these aspects. For example, higher-SES students and those whose parents are perceived to be highly involved by the teacher, already receive higher track recommendations than lower-SES students because of the main effects of SES and parental involvement. When placed with a teacher who considers parental involvement more strongly than other teachers, this effect could become even more substantial. These cumulative effects of students' background characteristics and perceived student attributes may possibly also be true for secondary school performance, for example if perceived student attributes would predict students' secondary school performance more strongly for lower-SES students than for higher-SES students. Keeping in mind that the Netherlands has a highly tracked system for secondary education with different educational opportunities for students (Glock et al., 2012; Klapproth et al., 2012; Korpershoek et al., 2016; Timmermans, Kuyper, et al., 2015), it would be interesting to investigate these cumulative effects of aspects on students' educational career.

It is important to note that the present study focused only on teacher perceptions of student attribute because these are the indicators that teachers can rely on when formulating track recommendations. Hence, the finding that teacher perceived student attributes are only to a very small extent predictive of track recommendations and the subsequent conclusion that considering them when formulating track recommendations is undesirable, only applies to these subjective teacher perceptions. This does not exclude the possibility that student attributes when assessed in a different, less subjective way, may in fact be predictive of students' future educational success (e.g. Hustinx et al., 2009; Van Rooijen et al., 2016) and may have added value for track recommendations. For future research it would be interesting to examine whether other, more objective, measures for student attributes may have the potential to enhance teachers' track recommendations. Also, in order to make a good comparison by what is happening in actual student behaviour and the teacher's perception, it would be interesting in future research to examine how track recommendations (and track placement) and teacher perceptions of student attributes may affect developments in different student attributes and may subsequently play a role in students' achievement. Student attributes are subject to change and development, as prior research indicated (Bittmann & Schindler, 2021; Hornstra et al., 2013; Müller & Hofmann,

2016; Praetorius et al., 2017; Van Houtte, 2017; van Rooijen et al., 2016; Zhu & Urhahne, 2020). For example, prior research suggested changes in students' investment in primary school over time, which were related to students' reading comprehension achievement (Hornstra et al., 2013), and students with initially high educational aspirations who started secondary education in a non-academic track were likely to lower their educational goals and eventually ended up without a higher-level educational degree (Bittmann & Schindler, 2021).

Limitations and Future Research

Several limitations of the present study are worth noting and need to be acknowledged. One limitation of the present research concerns the sample of the present research. Because not all students were included in the dataset in primary education, and especially in secondary education, a smaller sample size for RQ2 was used. Nevertheless, there were still 4150 students included in the sample of RQ2. Furthermore, some variables such as the SLT contained a higher percentage of missing values. This was due to the fact that not all schools administered the school leavers' test or they did not administer it to all students as – at that time – it was not mandatory (Driessen et al., 2009, 2012). Similar missing values for the school leavers' test were identified in other research (Luyten & Bosker, 2004; Timmermans, Kuyper, et al., 2015). Although missing values were handled with FIML, the results of this study should be interpreted with caution. In addition, the data distribution of RQ2 was not totally representative for the national data as small significant differences between the COOL$^{5\text{-}18}$ dataset and national data (Timmermans & Zijsling, 2014; Zijsling et al., 2017). The COOL$^{5\text{-}18}$ dataset contained more students in the higher tracks than nationally, and an unequal distribution of schools across the provinces, which might affect generalizability to other Dutch schools and regions, and even countries. There were also some small differences (Cohen's $d \leq .21$) between the two subsets of the data used in the present research. Nevertheless, the present study does provide insight into the predictiveness of teacher perceptions of student attributes on students' secondary school performance.

Another limitation of the present study is the specific context of this research. The present study was conducted in the Dutch educational context, which might affect generalizability to other educational systems as school track recommendations may not be used or may be formulated differently, especially with regard to the extent to which teachers may or may not include their perceptions of student attributes. In a strong exam-oriented educational system as the Dutch educational system is, it is not surprising that teachers are strongly guided by prior student achievement (Ağirdağ et al., 2021), and that track recommendations are mostly based on students' prior achievement (Driessen, 2005; Feron et al., 2013; Luyten & Bosker, 2004; Timmermans, Kuyper, et al., 2015; van Leest et al., 2020, 2024). It may be that in other educational systems teachers' perceptions of student attributes play a more important role.

Although choosing ninth grade as benchmark for students' secondary school performance was based on recent research indicating that ninth grade is indeed an important benchmark in students' educational career, there are also arguments for choosing another grade in secondary education as reference point (de Boer, 2009; de Jong & Steenbeeke, 2020). For example, research could include tenth grade, i.e. the grade in which students are divided based on their chosen study programme (van Rooijen et al., 2017), or the last grade of secondary education, so that it is clear what educational qualification students acquire (de Boer, 2009). It would be interesting to include multiple secondary school years or at least at a later stage in students' educational career to examine the extent to which students' track recommendation or their placement in the first year of secondary education was predictive of the educational qualification they acquired.

Furthermore, teachers had to fill out a student profile questionnaire, including the perceived student attributes examined in the present study for each student. These perceived student attributes were rated on a 5-point Likert type scale. When doing that for a class of approximately 25 students, it could be that teachers based their answers on making relative comparisons between students in their own class. As a result, different teachers may interpret these student attributes differently (e.g. Pendergast et al., 2018; Rubie-Davies, 2010) or use their class composition as a frame of reference (Boone et al., 2018), and rate them differently than other teachers on the Likert scale. This may have led to differences between teachers in interpretation, which may account for some of the between-teacher differences that were found. In the Netherlands, track recommendations are – most of the time – already formulated within a school team of multiple teachers and a special needs coordinator, but their different perceptions on student attributes are not visible in the track recommendation itself. Future research could include teacher perceptions of multiple teachers on these attributes to examine the extent to which these different perceptions are weighed in track recommendations. In addition, the findings of the present study are limited to the attributes that were included. It could also be that teachers also include perceptions of other student attributes that were not included in the present research in their track recommendations, such as effort (Vanlommel et al., 2020) or motivation (Paul W.J. Hustinx et al., 2009). Future research could include other teacher perceptions of student attributes.

Besides that, self-fulfilling prophecy effects may be at play. As mentioned before, students may perform according to the perceptions teachers have about them (Brophy, 1983; de Boer et al., 2010; Jussim & Harber, 2005). This may not only have an impact on their behaviour (and thus student attributes), but also on their achievement in primary education, and eventually on their track recommendation. In that case, the effect of student attributes is no longer visible in track recommendations, because they are already nested in students' achievement in primary education. In the present study, these different effects cannot be separated. Future research could include earlier teacher perceptions in primary education to try to get a grip on the teacher perceptions and their impact throughout primary education and beyond.

Conclusion

Altogether, the findings of the present study shed light on the question whether teachers consider their perceptions of student attributes in their track recommendations, and whether this is desirable or not. The findings indicated that teacher perceptions of student attributes (work habits, classroom behaviour, popularity, the dependency attribute of teacher-student relationship, and parental involvement) were hardly predictive of secondary school performance on top of the predictive value of achievement in primary education. Also, the differences – albeit very small – between students with different backgrounds and between teachers in the extent to which student attributes are considered in track recommendation suggest that it may be better not to include teacher perceptions of student attributes in track recommendations. Hence, the findings of the present study suggest that most teachers seem to base their track recommendations on prior achievement.

"Soms denk ik ook weleens dat ik niet te erg naar werkhouding moet kijken, omdat sommige kinderen het onderwijs op de basisschool minder leuk lijken te vinden en dat ze misschien op een vo-school bijvoorbeeld in een paardenklas tot hun recht komen. Dat zie ik dan weleens bij bepaalde opdrachten die ik geef. Ik zie dan dat ze daar heel goed in zijn. Hun houding kan dan veranderen als ze naar een vo-school gaan die past bij hun interesses, omdat ze op die school een andere manier onderwijs krijgen. Soms is dat wel een lastige overweging en dan moet je het echt inschatten."

"Ik zou moeten zien hoe ze aan haar opdrachten werkt. Hoe ze er mee bezig is. Je kunt aan haar werkhouding ook aflezen of iets veel te gemakkelijk is."

"Als je een leerling hebt die heel gemotiveerd is en je hebt ouders die het ondersteunen dan is het een win-win situatie."

"We hebben wel een groep leerlingen die hoog scoren en daar weinig voor hoeven te doen. De werkhouding is dan niet heel goed. Wij hebben wel het beeld op school dat kinderen die ervoor moeten werken verder komen in het vo dan kinderen die nu niet zoveel hoeven te doen (achterover hangen en even snel snel). Bij deze laatste groep is het cognitief allemaal goed, maar ze lijken toch sneller af te stromen omdat ze geen goede werkhouding hebben. Het dilemma is dan: cognitie zegt dit, werkhouding zegt dat. Wat gaan we daarmee doen? Welk advies geef je? Ik denk dat het uiteindelijke advies 50/50 is: de ene keer naar boven, de andere keer naar beneden."

6 General DISCUSSION

Introduction

In hierarchically tracked educational systems, students attend primary education together, but follow distinct secondary school tracks, based on ability level, which differ hierarchically in the educational certificates or qualifications students can acquire for tertiary education (Boone & Demanet, 2020; Contini & Scagni, 2011; Driessen et al., 2008). In most of these systems, the academically more challenging tracks are referred to as 'higher' levels of education.[16]

The Dutch educational system is such a hierarchically tracked system. In the Dutch system, students enter a secondary school track at a the age of twelve, which is relative early compared to other countries (Inspectie van het Onderwijs, 2007; OECD, 2020; Strello et al., 2021). Among politicians, teachers, school leaders, and scientists, the adverse effect for educational equality of this early tracking has been an ongoing debate for more than half a century, with the discussion about the Middenschool in the 1970s, the Basisvorming in the 1990s, and more recently the 10-14 schools in 2017 as milestones.

Not only early tracking, but also the best way to allocate students to secondary school tracks have been the topic of ongoing debates (Naayer et al., 2016; OECD, 2020; Onderwijsraad, 2021). The ongoing attention for this so-called track recommendation is justified, because in which secondary school students end up has major consequences. Mobility between tracks is rather limited and which track students complete largely determines their future (educational) careers (Glock et al., 2012; Korpershoek et al., 2016; Timmermans, Kuyper, et al., 2015). It is therefore important that students receive the 'right' track recommendation. The 'right' track recommendation is generally considered to be a recommendation that best reflects students' potential for educational performance in secondary education (Tieben & Wolbers, 2010).

The debate about teachers' track recommendation focusses on the type of information that teachers should take into account when formulating their track recommendations (Lek & van de Schoot, 2019; OECD, 2016b; Timmermans et al., 2018). How track recommendations are formulated varies from one educational system to another, but two main types of

16 Words like 'higher' and 'lower' in the context of education can have the connotation of more or less valuable. In this dissertation, the use of higher and lower are chosen for the sake of readability, but it is not intended to pass judgement on the value of different educational tracks.

track recommendations can be distinguished: (1) test-based recommendations, which are completely or primarily based on students' results on a standardised test, usually a school leavers' test, and (2) judgement-based recommendations, which are based on teachers' expectations or judgements of students' future performance. In the Netherlands, judgement-based track recommendations are used for allocating students to a secondary school track nowadays. Judgement-based track recommendations provide teachers with the opportunity to consider student achievement on standardised tests, as well as other factors that might be important for school success, such as work habits and classroom behaviour. However, there are concerns that teachers (unintentionally) also take students' background characteristics into account. Indeed, it has been repeatedly found that students' gender, SES, and migration background play a role in teachers' track recommendations (Boone & Van Houtte, 2013; Klapproth et al., 2013; Timmermans et al., 2018). This may affect the educational opportunities of certain groups of students in an undesired way.

This Dissertation

To better understand how and to what extent student characteristics play a role in teachers' track recommendations, more insight is needed into the interplay between student achievement, student attributes, such as work habits and classroom behaviour, and their background characteristics. Previous research mainly examined the predictive role of each of these factors on teachers' track recommendations, without considering their potential interplay. This dissertation contributes to the existing knowledge on track recommendations by also including the interplay between these factors. For example, in case of a lower-SES student, teachers may formulate lower track recommendations when they perceive this student to have poor work habits, whereas this issue may be overlooked or given less weight for higher-SES students. Therefore, the first aim of this dissertation was to determine what factors teachers consider when formulating track recommendations and how these factors together predict teachers' track recommendations. This dissertation discusses the extent to which teachers consider student achievement (Chapters 2 to 5), student attributes (Chapter 5), and student background characteristics (Chapters 2 to 5) when formulating track recommendations, either individually or in conjunction with each other. Additionally, teachers may differ from one another in how strongly they weigh each of these factors in their track recommendations.[17] Most prior research has disregarded these differences, which may lead to an underestimation of the role of these factors. The studies in this dissertation therefore also examine differences among teachers in their consideration of these factors when formulating track recommendations (Chapters 2, 3 and 5).

17 Teacher and school level cannot always be separated in this dissertation, especially in Chapters 2 to 4, due to unavailability of teacher data. However, most primary schools had only one Grade 6 class. Therefore, differentiating between teachers and schools would not have added value.

Since the goal of a track recommendation is to give an optimal reflection of students' potential educational performance in secondary education (Boone & Van Houtte, 2013; Glock et al., 2012; Klapproth et al., 2012; Korpershoek et al., 2016; Timmermans, Kuyper, et al., 2015), it is also important to know which factors are predictive of students' secondary school performance. Therefore, the second aim of this dissertation was to examine how predictive these factors actually are for students' performance in secondary school. In this dissertation, the extent to which (growth in) student achievement (Chapter 4), teacher perceived student attributes (Chapter 5), and student background characteristics (Chapters 4 and 5) predict students' secondary school performance is examined.

To answer these questions, this dissertation contains four empirical studies, using two large Dutch longitudinal datasets, which are presented in Chapters 2 to 5. In the following paragraphs, the findings of these studies were discussed in more detail following the two aims of this dissertation. After that, practical and scientific implications, implications for educational policy, limitations, and directions for future research are presented.

Findings and Discussion

Factors Considered by Teachers When Formulating Track Recommendations

Differences in educational opportunities between certain groups of students contradict the meritocratic ideal that educational opportunities should be based solely on (potential) ability rather than students' background characteristics, such as their gender, SES or migration background (Themelis, 2008). To prevent bias in track recommendations, many countries introduced (high-stakes) standardised tests in schools during the mid- and late 20th century (Au, 2013; Faasse et al., 1987; Luijkx & de Heus, 2008). These tests are intended to give objective information of students' abilities, thereby reducing the role of student background characteristics in track recommendations. In the Netherlands, Cito[18] was established for developing such standardised tests at the national level, such as the school leavers' test (1968) and the student monitoring system (1990s) (Cito, n.d.-a; Faasse et al., 1987; Luijkx & de Heus, 2008). The standardised school leavers' test was introduced in 1968 to provide teachers with an objective measure of students' abilities to inform their track recommendations. Tracks were defined in the Mammoth Act in 1968 which still forms the basis for the current Dutch educational system (Stiggins & Stiggins, 2002).

The question guiding the research described in Chapters 2 to 5 was what aspects of student achievement, perceived student attributes, and student background characteris-

18 Cito is the Central Institute for Test Development, established in 1967 (Cito, n.d.-a; Faasse et al., 1987; Luijkx & de Heus, 2008).

tics teachers consider when formulating track recommendations and how do these factors conjointly predict teachers' track recommendations? In addition, it was examined how predictive these factors are of students' performance in secondary education.

The Role of Student Achievement

In line with prior research (Driessen, 2005; Feron et al., 2013; Timmermans, Kuyper, et al., 2015), the findings of the research described in Chapters 2 to 5 suggest that teachers primarily base their track recommendations on the most recent test results available to them. Higher student achievement resulted in higher track recommendations.
Students' test results explained between 75% and 83% of the differences *(explained variance)* in track recommendations, depending on the policy regarding track recommendations in place.

The Policy Change of 2015

In the research described in Chapter 2, the situation before and after a policy reform in 2015 was compared. In the situation before 2015, teachers were able to base their recommendation on a school leavers' test which all students completed towards the end of the final year of primary education and on the test results from the student monitoring system. Teachers could divert from these test scores when formulating a track recommendation, but, in practice, they hardly seemed to use that option: 83% of the differences *(explained variance)* in track recommendations was explained by these test results. After the policy reform in 2015, teachers no longer had access to the results of the school leavers' test when formulating their track recommendations, but they still did have access to students' test results from the monitoring system. Our analyses showed that most recent test results of the monitoring system still explained 75% of the differences in track recommendations. We concluded that teachers tend to rely strongly on test results when formulating track recommendations, and when they have more test results available, they do use these test results, and - as a consequence - rely less on other factors. From a meritocratic perspective, it is considered desirable that teachers rely primarily on test results when formulating their track recommendations. Hence, from this perspective, the 2015 policy reform was not an improvement.

Although track recommendations are largely determined by students' most recent achievement, still 17 to 25% of the differences *(explained variance)* in track recommendations remain unexplained. This suggests that teachers also consider other more 'subjective' factors besides achievement when formulating track recommendations, such as perceived work habits or parental involvement, or even (unintentionally) student background, or they may consider other achievement-related aspects, such as achievement inconsistency and growth. Overall, the findings of this dissertation suggest that the role of these achievement-related and more 'subjective' factors in track recommendations is limited, emphasising the large impact of students' most recent achievement on track recommendations.

Other Achievement-Related Aspects

Prior research has mainly focused on the role of students' most recent achievement on track recommendations. In this dissertation, it was also examined how differences in achievement across subject domains (Chapter 3) and students' achievement growth over the years (Chapter 4) affect teachers' track recommendations.

Inconsistent Achievement

Although most students perform rather similar (i.e. consistently) across different subject domains, there are also students whose achievement differs across subject domains (e.g. high achievement in language and lower achievement in mathematics or vice versa; Luyten, 1998; Timmermans et al., 2015). The results presented in Chapter 3 showed that for about 20% of the students, the difference between the test results for reading comprehension and mathematics achievement was one standard deviation or higher (≥ 1 *SD*).

The overall effect of achievement inconsistency on track recommendations was very limited: it only explained an additional 0.2% of the differences *(explained variance)* in track recommendations beyond students' most recent achievement. This implies that when students show achievement inconsistency, teachers mostly formulate 'aggregated recommendations' (i.e. a track recommendation based on the mean achievement across domains) and tend to be slightly cautious (i.e. a 'careful' track recommendation based on students' weakest subject domain). Especially, when the achievement inconsistency was large (≥ 2 *SD*), teachers formulated such 'careful' track recommendations, particularly when students' reading comprehension achievement was lower than their mathematics achievement.

When students perform consistent across domains, it is easier for teachers to formulate a track recommendation compared with inconsistent student achievement (Böhmer et al., 2015). Because the Dutch educational system is a tracked system where students are allocated to a single level of secondary education and take all their courses at that level, it is understandable that teachers may be somewhat cautious in their track recommendations when students perform inconsistently. The higher the inconsistency, the more cautious they were. Teachers may be concerned that students will be unable to pass tests in their weaker subject domains if they are placed in a higher track. In all, the findings of Chapter 3 suggest that, at the time of the present research, teachers did not give students with inconsistent achievement the 'benefit of the doubt' based on their achievement in their strongest domain. Since 2021, which was after the data of Chapter 3 were gathered, the Dutch government has encouraged teachers to formulate high ('promising') track recommendations ('kansrijk adviseren') when they have doubts about the most appropriate secondary school track for a student. This policy is aimed at providing more opportunities for students from disadvantaged backgrounds (Dutch Education Council, 2019, 2021; Dutch Inspectorate of Education, 2020, 2021a). These higher track recommendations usually turn out to be advantageous for students, as research has indicated that most students who were given a 'benefit of the doubt' recommendation are able to succeed at this level of

education (Smeets & van Langen, 2022). Fortunately, preliminary findings from the Dutch Inspectorate of Education (Inspectie van het Onderwijs, 2021b) indicated that the number of 'promising' track recommendations has increased in recent years.

Furthermore, there were also small differences between primary school teachers in the extent to which they considered achievement in different subject domains in their track recommendations, with some teachers relying more strongly on certain subject domains than other teachers. These differences explained an additional 2.5% of differences *(explained variance)* in track recommendations on top of student achievement. As a result, two students with similar levels of achievement can receive different track recommendations depending on the teacher they have. Although local and national guidelines do exist (Inspectie van het Onderwijs, 2014; Oomens et al., 2019; Smeets et al., 2014), primary schools are not bound by them and are permitted to formulate their own guidelines, and, consequently, differences between teachers or schools can occur. In addition, these guidelines are not always applied in practice, especially in situations where students' achievement does not clearly indicate a specific secondary school track, teachers tend to deviate from the guidelines (Inspectie van het Onderwijs, 2014).

Achievement Growth

Students can also have different growth curves in their test results throughout primary education. These growth curves may be relevant to consider because students who started primary school with a disadvantage but managed to catch up by the end of primary education may have more potential than their test scores at the end of primary education show. Contrary to prior research (Caro, Lenkeit, et al., 2009; Klapproth & Fischer, 2019), the findings in Chapter 4 indicated that teachers only considered students' most recent achievement in their track recommendations, and not students' achievement growth over the years. This may be due to the large fluctuations in students' achievement during primary education and the considerable variation in growth curves between individual students. Hence, it may be challenging for teachers to compare growth curves between students, and, in turn, to take students' growth curves into account when indicating an appropriate level of secondary education. Moreover, scores from a single test can be directly translated to a particular school level, whereas it is more difficult to determine a school level based on multiple scores over the years.

Perceived Student Attributes

It is also possible for teachers to consider other, more 'subjective', information in their track recommendations, for example how they perceive individual students' work habits or classroom behaviour. The findings of Chapter 5 indicated a small overall effect of perceived student attributes. Students who were perceived by their teachers as having stronger work habits, displaying more negative (i.e. disruptive) classroom behaviour, underachieving, being relatively popular, having a relatively less dependent relationship with the teacher, or having relatively more involved parents received slightly higher track recommendations.

Together, these perceived student attributes explained less than 1% of the differences *(explained variance)* in track recommendations.

Hence, these findings suggest that primary school teachers do not substantially consider perceived student attributes when formulating track recommendations, which is in line with previous research (Driessen, 2005; Driessen et al., 2008; Timmermans et al., 2016, 2019). This does not mean that these factors do not matter for track recommendations. Instead, it seems plausible that these factors are associated with and predictive of achievement (Kpolovie et al., 2014; Malecki & Elliott, 2002; McKinney et al., 1975), and as such, do not have *unique* predictive value for students' track recommendations when students' achievement is also considered.

Furthermore, in line with previous research (Timmermans et al., 2016, 2019), the research in Chapter 5 indicated that there were also small differences between teachers in the extent to which they weighed different student attributes in their track recommendations. Some teachers weighed perceived work habits and parental involvement more heavily in their track recommendations than other teachers, explaining an additional 2.3% of the differences *(explained variance)* in track recommendations. For instance, when teachers perceived students' parents to be less involved in school (work), some teachers formulated a higher track recommendation for the student (e.g. because the student has come so far despite the support or involvement of those parents), whereas other teachers formulated lower track recommendations (e.g. because the parents cannot provide support in secondary education). According to the Dutch Inspectorate of Education (Inspectie van het Onderwijs, 2014) this could be explained by the fact that there are no clear guidelines or decision rules on the weighing of these types of factors in track recommendations.

Student Background Characteristics

When formulating their track recommendations, teachers may also (unintentionally) take students' background characteristics, such as gender, SES, or migration background, into account. The results of this dissertation indicated that teachers' track recommendations were slightly biased by students' background characteristics. Students' background characteristics accounted for less than 1% of the differences *(explained variance)* in track recommendations on top of students' recent test results. In addition, there were small differences between teachers in the extent to which they considered students' background characteristics in their track recommendations.

However, there were large indirect effects of students' background characteristics on track recommendations, meaning that differences in achievement between students with different backgrounds explained most of the differences in track recommendations, rather than a direct effect of these background characteristics. In addition, student background characteristics were not only related to achievement levels at the end of primary education (Chapters 2 to 5), but also to achievement inconsistency (Chapter 3) and to different achievement growth curves (Chapter 4). From a meritocratic perspective that students' background should not matter (Themelis, 2008), it can be considered desirable that

teachers' track recommendations are hardly biased by students' background characteristics. Nevertheless, the large gap in achievement and the subsequent differences in track recommendations can be considered very worrisome. In the next paragraphs, the findings for each background characteristic will be discussed in more detail.

Students' Gender

Aligning with previous research (Gentrup & Rjosk, 2018; Helbling et al., 2019; Spinath et al., 2014), the findings reported in this dissertation indicated that achievement differs between boys and girls: boys performed higher in mathematics, and showed more achievement growth in reading comprehension, and vice versa for girls (Chapters 3 and 4). The results of two of the three studies in this dissertation (Chapters 3 and 4, which were based on data gathered in a large Dutch city) indicated that gender did not affect track recommendations beyond students' recent test results. Hence, boys and girls with similar achievement, achievement inconsistency, or achievement growth were found to receive similar track recommendations. This is in line with studies (e.g. Boone & Van Houtte, 2013; Klapproth et al., 2013; Krolak-Schwerdt et al., 2017) that did not find an effect of gender on track recommendations. In Chapter 5, however, in which the findings were based on the national COOL5-18 dataset, small differences in track recommendations between boys and girls were found: while controlling for achievement, girls received higher track recommendations than boys. This finding is in line with research (e.g. Feron et al., 2016; Timmermans et al., 2016, 2018; Van Rooijen et al., 2016) which also found evidence of gender bias in track recommendations in favour of girls. These mixed findings of prior research and across the studies in this dissertation suggest that the effect of gender on track recommendations may not be robust as it tends to be small and varies across studies.

Furthermore, these discrepancies between the findings reported in Chapters 3 and 4 compared with those in Chapter 5 may be explained by differences in data sources used. Specifically, in the research described in Chapter 5, data from the national COOL5-18 dataset were used, whereas the data used in the research described in Chapters 3 and 4 were from one large Dutch city. Schnepf (2002) demonstrated significant differences in track recommendations between rural and urban areas in Germany for both genders. Girls in urban areas were more frequently recommended to the highest secondary school track. This may explain the gender differences we found between the findings of different chapters.

Additionally, Timmermans et al. (2018) suggest that the gender bias in track recommendations decreased the past decades. When we compare the results of the chapters with the time the data of these chapters were gathered, we see a similar pattern of decrease in gender bias. The data used in Chapter 5, in which we found a small gender bias, are from academic years 2007-2008 and 2010-2011, while the data used in Chapters 3 and 4 are from academic years 2014-2015 and 2015-2016, which did not show gender bias. According to Timmermans et al. (2018), the diminishing gender bias could be due to increased awareness of boys' underachievement in the classroom. This is commonly referred to as the 'boys problem' (Volman, 1999). The increased awareness of this issue has led to improved oppor-

tunities and greater attention to the talents and needs of boys (Heemskerk et al., 2012; Timmerman & van Essen, 2004), and may thereby potentially have reduced the gender bias in track recommendations. Another explanation for the differences between the findings across our studies could be that the large city our data originates from, applied local guidelines for track recommendations that differed from national guidelines (Inspectie van het Onderwijs, 2014; Oomens et al., 2019; Smeets et al., 2014).

Finally, in Chapter 5, it was found that there were no differences between boys and girls in the extent to which perceived student attributes were taken into account when formulating track recommendations. Hence, factors such as classroom behaviour were given equal weight in girls' and boys' track recommendations. Nevertheless, this does not exclude the possibility that teachers' perceptions of student attributes differ between boys and girls, and could even be biased (Beaman et al., 2006; Dee, 2005), and thereby indirectly affect their track recommendations.

Students' SES

In line with previous research (e.g. Batruch et al., 2023; Pit-ten Cate et al., 2016; Timmermans et al., 2016, 2023), track recommendations were found to be lower for lower-SES students compared to higher-SES students (Chapters 2 to 5). As explained before, these differences in track recommendations were mostly 'indirect'. That is, SES differences in students' achievement explained most of the differences track recommendations.

The additional bias of students' SES (direct effects) was small. It only accounted for 1% of the differences *(explained variance)* in track recommendations (Chapters 2 to 5). The research reported in Chapter 2 indicated that the role of SES was very small for test-based track recommendations (i.e. track recommendations given before the policy change in 2015 when teachers could rely on the results of the school leavers' test) as well as for judgement-based track recommendations (i.e. track recommendations given after the policy change when teachers did not have access to the school leavers' test and relied on students' last test scores of the student monitoring system). In both situations, SES explained 1% or less of the differences *(explained variance)* in track recommendations.

Overall, these findings suggest that the extent and type of student achievement data to which teachers have access is not very important. As long as they have access to standardised test results, they strongly rely on it. In previous research, the comparison between the use of standardised tests and/or the school leavers' test within a test-based system was also made. Luyten and Bosker (2004) showed that in schools that did not administer a school leavers' test, but where teachers had access to standardised test results, the effects of students' background characteristics on (test-based) track recommendations were significantly stronger and the effects of achievement on (test-based) track recommendations somewhat weaker. Additionally, research of Borghans and Diris (2021) has shown that after the policy reform, the effect of students' SES (measured by parental education) on track recommendations became larger. The findings reported in Chapter 2 also showed that teachers' judgement-based track recommendations rely somewhat less on achievement

compared to test-based track recommendations. However, our results do not confirm the strengthened impact of SES.

There are also other ways in which background variables can affect track recommendations. Besides the previously reported direct and indirect effects, it is also possible that teachers may weigh achievement or more 'subjective' student characteristics differently for different groups of students. Indeed, achievement inconsistency seemed to have a slightly stronger impact for higher-SES students than for lower-SES students (Chapter 3). That is, teachers seemed to formulate somewhat more careful (i.e. lower) track recommendations for higher-SES students with achievement inconsistency than for lower-SES students with achievement inconsistency. They also weighed work habits somewhat more strongly for lower-SES students than higher-SES with similar achievement (Chapter 5). This suggests that perceived poor work habits had a stronger negative impact on track recommendations of lower-SES than of higher-SES students. Although the effect was small, this can be considered worrisome, especially because previous research (Bakker et al., 2007; Riley & Ungerleider, 2012; Sui-Chu & Willms, 1996) indicated that teacher perceptions of such student attributes can be biased by students' SES. The extent to which teachers considered other student attributes (Chapter 5) or achievement growth (Chapter 4) in their track recommendations did not depend on students' SES. In all, the findings concerning students' SES suggest that achievement differences are by far the most important reason for SES differences in track recommendations.

Students' Migration Background

Previous research has shown mixed findings regarding bias towards migration background in track recommendations. Some studies found no evidence for bias (e.g. Boone et al., 2018; Feron et al., 2016; Timmermans et al., 2016), whereas other studies found evidence for 'overrecommendations', indicating that students with a migration background received higher track recommendations than students without a migration background in case of similar achievement (e.g. Caro et al., 2009; Dumont et al., 2019; Timmermans et al., 2018, 2019), while there are also studies that found that students with a migration background received lower track recommendations than students without a migration background in case of similar achievement (e.g. Krolak-Schwerdt et al., 2017; Lüdemann & Schwerdt, 2013; Van Rooijen et al., 2016).

In Chapter 5, we included students' migration background as a predictor of track recommendations. Aligning with the findings of some of the previous studies (e.g. Boone et al., 2018; Feron et al., 2016; Timmermans et al., 2016), we did not find a bias based on students' migration background in teachers' track recommendations. However, we did find a difference between the two cohorts included in the study in Chapter 5. The data from the academic year 2007-2008 indicated 'overrecommendation' (i.e. higher track recommendations than their achievement levels indicated) for students with a migration background, but this effect was not present in the data from the academic year 2010-2011. This aligns with the findings by Timmermans et al. (2018) who showed a reduction in positive track recommendation bias for students with a migration background between 1995 and 2014.

Timmermans et al. (2018) provide several explanations for the decrease in track recommendation bias towards students with a migration background. First, the initially higher track recommendations for students with a migration background could be due to positive discrimination. Until 2007, Dutch educational policies encouraged teachers to provide more educational opportunities for these students, for example, through additional funding. Since these students are no longer a specific target group for providing equal educational opportunities, the effects of migration background on track recommendations may have diminished. Second, teachers may have had positive perceptions towards these students due to the fact that their parents held relatively high ambitions for them (De Boer & van der Werf, 2015), but the growing cultural intolerance towards Muslims and other migration background groups may have altered these positive perceptions.

Another reason for the non-significant finding of migration background in the second cohort could be that students' SES and migration background often overlap: students with migration backgrounds often have a lower SES (Luyten & Bosker, 2004; Van der Veen, 2003). Hence, when including both SES and migration background, there may not be a unique effect of students' migration background.

Finally, in the study described in Chapter 5, a small difference was found between students with and without a migration background in the extent to which perceived interpersonal closeness in the teacher-student relationship was taken into account in track recommendations. A close relationship with the teacher led to higher track recommendations for students with a migration background, but not for students without a migration background. However, it is important to note that these effects are only small.

Overall, the findings suggest that achievement differences are by far the most important reason for differences in track recommendation between students with and without a migration background.

Predictors of Secondary School Performance

Teachers' track recommendations are aimed at allocating students to a secondary school track in which they have the best chance to realise their potential. Therefore, it is important to understand which factors included in teachers' track recommendations are actually predictive of students' secondary school performance. Overall, the findings described in Chapters 4 and 5 suggest that the factors that teachers considered when formulating their track recommendations corresponded to a large extent with factors that were predictive of students' secondary school performance.

Students' Most Recent Achievement

In line with previous research (de Boer et al., 2010; Poncelet & Metis Associates, 2004; van Rooijen et al., 2016), teachers primarily considered students' most recent test results when formulating track recommendations. In our studies, we found that achievement at the end of primary education was also a strong predictor of students' secondary school performance, explaining 40 to 55% of the differences in secondary school performance (Chapters 4 and 5). Higher achievement in primary school was associated with higher achievement in

secondary school and with a higher level of secondary education, even after controlling for differences in track recommendations or secondary school track placement.

Achievement Growth

It was also expected that achievement growth would be predictive of secondary school success, as achievement growth can be indicative of a student's potential for further achievement growth. That is, a student who learns and acquires new knowledge or skills quickly may have a higher growth rate in both primary and secondary education compared to a student who acquires new skills more slowly. However, the findings reported in Chapter 4 showed that achievement growth in primary education did not predict students' secondary school performance after taking students' most recent achievement in primary education into account.

(Early) tracking may be the reason why students' relatively high level of achievement growth was not continued in secondary education. That is, students' growth potential may be limited by the track they are allocated to, as track placement – not only in the first year of secondary education but also later in secondary school – is based on achievement thresholds (Borghans et al., 2019; Dockx et al., 2019; Hanushek & Wössmann, 2006). This restricts what students can achieve within a particular secondary school track. In other words, it seems plausible that students may perform according to the level of education they are assigned to, even if they have more potential. To illustrate, a student who has shown significant achievement growth in primary education but receives a cautious recommendation based on their achievement at the end of primary education, may be performing adequately for his or her current educational level. However, if this student had been placed at a higher track, he or she might have been challenged to show further achievement growth. Hence, in an untracked or less rigidly tracked educational system, the predictive value of students' achievement growth in primary education for their educational success in secondary education might be higher.

Perceived Student Attributes and Background Characteristics

We also examined the role of student attributes, as perceived by the primary school teacher, for performance in secondary school. The findings of Chapter 5 indicated that more 'subjective' factors, such as perceived work habits, classroom behaviour and parental involvement, only had a very little predictive value for students' performance in secondary school after accounting for student achievement in primary education. Together, such perceived student attributes explained less than 2% of the differences in secondary school performance.

Furthermore, the findings of Chapters 4 and 5 showed that students' secondary school success was impacted by their background, after accounting for prior achievement. For instance, girls and higher-SES students tended to perform better than boys and lower-SES students. Again, these effects explained only 1 to 3% of the differences *(explained variance)* in secondary school performance.

Implications for Educational Practice and Policy

The findings of this dissertation have potential implications for educational policy and practice. The implications that are discussed first, focus on the way in which track recommendations are formulated. In addition, the findings of this dissertation suggest that educational inequality may be more strongly rooted in how education is organised before and after the transition to secondary education (i.e. before refers to preschool and primary education, and after refers to secondary education). Therefore, the focus is also on (in)equality before and after students receive their track recommendations.

At the Transition From Primary to Secondary Education

Having a High-Stakes Test for Determination

Is a high-stakes test at the end of primary education desirable? There is no straightforward answer to this question. There are several reasons why the use of a standardised test is desirable, but also why it is not. Firstly, when teachers had access to more standardised test results which they could consider in their track recommendations, i.e. the school leavers' test on top of the test results from the student monitoring system, they based their track recommendations to a larger extent on students' achievement. That is, the findings of Chapter 2 indicated that the explained variance in track recommendations was 8% higher when results from the school leavers' test were available. This could suggest that there is less room for bias based on students' background characteristics. However, the findings in Chapter 2 did not support this idea: in both situations (with and without the school leavers' test), the direct effect of SES on track recommendations explained only 1% of the differences *(explained variance)* in track recommendations. Nevertheless, due to the larger explained variance in track recommendations when the school leavers' test is available, SES might play a larger role through the indirect effect of achievement (i.e. lower achievement for lower-SES students).

Secondly, standardised tests set the standard at the same level for all students. Therefore, it is possible to make comparisons in test results between students, both at school and national level. Furthermore, a school leavers' test usually already indicates a 'right' secondary school level based on students' achievement on this test. This can help to reduce the differences between primary school teachers' track recommendations, as they can compare their own students' achievements with those of others. It is especially the case when only one test is used such as a school leavers' test. When using the student monitoring system, which consists of multiple tests across time and domains, it is more difficult to compare the individual achievement (trajectory) of students during primary education with that of other students. Moreover, there is more room for differences in how teachers weigh these different tests. For example, the findings of Chapter 3 showed that

some teachers weighed achievement in some subject domains more heavily in their track recommendations than others.

Thirdly, administering a high-stakes test(s) could induce performance pressure, causing students to experience stress when they have to take such a high-stakes test(s) (Dopmeijer et al., 2023; Kleinjan et al., 2020; Mijs, 2016). This is one of the main reasons mentioned as a disadvantage of having a school leavers' test. However, not administering a school leavers' test does not solve the problem of performance pressure. If a school leavers' test is not used for formulating track recommendations, the significance of the school leavers' test shifts to the tests of the student monitoring system, as shown in Chapter 2. Consequently, the performance pressure students experience may also shift to the tests from the monitoring system, which may lead to performance pressure being spread over a longer period of time.

Lastly, a similar scenario could arise with regard to shadow education: not administering a school leavers' test does not solve the problem of shadow education, as the use of shadow education would shift to the tests of the monitoring system. An even greater increase in tutoring and training institutes that prepare students for these types of tests (Elffers, 2017; Elffers & Jansen, 2019) may be the result, since students need to be prepared for more tests over a longer period of time. In addition, the use of shadow education has major negative implications for equal educational opportunities for all students, because lower-SES students, who often come from less affluent families, may not have access to such tutoring and test training (de Geus & Bisschop, 2018). As a result, lower-SES students may be less prepared for and may perform not as good on the test(s) as higher-SES students who had extra test training. It is important to consider the financial barriers that may prevent these students from accessing additional resources.

Timing of the High-Stakes Test

In addition to the desirability of administering a high-stakes test at the end of primary education, the timing of such a test is also important. The school leavers' test and how it is used in the Dutch educational system has often changed in recent years. In the 2023-2024 academic year, the most recent policy reform regarding the use of the test in the transition from primary to secondary education was implemented. The school leavers' test was replaced by a so-called 'transition' test ('doorstroomtoets' in Dutch). This transition test can be considered a rebranding of the school leavers' test to suggest that it is not a high-stakes test after which development stops, but that development continues after the transition to secondary education (Ministerie van Onderwijs Cultuur en Wetenschap, 2023).

However, a more significant change is that when results on this test indicate a higher level of secondary education compared to the track recommendation, the teacher's track recommendation is adjusted automatically, except in the case of compelling reasons not to do so. Previously, adjustment of the track recommendation was optional. This often meant that many students who were eligible for an upward adjustment of their track recommendation did not receive such a correction (Ministerie van Onderwijs Cultuur en Wetenschap, 2019, 2023; Oomens et al., 2019; van Look et al., 2018). Additionally, adjustments were

more often made for higher-SES students (Swart et al., 2019). Research suggests that parents of higher-SES students may be more likely to question the track recommendations and exert more pressure on teachers than parents of lower-SES students (Batruch et al., 2023; EenVandaag & CNV Onderwijs, 2018; Inspectie van het Onderwijs, 2014; van Grinsven & van Rossum, 2022). Therefore, these new adjustment rules may help lower-SES students who are eligible for upward corrections to also receive higher track recommendations, and consequently, it may help to reduce the direct impact of background characteristics on track recommendations.

Positioning the test in time *after* students receive the track recommendation, the focus shifts from a single test to a more comprehensive view of the student's progress over a longer period of time. Moreover, there is always the option of being able to 'repair' a biased teacher recommendation with the results of the transition test. Summarising, the new system seems promising, but further research is necessary to explore if this new system works as intended.

Promising Track Recommendations

The current educational policies aimed at 'promising' track recommendations (known as 'kansrijk adviseren') for students for whom there is doubt about the right track may also reduce differences in track recommendations between teachers and limit bias by background characteristics. That is, differences between teachers may come about in cases where the track recommendation is not straightforward. Similarly, bias may especially occur only in these cases where there may be doubt between multiple tracks. As such, since the academic year 2020-2021, Dutch policymaker's general advice to teachers is to formulate promising track recommendations that may help to reduce these differences and bias (Inspectie van het Onderwijs, 2019, 2021a; Onderwijsraad, 2019, 2021). Formulating promising track recommendations has been found to be beneficial for students, since higher track recommendations and, subsequent track placements, lead to higher secondary school performance (Hebbink et al., 2022; Lenhard & Schröppel, 2014; Tolsma & Wolbers, 2010). An explanation may be that students 'adapt' their effort to meet the standards of the track they have been allocated to (de Boer, 2009; Hustinx, 2002; Maaz et al., 2008). Research has shown that, at least for the first year after this encouragement policy, teachers increasingly formulated more promising track recommendations (Inspectie van het Onderwijs, 2021b). Based on these findings, this policy for promising track recommendations can be described as effective in encouraging teachers to formulate high track recommendations, thereby improving the educational chances of students. It is recommended that policymakers continue to encourage towards promising track recommendations and that teachers continue to formulate them.

Supporting Teachers With Clearer Guidelines

To limit bias and differences between teachers in track recommendations, clear guidelines for schools on formulating track recommendations are necessary. Although national and

local guidelines and procedures do exist (Inspectie van het Onderwijs, 2014; Oomens et al., 2019; Smeets et al., 2014), primary schools are not bound by them and are permitted to create their own guidelines for formulating track recommendations. In pursuing equal educational opportunities for all students, it is important that guidelines do not differ (for the most part) between schools and that they are implemented similarly in all schools to limit bias and reduce differences in track recommendations between teachers. It is also important that such guidelines also focus on, for instance, differences in achievement between subject domains and clear (steep) growth curves. As part of implementing such guidelines teachers would benefit from additional support on how to interpret information on student achievement that is not straightforward.

In the Netherlands, primary school teachers have access to results of standardised tests from a monitoring system throughout primary education, visualised in an online teacher dashboard. Such dashboards contain mostly information on student achievement but sometimes also information on social and emotional student behaviour (Ministerie van Onderwijs, Cultuur en Wetenschap, n.d.). Next to information on the level of specific students, sometimes national data are available to facilitate comparison. However, it might be difficult for teachers to translate all this student information to a best fitting track recommendation, especially when not all student information points to a specific secondary school track. More advanced and easy-to-use dashboards could potentially help teachers, when making educational decisions such as formulating a track recommendation (Holstein et al., 2019; Molenaar & Knoop-van Campen, 2017).

Formulating Track Recommendations Together

To reduce bias in track recommendation, a practical solution is letting educational professionals formulate track recommendations together. Research showed that track recommendations in the majority of primary schools is formulated by several educational professionals together (Inspectie van het Onderwijs, 2014; Smeets et al., 2014). The Grade 6 teacher or teachers are always involved, sometimes assisted by the primary school principal, the internal supervisor and/or teachers that have taught the student in previous years. The number of educational professionals that were involved in formulating track recommendations differed between schools but typically at least two educational professionals were involved (Inspectie van het Onderwijs, 2014). Although it is not exactly known to what extent the size and composition of this group formulating track recommendations decreases the bias in track recommendations, research on decision making does suggest that expert groups providing judgements together may benefit from an increase of accuracy in their decision making (Kerr & Tindale, 2004, 2011).

Moreover, collaboration between schools or school boards may help to reduce differences between schools, since they set policies and procedures for track recommendations at the moment (Timmermans et al., 2023). Schools and school boards could discuss and compare their track recommendations, for example through a peer review format

where specific cases of students can be discussed, for instance cases about which teachers have more doubts regarding their track recommendation. This allows schools to compare whether they would provide similar track recommendations for particular students for whom teachers have more doubts regarding their track recommendation.

Before the Transition From Primary to Secondary Education

The findings presented in this dissertation indicate that there are substantial and persistent gaps in achievement between students from different backgrounds (Chapters 2 to 5). Students with lower-SES backgrounds or a migration background, on average, have lower achievement levels than other students. This is reflected in the track recommendations they receive. The track recommendations themselves hardly reinforce these differences. Hence, to reduce differences in track recommendations between groups of students with different background characteristics, it seems more promising to focus on reducing achievement differences in primary school rather than solely focusing on the track recommendations themselves.

Differences in achievement between students with different backgrounds are already visible at the start of primary education (Magnuson & Duncan, 2016; Roeleveld, Mulder, et al., 2011; van Look et al., 2018). More specifically, research has demonstrated that there are already variations in capabilities between Dutch children with different background characteristics when these children are as young as two years old (Mulder et al., 2015). Previous research (Luyten et al., 2009; van Huizen, 2018), as well as the results from this dissertation (Chapter 4), suggest that these gaps are not reduced during primary education. Instead, they are maintained or even increase over time.

Bernstein wrote in 1970 that "education cannot compensate for society", but that should not mean that schools should not try to reduce the gaps between students from different backgrounds. Prior research suggests that focusing on preschool children could for instance be one way to reduce existing achievement gaps. Options are prolonging the time that children spend in preschool (Sierens et al., 2020), ensuring better-quality preschool facilities (Leseman et al., 2017; van Huizen & Plantenga, 2018; Veen & Leseman, 2015), or ensuring a smooth transition from preschool to primary school (OECD, 2017, 2023).

To further contribute to reducing achievement gaps during primary school, even more additional resources need to be dedicated to students with disadvantaged backgrounds. This can include increased funding, time, attention, implementing effective interventions, such as tutoring, cooperative learning, monitoring student progress (Dietrichson et al., 2017), and facilitating or encouraging parental involvement in home-based educational activities (Watkins & Howard, 2015). For example, Dietrichson et al. (2017) suggested that (feedback to teachers on) student progress monitoring can have positive effects for lower-SES students, mainly because it gives teachers more knowledge about students' progress and allows teachers to adapt educational materials and teaching instruction targeted at a particular group of students accordingly. Moreover, for lower-SES students

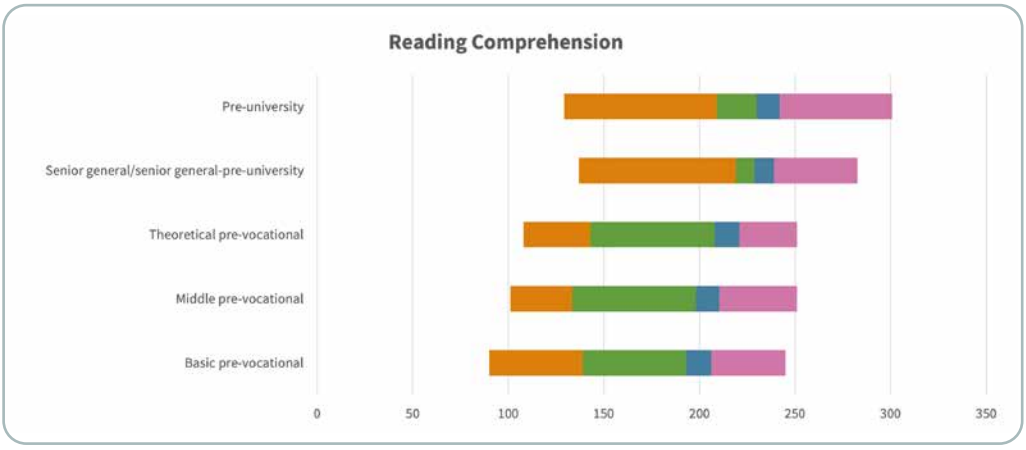
and students with a migration background especially interventions that prioritise reading comprehension and language skills may be helpful, as these students typically perform lower in these areas (Chapter 3).

After the Transition From Primary to Secondary Education

In the debates about unequal educational opportunities, teachers' track recommendations have often been the focus. This allocation process, based on the teacher' track recommendation, is under such scrutiny because it is very decisive for students' future educational careers. However, this research shows that teachers' track recommendations in itself make at most a modest contribution to unequal educational opportunities. Rather, the way Dutch secondary education is organised may be a greater factor which limits the educational opportunities of students from groups who are at risk of educational disadvantage. Students in the Netherlands are tracked at a very early age and there is limited mobility in Dutch secondary education. Mobility between tracks is rather uncommon, especially upwards (Driessen et al., 2005; Timmermans et al., 2013; van Rooijen et al., 2017). This is also referred to as the strong 'path dependency' of the Dutch educational system. As a result, the achievement gaps between students from different backgrounds in primary education are perpetuated in secondary education (Inspectie van het Onderwijs, 2021a; Scheerens et al., 2019). According to the findings of this dissertation, this especially affects lower-performing students, for example lower-SES students, negatively, since they receive lower track recommendations (Chapters 2 to 5).

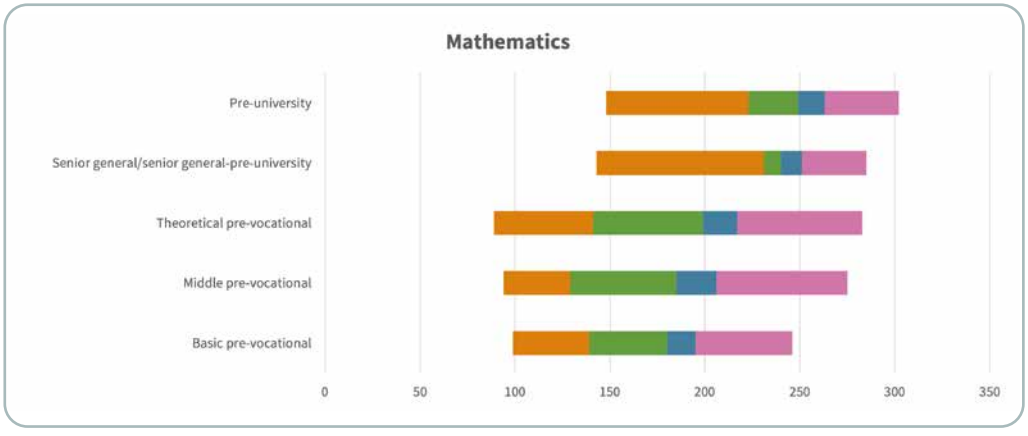
If it is not the track recommendation, the logical question to ask is whether changes should perhaps be made to the secondary school system to make it less rigid and to reduce the importance of the track recommendation. Moreover, there is a significant overlap in achievement between students in different tracks of secondary education (Hallinan, 1994; Inspectie van het Onderwijs, 2017a; OECD, 2016a). Similarly, data from our Utrecht dataset also show this overlap. The top 50% of pre-vocational middle track *('vmbo-k')* students perform similarly to the average pre-vocational theoretical track *('vmbo-g/t')* student, and the top 50% students in the senior general education track *('havo')* perform similarly to the average pre-university education track *('vwo')* student. In Figure 1 and 2, the overlapping achievement of respectively reading comprehension and mathematics of the Utrecht data used in this dissertation can be seen. In these figures, there is also substantial overlap between students in the lowest (i.e. pre-vocational basic track) and highest (i.e. pre-university track) track. This overlap in achievement suggests that a less rigid system may be desirable. Several implications to make the Dutch educational system less rigid will be addressed briefly.

Figure 1



Overlapping Achievement in Reading Comprehension Between Secondary School Tracks

Figure 2



Overlapping Achievement in Mathematics Between Secondary School Tracks

Postponing the Selection Moment

First, postponing the selection moment for students transitioning from primary to secondary education would be beneficial for students with lower achievement levels or lower-SES backgrounds, since early tracking especially affects them (Bauer & Riphahn, 2006; Borghans et al., 2020; Brunello & Checchi, 2007; Hopwood et al., 2016; Pekkala Kerr et al., 2013). International comparison studies suggest that early tracking increases educational inequality (Hanushek et al., 2006; Terrin & Triventi, 2023; van de Werfhorst, 2018). Hence, postponing the selection moment may help to reduce the divergent achievement gap, as students have more time to develop their full potential, prior to being separated into different tracks (van de Werfhorst & Heath, 2019). In 2021, this was also recommended by the Dutch Education Council (2021). They recommended that students should be assigned to a track only after three years of secondary education. However, postponing the track placement decision must be accompanied by additional interventions aimed at reducing the achievement gap between lower- and higher-SES students, since this dissertation (especially Chapter 4) indicated that this SES-based achievement gap widens with age. Hence, if this gap continues to widen, postponing the track placement decision could cause even greater differences in track allocation, unless effective interventions are implemented that reduce socioeconomic achievement gaps.

Introducing Heterogeneous Lower Secondary Education

Second, postponing the selection moment goes hand in hand with introducing comprehensive heterogeneous lower secondary education, where students are not allocated to specific tracks until later in their school career (commonly known as '(brede) brugklassen' in Dutch). The positive effect of heterogeneous classes is partly due the influence of students on each other. In a heterogeneous class, lower-performing students tend to perform better (Grift et al., 2010; van Elk et al., 2009; Zimmer, 2003). It needs to be mentioned here that an increasing number of Dutch secondary schools already have a bridging period aiming to mitigate the effects of early tracking. In this bridging period, students with adjacent track recommendations are not placed in specific tracks right-away but are kept in the same group for this period. However, this bridging period is sometimes limited to one year, or only includes some classes or tracks within a school. Therefore, lower secondary education can be optimalised in how heterogeneous education is organised. It could be beneficial for students who perform lower such as lower-SES students, as shown in Chapters 2 to 5.

Making Switching Upwards Easier

Furthermore, the rigidity of the system and thereby the importance of the track allocation may be reduced when switching upwards is made easier. This would be especially beneficial for students who received a track recommendation below their potential, which may be the case for students with a steep achievement growth curve in primary education (Chapter 4). If students are able to show that they are able to attend a higher track, it should be possible for them to switch upwards throughout secondary school. Switching is often difficult due to

institutional reasons, for example when schools do not offer alternative tracks or because the curricula of school subjects are not aligned between different tracks. This is becoming more prevalent, as an increasing number of schools offer only one particular level of education (i.e. 'categorale scholen' in Dutch; Inspectie van het Onderwijs, 2016). Switching tracks would then also involve a transfer to another school.

In addition, the possibilities for upwards transitions are often limited by school regulations. Most schools have formulated specific rules about when students are eligible to move up. These rules differ between schools. Furthermore, these rules can also be quite difficult for students to meet. For example, a student in a mixed pre-vocational theoretical and senior general track ('*vmbo-t/havo*') must achieve an average grades of 7.5 or 8.0 with no failures and meet certain minimum grades in core subjects, such as mathematics, Dutch and English before being eligible to transfer to a homogeneous senior general track. Students in the homogeneous senior general track, on the other hand, only need to meet the minimum requirements to be promoted to the next grade, and usually they are allowed to fail some subjects (de Winter-Koçak & Reches, 2022). It would be helpful if these types of regulations were streamlined within and between schools to promote opportunities and reduce differences between secondary schools.

Providing the Ability to Attend Courses at Different Levels

Finally, the findings of this dissertation demonstrate that a significant proportion of students show substantial achievement differences across different subject domains (Chapter 3). Yet, students follow all courses at the same educational level, which could imply that they may struggle to pass some of their courses and follow other courses at a level below their abilities. Ability grouping by subject would enable students to attend classes at a level that better aligns with their abilities. Ability grouping by school subject is also practised in several other countries, for example, the USA (Irizarry, 2021; Yaluma & Tyner, 2021).

Limitations and Suggestions for Future Research

Limitations of This Dissertation

There are several limitations of this dissertation to take into account.

Generalisability and Representativeness of the Sample

The first limitation concerns the generalisability and representativeness of the sample. The context of this dissertation was the Dutch educational system. Therefore, results of the studies of this dissertation may not be fully generalisable to other countries where track recommendations may be formulated differently and at a different moment in students'

educational career. Additionally, Chapters 2 to 4 were based on data from one large Dutch city and the findings of these studies may not be fully generalisable to other regions in the Netherlands. That is, small variations in findings were noted between the datasets with local and national data used in the present study, for example outcomes regarding gender. Hence, findings from studies using the local dataset may be specific to this city where schools had access to an allocation guide, known as the 'plaatsingswijzer' (POVO, 2013, 2015). In addition, the data distribution of the national dataset used in Chapter 5 was not entirely representative of the Netherlands due to slight but significant differences between the COOL5-18 dataset and other national data (Timmermans & Zijsling, 2014; Zijsling et al., 2017). The COOL5-18 dataset for secondary education included a higher number of students in upper tracks compared to the national average. Additionally, there was an uneven distribution of schools among the provinces, potentially impacting the generalisability of the findings to other Dutch schools, regions, or even countries. Nonetheless, Chapter 5 offers valuable insights into the predictive value of teachers' perceptions of student attributes for teachers' track recommendations and students' performance in secondary school.

Operationalisation of SES

Second, in Chapters 2 to 4, SES was operationalised based on a six-digit postal code. Although these six-digit postal codes are only shared by a very small number of households (Deckers et al., 2016; Guhn et al., 2010; van Hattem et al., 2009) and therefore provide a relatively accurate indication of the socioeconomic status of households, they do not distinguish between each individual household. Additionally, the six-digit classification included some missing values. For the missing values, we utilised the five-digit postal codes and to a lesser extent the four-digit codes, which offer less precision. Additionally, assessing SES through postal codes rather than individual measures could have created an overestimation of the variance at school level. As a result, the outcomes could be impacted: the school-level effect of SES on track recommendations that was found in Chapter 2 may be slightly overestimated and partly be located at the individual level. Thus, it is advisable to exercise caution when interpreting these outcomes.

Teachers' Initial Track Recommendations

Third, we focused on the teacher's initial track recommendations formulated in March rather than the final track recommendation or the track placement in secondary education. These initial track recommendations are indicative of the process that teachers undergo in formulating their track recommendations. Focusing on these initial track recommendations thereby enabled us to study how teachers weigh different factors and to compare test-based and judgement-based track recommendations in Chapter 2. We did not focus on adjusted track recommendations, as this was beyond the scope of the research we report about in the present dissertation. When considering the final track recommendations, other factors may be at play such as parental pressure (EenVandaag & CNV Onderwijs, 2018; Inspectie van het Onderwijs, 2014; van Grinsven & van Rossum, 2022). In addition,

the relation between the initial track recommendation and track placement is even more complicated because secondary schools are not required to place students at a level that exactly matches their track recommendation (Inspectie van het Onderwijs, 2023). Nevertheless, prior research (Hebbink et al., 2022) showed that the differences between the initial and final track recommendations are very small, and both track recommendations seemed to be equally predictive of students' educational level in Year 3 of secondary education.

Secondary School Performance

Furthermore, secondary school performance was assessed in the initial year(s) of secondary school, as data on performance in later years was not available. We could only include data from Year 2 of secondary education in Chapter 4, and data from Year 3 in Chapter 5. Therefore, we could not take into account students' academic progress in subsequent years in secondary school. This may have limited the variance in the outcome measure secondary school performance (i.e. reading comprehension and mathematics achievement, and level of education) as most of the switching between different levels of secondary education or the repetition of grades in secondary education occurs in later years (van Vuuren & van der Wiel, 2015; Veenstra, 1999). It would be informative for future research to incorporate multiple years of secondary school performance to contribute to more in-depth knowledge about the predictive value of different factors that are considered in track recommendations for students' future school success.

Measurement Errors in Track Recommendations

Findings of this dissertation also suggest that there are differences in track recommendations between students from different backgrounds (Chapters 2 to 5). This is mostly due to differences in achievement between these students, but we also found a small direct effect of students' background on track recommendation. Besides the fact that teachers could be biased or include other factors in their track recommendations (what we did not find in our studies), research of Van Huizen (2021) suggests that these differences in track recommendations can to some extent also be due to a measurement error in achievement tests. For instance, two students with different backgrounds performed similarly on the school leavers' test but received different track recommendations. Measurement errors can particularly affect students whose performance deviates from the mean achievement. However, in our studies, we included multiple achievement tests, thereby reducing the role of measurement errors. Future research on track recommendations should preferably also include multiple achievement tests since it reduces the possibility of measurement errors.

Quantitative Data

Moreover, the studies in this dissertation are all based on large-scale quantitative data. While this can provide important insights into general patterns and group differences, it does not provide insights into the decision-making processes itself, reasons behind particular choices of teachers or their (implicit) attitudes and stereotypes towards students from

different backgrounds. Qualitative research, for example, can provide deeper insights into teachers' personal considerations when formulating track recommendations.

Suggestions for Future Research

Examine the Interplay of Different Factors

The different studies in this dissertation have provided more insights into factors teachers consider in their track recommendations. Whereas previous research has typically focused on the individual contribution of different factors, the results of this dissertation have shown that it is of importance for researchers to also examine their interplay. In this dissertation, for example, we have shown that the interplay between students' SES and achievement inconsistencies had an effect on teachers' track recommendations beyond their individual effects (Chapter 3). It could be that teachers more often assign different factors or attributes to specific groups of students, and consider that in their track recommendations, or weigh these factors differently for different groups.

Examine Cumulative Effects of Different Factors

It would also be interesting for future research to examine the cumulative effects of different factors together. The different studies in this dissertation have shown which factors teachers consider in their track recommendations. However, each study focused on a particular set of factors. Combining them would create a more in-depth picture of the different factors and their cumulative impact on track recommendations. For instance, overall, lower-SES students perform lower on the school leavers' test, reading comprehension and mathematics tests but also more often showed achievement inconsistencies and lower work habits, all resulting in lower track recommendations. Additionally, we also found a small SES bias in the track recommendations, disadvantaging lower-SES students. Although these separate effects are small, they could create substantial differences in the track recommendations for lower- versus higher-SES students when they are considered altogether.

Examine School and Teacher Factors

Furthermore, the findings of this dissertation indicate that there are differences between teachers in how they consider different factors when formulating track recommendations. Therefore, we strongly recommend researchers to incorporate such random effects in their models. Otherwise, the role of certain factors may be underestimated. That is, non-significant results of factors predicting track recommendations may not always mean that teachers do not consider a certain factor when formulating their track recommendations, it can also be that some teachers weigh a certain factor positively and others negatively. For example, one teacher may give a student with parents who are perceived to be less involved a higher track recommendation. The reasoning may be that the student already performed at this level without parental support, showing that the student has a lot of potential ('the

benefit of the doubt'). Another teacher may think the exact opposite: if this student does not have a lot of parental support, they may struggle if they are placed at a higher track. Therefore, this teacher formulates a lower track recommendation ('careful'). As a result, two students with similar characteristics end up with different track recommendations. Such positive and negative effects may cancel each other out. Therefore, it is important for future research to include teacher or school level effects. Recent research (Timmermans et al., 2023) even suggests that the school board level should be included as well, since their study has shown large differences between school boards in their track recommendations. Future research could not only include a school (board) level in multilevel quantitative analyses, but it might also be interesting to examine differences in procedures and guidelines regarding the formulation of track recommendation, since schools are able to formulate their own procedures and guidelines, and that this might result in differences in track recommendations between primary schools.

Focus on Track Placement

Finally, in the present dissertation, we only focused on track recommendations of primary school teachers. However, the relation between the track recommendation and the actual track placement in secondary education is not always straightforward. Thereby, track placement may have an additional effect on students' educational careers on top of their track recommendations. Although secondary schools have been required since the 2014-2015 academic year to allocate students to the educational level of the track recommended by the primary school teacher rather than to the level indicated by the school leavers' test (Ministerie van Onderwijs Cultuur en Wetenschap, 2014), in practice secondary schools are not required to place students at a level that exactly matches their track recommendation. This means that students in the same secondary school class can have a diversity of track recommendations. Research (Inspectie van het Onderwijs, 2023) has shown that secondary schools take different factors into account when allocating a student to a secondary school track. For example, a large proportion of secondary schools (43%) indicated that they allocate pupils based on the score on the school leavers' test instead of the track recommendation. In addition, based on students' track recommendations, there appear to be many different options for track placement that are still in line with the track recommendation that students received since secondary schools differ in the number and types of tracks they offer. For instance, if a secondary school offers both heterogeneous and homogeneous classes, they can choose which class a student will be allocated to, as long as the recommended educational level is part of the class level. In this sense, a student with a senior general track *('havo'*) recommendation may end up in three different classes: (1) a heterogeneous pre-vocational theoretical/senior general track (*'vmbo-(g)t/havo'*), (2) a homogeneous senior general track (*'havo'*), or (3) a heterogeneous senior general/pre-university track (*'havo/vwo'*). And a student with a mixed track recommendation for pre-vocational theoretical/senior general education (*'vmbo-(g)t/havo'*), may even end up in five different classes: (1) a heterogeneous pre-vocational middle/theoretical track

(*'vmbo-k/vmbo-(g)t'*), (2) a homogeneous pre-vocational theoretical track (*'vmbo-(g)t'*), (3) a heterogeneous pre-vocational theoretical/senior general track (*'vmbo-(g)t/havo'*), (4) a homogeneous senior general track (*'havo'*), or (5) a heterogeneous senior general/pre-university track (*'havo/vwo'*). As a result, students with similar track recommendations may end up in very different secondary school classes. It is unknown to what extent these placement procedures contribute to unequal educational opportunities for students with different background characteristics.

Are we on the Right Track?

The findings of this dissertation - along with those of other studies - suggest that educational inequalities are not primarily caused by the track allocation process but are present prior to and during primary education and are solidified by the way in which our secondary educational system is organised. Because of early tracking, and the hierarchical and rigid system, the Dutch secondary school system offers few opportunities to reduce differences between students from different backgrounds. Although several policy changes have been made in the last decades, the achievement gap between students from different backgrounds persists. More drastic interventions than changes in the time schedule for administering the school leavers' test or mandatory versus non-mandatory adjustments of the track recommendations are needed to effectively reduce educational inequalities before and after the transition from primary to secondary education, including postponing the selection moment by introducing heterogeneous lower secondary education.

When the Teacher is in Doubt: Some Illustrative Cases

In the introduction of this dissertation, we discussed four 'real' examples of student pairs, derived from our own 'Utrecht' data, for whom teachers might find it difficult to formulate a track recommendation. We explained per student pair what the similar characteristics were, related to their achievement, student attributes and their background, and on which characteristic both students differed. The choice of these characteristics was related to the topics of the chapters of the present dissertation. We now return to these examples and describe which track recommendations these students received and discuss these cases in light of the overall findings of the dissertation.

Chapter 2: Socioeconomic Status

Destiny, a student from a lower-SES background, and Joanne, a student from a higher-SES background, had almost similar primary school achievement. Their achievement on the school leavers' test could suggest the pre-university track (*'vwo'*), but Joanne's reading comprehension and mathematics test results from the student monitoring system LVS were lower. Since Joanne performed lower, her teacher could have been in doubt about the level of education to recommend to Joanne: similar or lower (because of her lower achievement on the reading comprehension and mathematics tests) to that of Destiny. However, Joanne received a pre-university track (*'vwo'*) recommendation, whereas Destiny received a senior general track (*'havo'*) recommendation. This difference in track recommendations suggests, in line with our findings in Chapter 2, that on top of prior achievement, students' SES may (unintentionally) be considered by teachers. In this example, the lower-SES student Destiny received a lower track recommendation than the higher-SES student Joanne, while their prior achievement was similar, or even better for Destiny.

Chapter 3: Inconsistency in Achievement

Paul and Mark, both from a middle-SES background, had similar primary school achievement. Both boys showed substantial inconsistencies in their achievement with higher mathematics than reading comprehension achievement. Because of their inconsistencies in achievement, their teachers could have been in doubt about the level of education to recommend to the boys. Paul and Mark received different track recommendations while having similar prior achievement. Paul received a pre-vocational middle track (*'vmbo-k'*) recommendation and Mark a pre-vocational theoretical track (*'vmbo-t'*) recommendation. These differences in track recommendations suggest, in line with our findings in Chapter 3, that some teachers may give more careful (i.e. lower) track recommendations than others because of the inconsistent achievement. In this example, Paul received a more careful track recommendation than Mark.

Chapter 4: Achievement Growth

Olivier, from a higher-SES background, and Thomas, from a middle-SES background, had similar initial achievement and a similar growth curve for reading comprehension. However, although their mathematics achievement at the end of primary school was similar, their growth curves for mathematics were different with Olivier starting with high initial achievement and showing slower achievement growth, and Thomas starting with low initial achievement and showing faster achievement growth. Since the boys showed different growth curves for mathematics, their teachers could have been in doubt about the level of education to recommend to the boys. Despite their different growth curves, both boys received a pre-university track (*'havo'*) recommendation. These similar track recommendations suggest, in line with our findings in Chapter 4, that teachers do not consider achievement growth in their track recommendation on top of students' most recent achievement. In this example, Olivier and Thomas received similar track recommendations.

Chapter 5: Teachers' Perception of Student Attributes

Emma, from a higher-SES background, and Kelly, from a lower-SES background, had similar reading comprehension and mathematics achievement at the end of primary education, and similar results on the school leavers' test (suggesting a senior general track ('*havo*') recommendation). Emma's teacher considered her parents to to be highly involved in her schooling, while Kelly's teacher considered Kelly's parents to be less involved. Kelly's teacher may have doubted which track to recommend due to the perceived lack of involvement of Kelly's parents. Kelly received a pre-vocational middle track ('*vmbo-k*') recommendation, while Emma received a senior general track ('*havo*') recommendation. These differences in track recommendations suggest, in line with our findings in Chapter 5, that teachers consider their perception of parental involvement in their track recommendation on top of students' most recent achievement. In this example, Emma, the student with more involved parents, received a higher track recommendation than Kelly, the student with less involved parents. Despite the similarities in achievement, the difference in track recommendations between both students was quite large.

"Wij hadden vroeger wel de neiging om te hoge adviezen te geven: te hoog in te zetten. Bij twijfel denk ik: laat ze het maar proberen. Tot we bij een aantal leerlingen, waarbij we echt twijfelden welke kant het op zou gaan, zagen dat het niet goed ging, die hadden wat meer onvoldoendes op de rapporten in het vo. Ik denk dat het komt sinds wij de grootste invloed hebben, sinds ons advies bindend is. Voorheen was het nog even spannend, want dan konden vo-scholen ook nog naar de scores en het hele dossier kijken. Tot het moment van plaatsing konden leerlingen nog op niveau afgewezen worden. Dat kan nu niet meer."

References

Aarnoutse, C. A. J. (2017). Onderwijs in begrijpend lezen. [Reading comprehension instruction]. *Orthopedagogiek: Onderzoek en Praktijk, 56*(11–12), 269–291. https://newsroom.didactiefonline.nl/uploads/PDF/Cor%20Aarnoutse%20Onderwijs%20in%20begrijpend%20lezen.%20OOP%20dec.%202017.-2.pdf

Aarnoutse, C. A. J., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation, 6*(3), 251–278. https://doi.org/10.1076/1380-3611(200009)6:3;1-a;ft251

Ağirdağ, O., Biesta, G., Bosker, R. J., Kuiper, R., Nieveen, N., Raijmakers, M., & van Tartwijk, J. (2021). *Examenprogramma's in perspectief: Tussenadvies 3 wetenschappelijke curriculumcommissie.* [Interim report examination programmes in perspective]. Curriculumcommissie. https://www.rijksoverheid.nl/documenten/rapporten/2021/06/30/examenprogrammas-in-perspectief

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*(3), 317–332. https://doi.org/10.1007/BF02294359

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*(4), 731–746. https://doi.org/10.1037/0022-0663.91.4.731

Au, W. (2013). Hiding behind high-stakes testing: Meritocracy, objectivity and inequality in U.S. education. *International Education Journal, 12*(2), 7–20. https://openjournals.library.sydney.edu.au/IEJ/article/view/7453/7812

Awang, Z. (2012). *A handbook on structural equation modeling using AMOS.* Universiti Technologi MARA Press.

Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology, 74*(4), 459–474. https://doi.org/10.1037/0022-0663.74.4.459

Bakker, J., Denessen, E., & Brus-Laeven, M. (2007). Socio-economic background, parental involvement and teacher perceptions of these in relation to pupil achievement. *Educational Studies, 33*(2), 177–192. https://doi.org/10.1080/03055690601068345

Barg, K. (2013). The influence of students' social background and parental involvement on teachers' school track choices: Reasons and consequences. *European Sociological Review, 29*(3), 565–579. https://doi.org/10.1093/esr/jcr104

Batruch, A., Geven, S., Kessenich, E., & van de Werfhorst, H. G. (2023). Are tracking recommendations biased? A review of teachers' role in the creation of inequalities in tracking decisions. *Teaching and Teacher Education, 123,* Article 103985. https://doi.org/10.1016/j.tate.2022.103985

Bauer, P., & Riphahn, R. T. (2006). Timing of school tracking as a determinant of intergenerational transmission of education. *Economics Letters, 91*(1), 90–97. https://doi.org/10.1016/j.econlet.2005.11.003

Beaman, R., Wheldall, K., & Kemp, C. (2006). Differential teacher attention to boys and girls in the classroom. *Educational Review, 58*(3), 339–366. https://doi.org/10.1080/00131910600748406

Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2006). Leistungszuwachs in Mathematik: Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? [Differential achievement trajectories in mathematics in a tracked school system]. *Zeitschrift für Pädagogische Psychologie, 20*(4), 233–242. https://doi.org/10.1024/1010-0652.20.4.233

Becker, R., & Schubert, F. (2006). Soziale Ungleichheit von Lesekompetenzen: Eine Matching-Analyse im Längsschnitt mit Querschnittsdaten von PIRLS 2001 und PISA 2000. [Social inequality of reading literacy]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 58*(2), 253–284. https://doi.org/10.1007/s11575-006-0055-0

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology, 85*(2), 347–356. https://doi.org/10.1037/0022-0663.85.2.347

Bentler, P. M. (1990). Comparative fit indices in structural equation models. *Psychological Bulletin, 107*(2), 238–246. https://escholarship.org/uc/item/2mk8r49v

Bittmann, F., & Schindler, S. (2021). Analysing diversion processes in German decondary education: School-track effects on educational aspirations. *Kolner Zeitschrift fur Soziologie und Sozialpsychologie, 73*(2), 231–257. https://doi.org/10.1007/s11577-021-00789-1

Böhmer, M., Glock, S., Gräsel, C., Hörstermann, T., & Krolak-Schwerdt, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? [An analysis of information search in the process of making school tracking decisions]. *Journal for Educational Research Online, 7*(2), 59–81. https://doi.org/10.25656/01:11490

Bol, T., Witschge, J., van de Werfhorst, H. G., & Dronkers, J. (2014). Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces, 92*(4), 1545–1572. https://doi.org/10.1093/sf/sou003

Bonizzoni, P., Romito, M., & Cavallo, C. (2016). Teachers' guidance, family participation and track choice: The educational disadvantage of immigrant students in Italy. *British Journal of Sociology of Education, 37*(5), 702–720. https://doi.org/10.1080/01425692.2014.982860

Boone, S., & Demanet, J. (2020). Track choice, school engagement and feelings of perceived control at the transition from primary to secondary school. *British Educational Research Journal, 46*(5), 929–948. https://doi.org/10.1002/berj.3606

Boone, S., Thys, S., Van Avermaet, P., & Van Houtte, M. (2018). Class composition as a frame of reference for teachers? The influence of class context on teacher recommendations. *British Educational Research Journal, 44*(2), 274–293. https://doi.org/10.1002/berj.3328

Boone, S., & Van Houtte, M. (2013). Why are teacher recommendations at the transition from primary to secondary education socially biased? A mixed-methods research. *British Journal of Sociology of Education, 34*(1), 20–38. https://doi.org/10.1080/01425692.2012.704720

Boonk, L., Gijselaers, H. J. M., Ritzen, H., & Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational Research Review, 24*(1), 10–30. https://doi.org/10.1016/j.edurev.2018.02.001

Borg, E. (2015). Classroom behaviour and academic achievement: How classroom behaviour categories relate to gender and academic performance. *British Journal of Sociology of Education, 36*(8), 1127–1148. https://doi.org/10.1080/01425692.2014.916601

Borghans, L., & Diris, R. (2021). Ongelijkheid in het Nederlandse onderwijs door de jaren heen. In A. Gielen, D. Webbink, & B. Ter Weel (Eds.), *Ongelijk Nederland: Een drieluik over migratie, onderwijs en de coronacrisis* (pp. 37–43). Koninklijke Vereniging voor de Staathuishoudkunde. https://esb.nu/wp-content/uploads/2022/11/GA0E4KSbRoTYQBHokxPbKQmKMys.pdf

Borghans, L., Diris, R., Smits, W., & de Vries, J. (2012). The impact of early tracking on later-life outcomes: An instrumental variable approach (*Discussion Paper,* Vol. 7, Issue 143). Society of Labor Economists. https://www.sole-jole.org/assets/docs/13506.pdf

Borghans, L., Diris, R., Smits, W., & de Vries, J. (2019). The long-run effects of secondary school track assignment. *PLoS ONE, 14*(10), e0215493. https://doi.org/10.1371/journal.pone.0215493

Borghans, L., Diris, R., Smits, W., & de Vries, J. (2020). Should we sort it out later? The effect of tracking age on long-run outcomes. *Economics of Education Review, 75,* Article 101973. https://doi.org/10.1016/j.econedurev.2020.101973

Bos, W., Lankes, E. M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (2004). *Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich.* [Germany in national and international comparison]. Waxmann.

Bosker, R. J., Hofman, W. H. A., & van der Velden, R. K. W. (1985). *Een generatie geselecteerd: Deel 1: De loopbanen.* [Educational careers of a selected generation]. RION.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*(1), 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Brame, R., Paternoster, R., Mazerolle, P., & Piquero, A. (1998). Testing for the equality of maximum-likelihood regression coefficients between two independent equations. *Journal of Quantitative Criminology, 14*(3), 245–261. https://doi.org/10.1023/A:1023030312801

Brandmiller, C., Dumont, H., & Becker, M. (2020). Teacher perceptions of learning motivation and classroom behavior: The role of student characteristics. *Contemporary Educational Psychology, 63*, Article 101893. https://doi.org/10.1016/j.cedpsych.2020.101893

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*(5), 631–661. https://doi.org/10.1037/0022-0663.75.5.631

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy, 22*(52), 781–861. https://doi.org/10.1111/j.1468-0327.2007.00189.x

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education, 8*(1), 158–233. https://doi.org/10.2307/1167125

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows.* Sage Publications.

Cain, M. K., & Zhang, Z. (2019). Fit for a Bayesian: An evaluation of PPP and DIC for structural equation modeling. *Structural Equation Modeling, 26*(1), 39–50. https://doi.org/10.1080/10705511.2018.1490648

Caro, D. H., & Lehmann, R. (2009). Achievement inequalities in Hamburg schools: How do they change as students get older? *School Effectiveness and School Improvement, 20*(4), 407–431. https://doi.org/10.1080/09243450902920599

Caro, D. H., Lenkeit, J., Lehmann, R., & Schwippert, K. (2009). The role of academic achievement growth in school track recommendations. *Studies in Educational Evaluation, 35*(4), 183–192. https://doi.org/10.1016/j.stueduc.2009.12.002

Caro, D. H., McDonald, J. T., & Willms, J. D. (2009). Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education, 32*(3), 558–590. https://doi.org/10.1017/CBO9781107415324.004

Centraal Bureau voor de Statistiek. (2022, April 7). *Statusscore per wijk en buurt obv welvaart, opleidingsniveau en arbeid.* [Status score by neighbourhood and district]. https://www.cbs.nl/nl-nl/achtergrond/2022/14/statusscore-per-wijk-en-buurt-o-b-v-welvaart-opleidingsniveau-en-arbeid

Chzhen, Y., Rees, G., Gromada, A., Cuesta, J., & Bruckauf, Z. (2018). *An unfair start: Inequality in children's education in rich countries* (M. Drohan (Ed.); Innocenti Report Card No. 15). UNICEF. https://www.unicef-irc.org/publications/pdf/an-unfair-start-inequality-children-education_37049-RC15-EN-WEB.pdf

Cito. (n.d.-a). *Historie: Cito sinds 1968.* [History: Cito since 1968]. https://cito.nl/over-cito/historie/

Cito. (n.d.-b). *Leerling in beeld PO.* [Student monitoring system primary education]. https://cito.nl/onderwijs/primair-onderwijs/leerling-in-beeld-lvs/

Cito. (n.d.-c). *Leerling in beeld VO*. [Student monitoring system secondary education]. https://cito.nl/onderwijs/voortgezet-onderwijs/leerling-in-beeld-vo/

Cito. (2016). *Tabellen tussenopbrengsten Cito LVS.* [Results of student monitoring tests in primary education]. http://www.sbzw.nl/userfiles/CITO_tabellen_tussenopbrengsten_- januari_2016.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

College voor Toetsen en Examens. (2015). *Verantwoording: Centrale eindtoets po.* [Accountability of end of primary education test]. https://www.centraleeindtoetspo.nl/binaries/centraleeindtoets/documenten/publicaties/2015/05/01/verantwoording-centrale-eindtoets-po/Centrale-Eindtoets_CvTE_verantwoording_12_mei_2015.pdf

Constantino, R. (2005). Print environments between high and low socioeconomic (SES) communities. *Teacher Librarian, 32*(3), 22–26. https://eric.ed.gov/?id=EJ711281

Contini, D., & Scagni, A. (2011). Inequality of opportunity in secondary school enrolment in Italy, Germany and the Netherlands. *Quality & Quantity, 45*(2), 441–464. https://doi.org/10.1007/s11135-009-9307-y

COOL5-18. (n.d.). *COOL5-18 cohortonderzoek onderwijsloopbanen.* [COOL5-18 cohort research on educational careers]. http://cool5-18.kohnstamminstituut.nl

Crosnoe, R., Morrison, F., Burchinal, M., Pianta, R. C., Keating, D., Friedman, S. L., & Clarke-Stewart, K. A. (2010). Instruction, teacher-student relations, and math achievement trajectories in elementary school. *Journal of Educational Psychology, 102*(2), 407–417. https://doi.org/10.1037/a0017762

Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009). Teacher-child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology, 101*(4), 912–925. https://doi.org/10.1037/a0016647

Danesh, J., Gault, S., Semmence, J., Appleby, P., & Peto, R. (1999). Postcodes as useful markers of social class: Population based study in 26,000 British households. *Journal of Epidemiology and Community Health, 33*(2), 843–845. https://doi.org/10.1136/jech.53.9.582

de Boer, H. (2009). *Schoolsucces van Friese leerlingen in het voortgezet onderwijs.* [School success of Frisian students in seconday education] [Doctoral Dissertation, Rijksuniversiteit Groningen]. https://hdl.handle.net/11370/d19b6683-4825-4a01-ab85-79150c6ef646

de Boer, H., Bosker, R. J., & van der Werf, M. P. C. G. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology, 102*(1), 168–179. https://doi.org/10.1037/a0017289

De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal, 29*(6), 841–859. https://doi.org/10.1080/0141192032000137330

de Geus, W., & Bisschop, P. (2018). *Licht op schaduwonderwijs: Onderzoek naar deelname aan en uitgaven voor schaduwonderwijs.* [Light on shadow education]. SEO economisch onderzoek, Oberon onderzoek/advies. https://www.oberon.eu/media/5fghohe0/licht-op-schaduwonderwijs.pdf

de Jong, W., & Steenbeeke, L. (2020). *Een derde van scholieren na drie jaar op ander niveau dan schooladvies.* [Different level of education after three years of secondary education]. NOS. https://nos.nl/artikel/2322854-een-derde-van-scholieren-na-drie-jaar-op-ander-niveau-dan-schooladvies.html

de Rooy, P. (2018). *Een geschiedenis van het onderwijs in Nederland.* [A history of education in the Netherlands]. Wereldbibliotheek.

de Winter-Koçak, S., & Reches, L. (2022). *Gelijke kansen bij tussentijds opstromen in het voortgezet onderwijs? Een verkennend onderzoek naar beleid en praktijk van tussentijds opstromen.* [Equal opportunities in mid-term transfer in secondary education?]. Kennisplatform inclusief samenleven. https://www.kis.nl/sites/default/files/2022-06/gelijke-kansen-bij-tussentijds-opstromen-voortgezet-onderwijs.pdf

Deckers, I. E., Janse, I. C., van der Zee, H. H., Nijsten, T., Boer, J., Horváth, B., & Prens, E. P. (2016). Hidradenitis suppurativa (HS) is associated with low socioeconomic status (SES): A cross-sectional reference study. *Journal of the American Academy of Dermatology, 75*(4), 755-759.e1. https://doi.org/10.1016/j.jaad.2016.04.067

Dee, T. (2005). A teacher like me: Does race, ethnicity, or gender matter? The American Economic Review, 95(2), 157–165. https://www.jstor.org/stable/4132809

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A. M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research, 87*(2), 243–282. https://doi.org/10.3102/0034654316687036

Dijks, M. A., Warrens, M. J., Fleur, E., Korpershoek, H., Wichgers, I. J. M., & Bosker, R. J. (2020). The predictive power of track recommendations in Dutch secondary education. *Pedagogische Studiën, 97*, 263–280. https://hdl.handle.net/11370/9bf0a048-19c9-4acd-8115-247e8466a4a6

Ditton, H., & Krüsken, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe I. [The transition from primary school to lower secondary level]. *Zeitschrift Für Erziehungswissenschaft, 9*(3), 348–372. https://doi.org/10.1007/s11618-006-0055-7

Dockx, J., De Fraine, B., & Vandecandelaere, M. (2019). Does the track matter? A comparison of students' achievement in different tracks. *Journal of Educational Psychology, 111*(5), 827–846. https://doi.org/10.1037/edu0000305

Dollmann, J. (2016). Less choice, less inequality? A natural experiment on social and ethnic differences in educational decision-making. *European Sociological Review, 32*(2), 203–215. https://doi.org/10.1093/esr/jcv082

Dopmeijer, J., Scheeren, L., van Baar, J., & Bremer, B. (2023). *Harder, better, faster, stronger? Een onderzoek naar risicofactoren en oplossingen voor prestatiedruk en stress van studenten in het hbo en wo.* (p. 156). [Risk factors and solutions for students' performance pressure and stress]. Trimbos-instituut, ECIO, & RIVM. https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2023/06/14/rapportage-harder-better-faster-stronger/rapportage-harder-better-faster-stronger.pdf

Driessen, G. (2005). De totstandkoming van de adviezen voortgezet onderwijs: Invloeden van thuis en school. [The impact of home and school when formulating track recommendations]. *Pedagogiek, 25*(4), 279–298. https://journal-archive.aup.nl/pedagogiek/geert_driessen,_de_totstandkoming_van_de_adviezen_voortgezet_onderwijs.pdf

Driessen, G. (2006). Het advies voortgezet onderwijs: Is de overadvisering over? [Secondary school track overrecommendation]. *Mens & Maatschappij, 81*(1), 5–23. https://journal-archive.aup.nl/mens-en-maatschappij/vol_81_no_1_-_het_advies_voortgezet_onderwijs.pdf

Driessen, G. (2011a). *Onderadvisering van allochtone leerlingen?* [Underrecommendation of minority students?]. Inspectie van het Onderwijs, Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.geertdriessen.nl/wp-content/uploads/2014/11/rap2011-onder-advisering-driessen.pdf

Driessen, G. (2011b). *The teacher recommendation: A valid instrument for selecting ethnic minority students into secondary school tracks?* [Conference Paper], ECER annual meeting, Berlin, Germany. https://www.researchgate.net/publication/236004048_The_teacher_recommendation_A_valid_instrument_for_selecting_ethnic_minority_students_into_secondary_school_tracks

Driessen, G. (2012). *De doelgroepen van het onderwijsachterstandenbeleid: Ontwikkelingen in prestaties en het advies voortgezet onderwijs.* [The target population of the educational deprivation policy]. ITS, Radboud Universiteit Nijmegen. http://www.geertdriessen.nl/wp-content/uploads/2014/11/rap2012-doelgroepen-onderwijsachterstandenbeleid-driessen.pdf

Driessen, G., Doesborgh, J., Ledoux, G., Overmaat, M., Roeleveld, J., & van der Veen, I. (2005). *Van basis- naar voortgezet onderwijs.* [From primary to secondary education]. ITS, Radboud Universiteit Nijmegen, & SCO-Kohnstamm Instituut, Universiteit van Amsterdam. http://hdl.handle.net/2066/211591

Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J., & van der Veen, I. (2009). Cohortonderzoek COOL5-18: Technisch rapport meting basisonderwijs, eerste meting 2007/08. [Technical report COOL5-18 data collection primary education 2007]. ITS, Radboud Universiteit Nijmegen, & SCO-Kohnstamm Instituut, Universiteit van Amsterdam. https://hdl.handle.net/2066/210877

Driessen, G., Mulder, L., & Roeleveld, J. (2012). *Cohortonderzoek COOL5-18: Technisch rapport basisonderwijs, tweede meting 2010/11.* [Technical report COOL5-18 data collection primary education 2010]. ITS, Radboud Universiteit Nijmegen, & SCO-Kohnstamm Instituut, Universiteit van Amsterdam. http://www.geertdriessen.nl/wp-content/uploads/2014/11/rap2012-cohortonderzoekCool5-18_2-driessenea.pdf

Driessen, G., Sleegers, P., & Smit, F. (2008). The transition from primary to secondary education: Meritocracy and ethnicity. *European Sociological Review, 24*(4), 527–542. https://doi.org/10.1093/esr/jcn018

Driessen, G., Smeets, E., Mulder, L., & Vierke, H. (2007). *De relatie tussen prestaties en advies: Onder- of overadvisering bij de overgang van basis- naar voortgezet onderwijs.* [The relation between performance and track recommendation]. In Onderadvisering in beeld (pp. 59–81). ITS, Radboud Universiteit Nijmegen. https://hdl.handle.net/2066/211423

Driessen, G., Van Langen, A., & Vierke, H. (2006). *Basisonderwijs: Veldwerkverslag, leerling-gegevens en oudervragenlijsten. Basisrapportage PRIMAcohortonderzoek. Zesde meting 2004/05.* [Report PRIMA data collection primary education 2004]. ITS, Radboud Universiteit Nijmegen. http://hdl.handle.net/2066/211473

Dronkers, J., van Erp, M., Robijn, M., & Roeleveld, J. (1998). *Krijgen leerlingen in de grote steden en met name in Amsterdam te hoge adviezen? De relaties tussen taal- en rekenscores en advies binnen en buiten de Randstand onderzocht.* [Do students in big cities receive too high school track recommendations?]. Tijdschrift voor Onderwijsresearch, 23, 17–30.

Dumont, H., Klinge, D., & Maaz, K. (2019). The many (subtle) ways parents game the system: Mixed-method evidence on the transition into secondary-school tracks in Germany. *Sociology of Education, 92*(2), 199–228. https://doi.org/10.1177/0038040719838223

Dyslexie Centraal. (2023a). *Toetsen en observeren van achterstanden.* [Testing and observation of backlogs]. https://dyslexiecentraal.nl/weten/signaleren-van-achterstanden/toetsen-en-observeren-van-achterstanden

Dyslexie Centraal. (2023b). *Toetsen in het basisonderwijs.* [Testing in primary education]. https://dyslexiecentraal.nl/weten/wet-en-regelgeving/toetsen-het-basisonderwijs

EenVandaag, & CNV Onderwijs. (2018). *Onderzoek: Leraren onder druk om schooladvies.* [Teachers under pressure because of track recommendation]. https://eenvandaag.assets.avrotros.nl/user_upload/PDF/Rapportage_Peiling_Leraren_Schooladvies.pdf

Eggen, T. J. H. M. (2004). *Contributions to the theory and practice of computerized adaptive testing* [Doctoral Dissertation, Universiteit Twente]. https://ris.utwente.nl/ws/portalfiles/portal/6120151/thesis_Eggen.pdf

Elffers, L. (2017). *De bijlesgeneratie: Opkomst van de onderwijscompetitie.* [The tutoring generation]. Amsterdam University Press.

Elffers, L., & Jansen, D. (2019). *De opkomst van schaduwonderwijs in Nederland: Wat weten we en welke vragen liggen nog open?* [The rise of shadow education in the Netherlands: What do we know and what questions remain open?]. Universiteit van Amsterdam. https://www.nro.nl/sites/nro/files/migrate/rapport-schaduwonderwijs-Elffers-Jansen-2019.pdf

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Entwisle, D. R., Alexander, K. L., & Steffel Olsen, L. (1994). The gender gap in math: Its possible origins in neighborhood effects. American Sociological Review, 59(6), 822–838. https://doi.org/https://doi.org/10.2307/2096370

Exalto, R., Klein, T., & Kooij, D. (2018). *Terecht overstaprecht? Doorstroom havo-vwo.* [The right to transfer from havo to vwo]. Oberon onderzoek/advies. https://www.oberon.eu/media/sukbxip4/terecht-overstaprecht-doorstroom-havo-vwo.pdf

Faasse, J. H., Bakker, B., Dronkers, J., & Schijf, H. (1987). The impact of educational reform: Empirical evidence from two Dutch generations. *Comparative Education, 23*(3), 261–277. https://doi.org/10.1080/0305006870230302

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *Model selection criterion: AIC and BIC.* In The basics of financial econometrics: Tools, concepts, and asset management applications (pp. 399–403). John Wiley & Sons. https://doi.org/10.1002/9781118856406

Feenstra, H., Kamphuis, F., Kleintjes, F., & Krom, R. (2010). *Leerling- en onderwijsvolgsysteem: Begrijpend lezen groep 3 t/m 6.* [Reading comprehension in grade 1-6]. Cito.

Feron, E., Schils, T., & ter Weel, B. (2013). *Test scores, teacher assessment and track placement in a secondary school system with early tracking.* http://www.iwaee.org/PaperValidi2014/20140226204440_Feron Schils terWeel.pdf

Feron, E., Schils, T., & ter Weel, B. (2016). Does the teacher beat the test? The value of teacher's assessment in predicting student ability. *De Economist, 164*(4), 391–418. https://doi.org/10.1007/s10645-016-9278-z

Feskens, R., Kuhlemeier, H., & Limpens, G. (2016). *Resultaten PISA-2015 in vogelvlucht: Praktische kennis en vaardigheden van 15-jarigen.* [Results of PISA-2015 in a nutshell]. Cito. https://cito.nl/media/qp3ban3a/resultaten-pisa-2015-in-vogelvlucht-praktische-kennis-en-vaardigheden-van-15-jarigen.pdf

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from catergory-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gamoran, A. (2009). Tracking and inequality: New directions for research and practice. In M. W. Apple, S. J. Ball, & L. A. Gandin (Eds.), *The routledge international handbook of the sociology of education* (pp. 213–228). Routeledge International Handbooks. https://doi.org/10.4324/9780203863701

Gelman, A. (2013). Two simple examples for understanding posterior p-values whose distributions are far from unform. *Electronic Journal of Statistics, 7*(1), 2595–2602. https://doi.org/10.1214/13-EJS854

Gentrup, S., & Rjosk, C. (2018). Pygmalion and the gender gap: do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling? *Educational Research and Evaluation, 24*(3–5), 295–323. https://doi.org/10.1080/13803611.2018.1550840

Geven, S., Batruch, A., & van de Werfhorst, H. G. (2018). *Inequality in teacher judgements, expectations and track recommendations: A review study.* Universiteit van Amsterdam. https://zoek.officielebekendmakingen.nl/blg-864911

Geven, S., Wiborg, Ø. N., Fish, R. E., & van de Werfhorst, H. G. (2021). How teachers form educational expectations for students: A comparative factorial survey experiment in three institutional contexts. *Social Science Research, 100*, Article 1102599. https://doi.org/10.1016/j.ssresearch.2021.102599

Glock, S. (2016a). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers' judgments of students. Social Psychology of Education, 19(3), 493–509. https://doi.org/10.1007/s11218-016-9349-7

Glock, S. (2016b). Stop talking out of turn: The influence of students' gender and ethnicity on preservice teachers' intervention strategies for student misbehavior. *Teaching and Teacher Education, 56*, 106–114. https://doi.org/10.1016/j.tate.2016.02.012

Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education, 16*(1), 111–127. https://doi.org/10.1007/s11218-012-9197-z

Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2012). Improving teachers' judgments: Accountability affects teachers' tracking decisions. *International Journal of Technology and Inclusive Education, 1*(2), 89–98. https://doi.org/10.20533/ijtie.2047.0533.2012.0012

Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013). Beyond judgment bias: How students' ethnicity and academic profile consistency influence teachers' tracking judgments. *Social Psychology of Education, 16*(4), 555–573. https://doi.org/10.1007/s11218-013-9227-5

Glock, S., Krolak-Schwerdt, S., & Pit-ten Cate, I. M. (2015). Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory. *European Journal of Psychology of Education, 30*(2), 169–188. https://doi.org/10.1007/s10212-014-0237-2

Good, T. L., & Brophy, J. E. (1997). *Looking in classrooms* (7th ed.). Longman.

Grift, Y., Rümke, T., & Verberne, P. (2010). *De invloed van heterogene of homogene brugklassen op de keuze tussen vmbo-tl en havo/vwo.* [The influence of heterogeneous or homogeneous classes on the choice between vmbo-tl and havo/vwo]. TPEdigitaal, 4(3), 101–125. https://www.tpedigitaal.nl/sites/default/files/bestand/de_invloed_van_heterogene_of_homogene_brugklassen_op_de_keuze_tussen_vmbo-tl_en_havovwo.pdf

Guhn, M., Gadermann, A. M., Hertzman, C., & Zumbo, B. D. (2010). Children's development in kindergarten: A multilevel, population-based analysis of ESL and gender effects on socioeconomic gradients. *Child Indicators Research, 3*(2), 183–203. https://doi.org/10.1007/s12187-009-9053-7

Hakkarainen, A., Holopainen, L., & Savolainen, H. (2013). Mathematical and reading difficulties as predictors of school achievement and transition to secondary education. *Scandinavian Journal of Educational Research, 57*(5), 488–506. https://doi.org/10.1080/00313831.2012.696207

Hallinan, M. T. (1994). Tracking: From theory to practice. *Sociology of Education, 67*(2), 79–84. https://doi.org/10.2307/2112697

Hamaker, E., van Hattum, P., Kuiper, R., & Hoijtink, H. (2011). Model selection based on information criteria in multilevel modeling. In J. Hox, Joop & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 231–255). Routledge/Taylor & Francis Group.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*(2), 625–638. https://doi.org/10.1111/1467-8624.00301

Hanushek, E. A., & Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal, 116*(510), C63–C76. https://doi.org/10.1111/j.1468-0297.2006.01076.x

Hanushek, E. A., & Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal, 116*(510), C63–C76. https://doi. org/10.1111/j.1468-0297.2006.01076.x

Hartgers, M., Traag, T., & Wielenga, L. (2021). *De schooladviezen in groep 8: Verschillen tussen groepen leerlingen.* [Differences in track recommendations between groups of students]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/longread/statistische-trends/2021/de-schooladviezen-in-groep-8-verschillen-tussen-groepen-leerlingen?onepage=true#c-3--Resultaten-van-het-onderzoek

Hebbink, P., Warrens, M. J., Fleur, E., Dijks, M. A., & Korpershoek, H. (2022). De voorspellende waarde van het initiële schooladvies, het toetsadvies en het definitieve schooladvies in het Nederlandse onderwijs. [The predictive power of track recommendations]. *Pedagogische Studiën, 99*(1), 20–35. https://pedagogischestudien.nl/article/view/13725

Heemskerk, I., van Eck, E., Kuiper, E., & Volman, M. (2012). *Succesvolle onderwijsaanpakken voor jongens in het vo.* [Successful educational approaches for boys in secondary school]. Kohnstamm Instituut, Universiteit van Amsterdam. https://kohnstamminstituut.nl/wp-content/uploads/2019/08/878-Succesvolle-onderwijsaanpakken-voor-jongens-in-het-vo.pdf

Heiskala, L., & Erola, J. (2019). *Tracking and sorting in the Finnish educational system* (No. 12; DIAL Working Papers). https://dynamicsofinequality.org/wp-content/uploads/2019/07/LIFETRACK-country-report_FIN-1.pdf

Helbling, L. A., Tomasik, M. J., & Moser, U. (2019). Long-term trajectories of academic performance in the context of social disparities: Longitudinal findings from Switzerland. *Journal of Educational Psychology, 111*(7), 1284–1299. https://doi.org/10.1037/edu0000341

Hjort, N. L., Dahl, F. A., & Steinbakk, G. H. (2006). Post-processing posterior predictive p values. *Journal of the American Statistical Association, 101*(475), 1157–1174. https://doi.org/10.1198/016214505000001393

Ho, J., & Lee, Y.-J. (2022). Primary and secondary education in Singapore: Bringing out the best in every learner. In Y.-J. Lee (Ed.), *Education in the Asia-Pacific region: Issues, concerns and prospects: Education in Singapore.* (Vol. 66, pp. 49–68). Springer. https://doi.org/10.1007/978-981-16-9982-5_5

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–ai complementarity. *Journal of Learning Analytics, 6*(2), 27–52. https://doi.org/10.18608/jla.2019.62.3

Hoover-Dempsey, K. V, Bassler, O. C., & Brissie, J. S. (1992). Explorations in parent-school relations. *The Journal of Educational Research, 85*(5), 287–294. https://www.jstor.org/stable/27540488

Hopwood, B., Hay, I., & Dyment, J. (2016). The transition from primary to secondary school: Teachers' perspectives. *The Australian Educational Researcher, 43*(3), 289–307. https://doi.org/10.1007/s13384-016-0200-0

Horn, D. (2013). Diverging performances: The detrimental effects of early educational selection on equality of opportunity in Hungary. *Research in Social Stratification and Mobility, 32*(1), 25–43. https://doi.org/10.1016/j.rssm.2013.01.002

Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2013). Developments in motivation and achievement during primary school: A longitudinal study on group-specific differences. *Learning and Individual Differences, 23*(1), 195–204. https://doi.org/10.1016/j.lindif.2012.09.004

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd Ed). Taylor & Francis Ltd.

Hu, L., & Bentler, P. M. (1995). *Evaluating model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues and applications.* (pp. 76–99). Sage Publications.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hustinx, P. W. J. (2002). School careers of pupils of ethnic minority background after the transition to secondary education: Is the ethnic factor always negative? *Educational Research and Evaluation, 8*(2), 169–195. https://doi.org/10.1076/edre.8.2.169.3860

Hustinx, P. W. J., Kuyper, H., van der Werf, M. P. C., & Dijkstra, P. (2009). Achievement motivation revisited: New longitudinal data to demonstrate its predictive power. *Educational Psychology, 29*(5), 561–582. https://doi.org/10.1080/01443410903132128

Inspectie van het Onderwijs. (2007). *Aansluiting voortgezet onderwijs op het basisonderwijs.* [Interconnection between primary and secondary education]. https://zoek.officielebekendmakingen.nl/kst-30800-VIII-127-b4.pdf

Inspectie van het Onderwijs. (2012). *De beoordeling van opbrengsten in het basisonderwijs.* [The assessment of educational achievement in primary education]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/brochures/2012/02/15/de-beoordeling-van-opbrengsten-in-het-basisonderwijs-printversie/printversie-beoordeling-opbrengsten-basisonderwijs-2012.pdf

Inspectie van het Onderwijs. (2014). *De kwaliteit van het basisschooladvies: Een onderzoek naar de totstandkoming van het basisschooladvies en de invloed van het basisschooladvies op de verdere schoolloopbaan.* [The quality of the track recommendation]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/publicaties/2014/10/14/de-kwaliteit-van-het-basisschooladvies/de-kwaliteit-van-het-basisschooladvies.pdf

Inspectie van het Onderwijs. (2016). *De staat van het onderwijs 2016: Onderwijsverslag 2014/2015.* [Report of education 2016]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/publicaties/2016/04/13/staat-van-het-onderwijs-2014-2015/de-staat-van-het-onderwijs-2014-2015.pdf

Inspectie van het Onderwijs. (2017a). *De staat van het onderwijs 2017: Infographic niveau en prestaties onderwijsverslag 2015/2016.* [Infographic Dutch education 2017]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/infographics/2017/4/12/niveau-en-prestaties/trends-2015-2016-svho-1-niveau-en-prestaties.pdf

Inspectie van het Onderwijs. (2017b). *De staat van het onderwijs 2017: Technisch rapport onderwijsverslag 2015/2016 hoofdstuk 1: Niveau en prestaties.* [Level of education and achievement of Dutch education 2017]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2017/04/12/technisch-rapport-gelijke-kansen/Technisch+rapport+Gelijke+kansen.pdf

Inspectie van het Onderwijs. (2018a). *De staat van het onderwijs 2018: Onderwijsverslag 2016/2017.* [Report of education 2018]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2018/04/11/rapport-de-staat-van-het-onderwijs/108126_IvhO_Staatvanhet Onderwijs_TG.pdf

Inspectie van het Onderwijs. (2018b). *De staat van het onderwijs 2018: Technisch rapportage onderwijskansen en segregatie onderwijsverslag 2016/2017.* [Technical report of educational opportunities and segregation 2018]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2018/04/11/technische-rapporten-hoofdlijnen/TECHNISCH+RAPPORT+ONDERWIJSKANSEN+EN+SEGREGATIE.pdf

Inspectie van het Onderwijs. (2018c). *Jaarwerkplan 2018: Doelen en activiteiten Inspectie van het Onderwijs.* [Annual work plan 2018]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/jaarplannen/2017/11/08/jaarwerkplan-2018/WEB_Jaarwerkplan+2018.pdf

Inspectie van het Onderwijs. (2018d). *Kansen(on)gelijkheid bij de overgangen po-vo: Bevindingen en bevorderende en belemmerende factoren.* [(In)equality of opportunity at the transition from primary to secondary education]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2018/06/13/onderzoeksverslag-kansenongelijkheid-bij-de-overgangen-po-vo/Kansen-on-gelijkheid_bij_de_overgangen_po-vo.pdf

Inspectie van het Onderwijs. (2019). *De staat van het onderwijs 2017/2018: Hoofdlijnen.* [Headlines report of education 2019]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2019/04/10/hoofdlijnen-de-staat-van-het-onderwijs-2019/116838_IvhO_StaatvanhetOnderwijs2019_H1_TG+A.pdf

Inspectie van het Onderwijs. (2020). *De staat van het onderwijs 2020: Onderwijsverslag 2018/2019.* [Report of education 2020]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2020/04/22/staat-van-het-onderwijs-2020/Staat+van+het+Onderwijs+2020.pdf

Inspectie van het Onderwijs. (2021a). *De staat van het onderwijs 2021: Onderwijsverslag 2019/2020.* [Report of education 2021]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2021/04/14/de-staat-van-het-onderwijs-2021/Staat-van-het-Onderwijs-2021.pdf

Inspectie van het Onderwijs. (2021b). E*erste uitkomsten doorstroomonderzoek po school-loopbanen gegeven.* [First results of the school career transition survey]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/onderwijssectoren/primair-onderwijs/doorstroomonderzoek-schoolloopbanen#:~:text=Verschillende scholen vonden het positief,meer over dergelijke informatie beschikken.

Inspectie van het Onderwijs. (2023). *Kansrijk onderwijs in de overgang van po naar vo: Onderzoeksresultaten en aanbevelingen.* [Promising education at the transition from primary to secondary education]. Ministerie van Onderwijs, Cultuur en Wetenschap. https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/publicaties/2023/04/05/kansrijk-onderwijs-in-de-overgang-van-po-naar-vo/Kansrijk+onderwijs+in+de+overgang+van+po+naar+vo.pdf

Irizarry, Y. (2021). On track or derailed? Race, advanced math, and the transition to high school. *Socius, 7,* Article 2378023120980293. https://doi.org/10.1177/2378023120980293

Izzo, C. V., Weissberg, R. P., Kasprow, W. J., & Fendrich, M. (1999). A longitudinal assessment of teacher perceptions of parent involvement in children's education and school performance. *American Journal of Community Psychology, 27*(6), 817–839. https://doi.org/10.1023/A:1022262625984

Jacob, M., & Tieben, N. (2009). Social selectivity of track mobility in secondary schools: A comparison of intra-secondary transitions in Germany and the Netherlands. *European Societies, 11*(5), 747–773. https://doi.org/10.1080/14616690802588066

Jacobs, B., & Wolbers, M. H. J. (2018). Inequality in top performance: An examination of cross-country variation in excellence gaps across different levels of parental socioeconomic status. *Educational Research and Evaluation, 24*(1–2), 68–87. https://doi.org/10.1080/13803611.2018.1520130

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenen-wiskunde voor groep 3 t/m 8.* (pp. 1–96). [Scientific justification of the mathematics test in grade 1-6]. Cito.

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review, 48*(3), 398–407. https://doi.org/10.2307/2095231

Jungbluth, P. (2003). *De ongelijke basisschool: Etniciteit, sociaal milieu, sekse, verborgen differentiatie, segregatie, onderwijskansen en schooleffectiviteit.* [The unequal primary school]. ITS, Radboud Universiteit. https://jungbluth.nl/paul/wp-content/uploads/2016/02/08-de-ongelijke-basisschool.pdf

Jürges, H., & Schneider, K. (2011). Why young boys stumble: Early tracking, age and gender bias in the German school system. *German Economic Review, 12*(4), 371–394. https://doi.org/10.1111/j.1468-0475.2011.00533.x

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3

Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction, 28,* 73–84. https://doi.org/10.1016/j.learninstruc.2013.06.001

Kamerstukken II, 35671, nr. 13. (2021). [Parliamentary Paper]. Tweede Kamer der Staten-Generaal. https://www.tweedekamer.nl/downloads/document?id=2021D27328

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*(1), 623–655. https://doi.org/10.1146/annurev.psych.55.090902.142009

Kerr, N. L., & Tindale, R. S. (2011). Group-based forecasting? A social psychological analysis. *International Journal of Forecasting, 27*(1), 14–40. https://doi.org/10.1016/j.ijforecast.2010.02.001

Klapproth, F., & Fischer, B. D. (2019). Preservice teachers' evaluations of students' achievement development in the context of school-track recommendations. *European Journal of Psychology of Education, 34*(4), 825–846. https://doi.org/10.1007/s10212-018-0405-x

Klapproth, F., Glock, S., Böhmer, M., Krolak-Schwerdt, S., & Martin, R. (2012). School placement decisions in Luxembourg: Do teachers meet the Education Ministry's standards? *The Literacy Information and Computer Education Journal, 1*(1), 765–771. https://doi.org/10.20533/licej.2040.2589.2012.0113

Klapproth, F., Glock, S., Krolak-Schwerdt, S., Martin, R., & Böhmer, M. (2013). Prädiktoren der sekundarschulempfehlung in Luxemburg: Ergebnisse einer large-scale-untersuchung [Predictors of recommendations for secondary school type in Luxembourg]. *Zeitschrift Fur Erziehungswissenschaft, 16*(2), 355–379. https://doi.org/10.1007/s11618-013-0340-1

Klapproth, F., Kärchner, H., & Glock, S. (2018). Do students' religion and school absences moderate the effect of ethnic stereotypes on school-placement recommendations? *Journal of Experimental Education, 86*(2), 173–194. https://doi.org/10.1080/00220973.2017.1293602

Kleinjan, M., Pieper, I., Stevens, G., van de Klundert, N., Rombouts, M., Boer, M., & Lammers, J. (2020). *Geluk onder druk? Onderzoek naar het mentaal welbevinden van jongeren in Nederland.* [Research on the mental well-being of young people in the Netherlands]. UNICEF Nederland. https://www.trimbos.nl/wp-content/uploads/sites/31/2021/09/af1785-geluk-onder-druk.pdf

Knoester, M., & Au, W. (2017). Standardized testing and school segregation: Like tinder for fire? *Race Ethnicity and Education, 20*(1), 1–14. https://doi.org/10.1080/13613324.2015.1121474

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices.* Springer-Verlag.

Koomen, H. M. Y., Verschueren, K., & Pianta, R. C. (2007). *Leerling-Leerkracht Relatie Vragenlijst: Handleiding.* [Manual Student-Teacher Relationship Questionnaire]. Bohn Stafleu van Loghum.

Korpershoek, H., Beijer, C., Spithoff, M., Naaijer, H. M., Timmermans, A. C., van Rooijen, M., Vugteveen, J., & Opdenakker, M.-C. (2016). *Overgangen en aansluitingen in het onderwijs. Deelrapportage 1: Reviewstudie naar de po-vo en de vmbo-mbo overgang.* [Transitions and interconnections in education]. GION Onderwijs/Onderzoek. https://www.nro.nl/sites/nro/files/migrate/Eindrapport-405-14-402-project-1-Reviewstudie-naar-de-po-vo-en-de-vmbo-mbo-overgang.pdf

Korthals, R. A., & Dronkers, J. (2016). Selection on performance and tracking. *Applied Economics, 48*(30), 2836–2851. https://doi.org/10.1080/00036846.2015.1130789

Kpolovie, J. P., Joe, A. I., & Okoto, T. (2014). Academic achievement prediction: Role of interest in learning and attitude towards school. *International Journal of Humanities Social Sciences and Education, 1*(11), 73–100. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b4816f3f0c344184ec9f75dd7a6aef039a2bb80f

Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: A social cognitive analysis. *Social Psychology of Education, 16*(2), 215–239. https://doi.org/10.1007/s11218-013-9215-9

Krolak-Schwerdt, S., Hörstermann, T., Glock, S., & Böhmer, M. (2017). Teachers' assessments of students' achievements: The ecological validity of studies using case vignettes. *The Journal of Experimental Education, 86*(4), 515–529. https://doi.org/10.1080/00220973.2017.1370686

Kuhfeld, M., & Soland, J. (2021). The learning curve: Revisiting the assumption of linear growth during the school year. *Journal of Research on Educational Effectiveness, 14*(1), 143–171. https://doi.org/10.1080/19345747.2020.1839990

Kuyper, H., Zijsling, D., & Keuning, J. (2014). *Cohortonderzoek COOL5-18: Basisrapport voortgezet onderwijs, meting 2011.* [Report COOL5-18 data collection secondary education 2011]. GION onderwijs/onderzoek, Rijksuniversiteit Groningen.

Le Métais, J. (2003). Transition from primary to secondary education in selected countries of the INCA website. In *INCA Thematic Study (Vol. 9). Northern Ireland Council for the Curriculum, Examinations and Assessment.* https://www.oecd-ilibrary.org/docserver/9789264257658-en.pdf?expires=1695906386&id=id&accname=oid038041&checksum=6BD582D5E5BCD-6FBD2754D42272FFDE1

Leahey, E., & Guo, G. (2001). Gender differences in mathematical trajectories. *Social Forces, 80*(2), 713–732. https://doi.org/10.1353/sof.2001.0102

Ledoux, G., Roeleveld, J., Driessen, G., Cuppen, J., & Meijer, J. (2011). *Prestaties en loopbanen van doelgroepleerlingen in het onderwijsachterstandenbeleid: Stand van zaken en ontwikkeling in de periode 1994-2007.* [Performance and careers of target group pupils in educational disadvantage policy]. Kohnstamm Instituut, Universiteit van Amsterdam & ITS, Radboud Universiteit. https://hdl.handle.net/2066/210411

Lee, S.-Y., & Song, X.-Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences.* John Wiley & Sons.

Leidelmeijer, K., & Burema, F. (2022). *Leefbaarometer 2020 Verdiepende analyses.* [In-depth analyses of liveability meter 2020]. Atlas Research. https://leefbaarometer.nl/resources/ThemarapportLeefbaarometer2020VerdiependeAnalyses.pdf

Lek, K., & van de Schoot, R. (2019). Wie weet het beter, de docent of de Centrale eindtoets? [Who knows best: The teacher or the End of primary school test]. *De Psycholoog, 4,* 10–21. https://www.tijdschriftdepsycholoog.nl/wp-content/uploads/2019/04/Lek-K.-van-de-Schoot-R.-2019.-Wie-weet-het-beter-de-docent-of-de-centrale-eindtoets.-De-Psycholoog-4-10-21..pdf

Lenhard, W., & Schröppel, D. (2014). Prediction of academic performance prior to intersections within a multi-tiered school system. *Educational Research and Evaluation, 20*(6), 454–468. https://doi.org/10.1080/13803611.2014.975136

Leseman, P. P., Mulder, H., Verhagen, J., Broekhuizen, M., van Schaik, S., & Slot, P. (2017). Effectiveness of Dutch targeted preschool education policy for disadvantaged children: Evidence from the pre-COOL study. In H.-P. Blossfeld, N. Kulic, J. Skopek, & M. Triventi (Eds.), *Childcare, early education and social inequality: An international perspective* (pp. 173–193). Edward Elgar. https://doi.org/10.4337/9781786432094

LeTendre, G. K., Hofer, B. K., & Shimizu, H. (2003). What is tracking? Cultural expectations in the United States, Germany, and Japan. *American Educational Research Journal, 40*(1), 43–89. https://doi.org/10.3102/00028312040001043

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research, 37*(3), 888–902. https://doi.org/10.1016/j.ssresearch.2008.03.004

Lüdemann, E., & Schwerdt, G. (2013). Migration background and educational tracking: Is there a double disadvantage for second-generation immigrants? *Journal of Population Economics, 26*(2), 455–481. https://doi.org/10.1007/s00148-012-0414-z

Luijkx, R., & de Heus, M. (2008). The educational system of the Netherlands. In S. L. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97): An evaluation of content and criterion validity for 15 European countries* (pp. 47–75). Mannheimer Zentrum für Europäische Sozialforschung. https://www.mzes.uni-mannheim.de/publications/misc/isced_97/luij08_the_educational_system_of_the_netherlands.pdf

Luyten, H. (1998). School effectiveness and student achievement, consistent across subjects? Evidence from Dutch elementary and secondary education. *International Journal of Phytoremediation, 4*(4), 281–306. https://doi.org/10.1076/edre.4.4.281.6950

Luyten, H. (2004). Succes in het voortgezet onderwijs: Capaciteiten, inzet of achtergrond? [Success in secondary education: Capacities, commitment or background?]. *Pedagogische Studiën, 81*(2), 151–166. https://pedagogischestudien.nl/article/download/14625/16108

Luyten, H., & Bosker, R. J. (2004). Hoe meritocratisch zijn schooladviezen? [To what extent are track recommendations meritocratic?]. *Pedagogische Studiën, 81*(2), 89–103. https://pedagogischestudien.nl/article/view/14622/16105

Luyten, H., Schildkamp, K., & Folmer, E. (2009). Cognitive development in Dutch primary education, the impact of individual background and classroom composition. *Educational Research and Evaluation, 15*(3), 265–283. https://doi.org/10.1080/13803610902955333

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106. https://doi.org/10.1111/j.1750-8606.2008.00048.x

Magnuson, K., & Duncan, G. J. (2016). Can early childhood interventions decrease inequality of economic opportunity? *RSF: The Russell Sage Foundation Journal of the Social Sciences, 2*(2), 123–141. https://doi.org/10.7758/rsf.2016.2.2.05

Malecki, C. K., & Elliott, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17*(1), 1–23. https://doi.org/10.1521/scpq.17.1.1.19902

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*(4), 471–490. https://doi.org/10.1016/j.ecresq.2006.09.003

McKinney, J. D., Mason, J., Perkerson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology, 67*(2), 198–203. https://doi.org/10.1037/h0077012

McKown, C., & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology, 32*(1), 159–184. https://doi.org/10.1111/j.1559-1816.2002.tb01425.x

Meghir, C., & Palme, M. (2005). Educational reform, ability, and family background. *American Economic Review, 95*(1), 414–424. https://doi.org/10.1257/0002828053828671

Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics, 22*(3), 1142–1160. https://doi.org/10.1214/aos/1176325622

Mijs, J. J. B. (2016). The unfulfillable promise of meritocracy: Three lessons and their implications for justice in education. *Social Justice Research, 29*(1), 14–34. https://doi.org/10.1007/s11211-014-0228-0

Miller, V. J. (2018). Access denied: Tracking as a modern roadblock to equal educational opportunity. *New York University Law Review, 93*(4), 903–940. https://www.nyulawreview.org/wp-content/uploads/2018/10/NYULawReview-93-4-Miller.pdf

Milne, A., & Plourde, L. A. (2006). Factors of a low-SES household: What aids academic achievement? *Journal of Instructional Psychology, 33*(3), 183–193. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b4816f3f0c344184ec9f75dd7a6aef039a2bb80f

Ministerie van Onderwijs Cultuur en Wetenschap. (n.d.). *Hoe legt de basisschool de ontwikkeling van mijn kind vast?* [How does the primary school record my child's development?]. https://www.rijksoverheid.nl/onderwerpen/basisonderwijs/vraag-en-antwoord/hoe-legt-de-basisschool-de-prestaties-van-mijn-kind-vast

Ministerie van Onderwijs Cultuur en Wetenschap. (2014). *Toelating voortgezet onderwijs gebaseerd op definitief schooladvies.* [Admission secondary education is based on final school track recommendation]. [Admission secondary education is based on final school track recommendation]. https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-eindtoets-basisschool/toelating-voortgezet-onderwijs-gebaseerd-op-definitief-schooladvies

Ministerie van Onderwijs Cultuur en Wetenschap. (2019). *Gelijke kansen in het onderwijs.* [Equal educational opportunities]. https://www.ocwincijfers.nl/onderwijs/dashboard-gelijke-kansen

Ministerie van Onderwijs Cultuur en Wetenschap. (2023). *Schooladvies toelating voortgezet onderwijs.* [Recommendation admission secondary education]. https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-doorstroomtoets-basisschool/toelating-voortgezet-onderwijs-gebaseerd-op-definitief-schooladvies

Molenaar, I., & Knoop-van Campen, C. (2017). *Teacher dashboards in practice: Usage and impact.* [Conference Paper], Data driven approaches in digital education: 12th European conference on technology enhanced learning, Tallinn, Estonia. https://doi.org/10.1007/978-3-319-66610-5_10

Moon, U. J., & Hofferth, S. L. (2016). Parental involvement, child effort, and the development of immigrant boys' and girls' reading and mathematics skills: A latent difference score growth model. *Learning and Individual Differences, 47*, 136–144. https://doi.org/10.1016/j.lindif.2016.01.001

Motiejunaite-Schulmeister, Akvile Sicurella, A., & Birch, P. (2022). *The structure of the European education systems 2022/2023: Schematic diagrams.* European Commision, European Education and Culture Executive Agency & Eurydice. https://doi.org/10.2797/21002

Mulder, H., Verhagen, J., & Leseman, P. P. (2015). Metingen van executieve functies en taal in de voorschoolse periode. [Measurements of executive functions and language in the preschool period]. In A Veen & P. Leseman (Eds.), *Pre-COOL cohortonderzoek: Resultaten over de voorschoolse periode.* (pp. 31–46). Kohnstamm Instituut, Universiteit van Amsterdam. https://kohnstamminstituut.nl/wp-content/uploads/2019/08/932-Pre-COOL-cohortonderzoek.-Resultaten-over-de-voorschoolse-periode.pdf

Müller, C. M., & Hofmann, V. (2016). Does being assigned to a low school track negatively affect psychological adjustment? A longitudinal study in the first year of secondary school. *School Effectiveness and School Improvement, 27*(2), 95–115. https://doi.org/10.1080/09243453.2014.980277

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide.* (8th ed.). Muthén & Muthén. http://www.ncbi.nlm.nih.gov/pubmed/22210541

Naayer, H. M., Spithoff, M., Osinga, M., Klitzing, N., Korpershoek, H., & Opdenakker, M.-C. (2016). *De overgang van primair naar voortgezet onderwijs in internationaal perspectief: Een systematische overzichtsstudie van onderwijstransities in relatie tot kenmerken van verschillende Europese onderwijsstelsels.* (pp. 2–5). [The transition from primary to secondary education in international perspective]. GION onderwijs/onderzoek, Rijksuniversiteit Groningen. https://www.nro.nl/sites/nro/files/migrate/NRO_OPRO_De-overgang-van-po-naar-vo-in-internationaal-perspectief_Opdenakker_projectleider.pdf

Norman, G. (2010). Likert scales, levels of measurement and the 'laws' of statistics. *Advances in Health Sciences Education, 15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

O'Connell, A. A., & McCoach, D. B. (Eds.). (2008). *Multilevel modeling of educational data.* IAP.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). Yale University Press.

OECD. (2008). School and system characteristics and student performance in science [Table]. In *PISA 2006: Science competencies for tomorrow's world* (Vol. 1, pp. 213–281). OECD Publishing. https://doi.org/10.1787/142127877152

OECD. (2012). *Equity and quality in education: Supporting disadvantaged students and schools.* OECD Publishing. https://doi.org/10.1787/9789264130852-en

OECD. (2016a). *PISA 2015 results: Excellence and equity in education.* (Vol. 1). OECD Publishing. https://doi.org/10.1787/9789264266490-en

OECD. (2016b). *Reviews of national policies for education: Netherlands 2016: Foundations for the future.* OECD Publishing. https://doi.org/10.1787/9789264257658-en

OECD. (2017). *Starting strong V: Transitions from early childhood education and care to primary education.* OECD Publishing. https://doi.org/10.1787/9789264276253-en

OECD. (2020). Sorting and selecting students between and within schools. In *PISA 2018 results: Effective policies, successful schools* (Vol. 5, pp. 69–86). OECD Publishing. https://doi.org/10.1787/ca768d40-en

OECD. (2023). *PISA 2022 results: Factsheets Netherlands (Vol. 3).* OECD Publishing. https://www.oecd.org/publication/pisa-2022-results/country-notes/netherlands-0941b029#section-d1e214

Olivares, M. A., & Forero, G. C. (2010). Goodness-of-fit testing. *International Encyclopedia of Education, 7*(1), 190–196. https://doi.org/10.1016/B978-0-08-044894-7.01333-6

Onderwijsraad. (2019). *Doorgeschoten differentiatie in het onderwijsstelsel: Stand van educatief Nederland 2019.* [Excessive differentiation in the educational system]. https://www.onderwijsraad.nl/binaries/onderwijsraad/documenten/adviezen/2019/02/22/doorgeschoten-differentiatie-onderwijs/Stand-van-educatief-Nederland-2019.pdf

Onderwijsraad. (2021). *Later selecteren, beter differentiëren.* [Later selection, better differentiation]. https://www.onderwijsraad.nl/binaries/onderwijsraad/documenten/adviezen/2021/04/15/later-selecteren-beter-differentieren/Advies+Later+selecteren+beter+differentiëren.pdf

Oomens, M., Scholten, F., & Luyten, H. (2017). *Evaluatie wet eindtoetsing PO: Tussenrapportage.* [Interim evaluation report of the end or primary school test policy]. Oberon onderzoek/advies & Universiteit Twente. https://ris.utwente.nl/ws/portalfiles/portal/59266690/evaluatie_wet_eindtoetsing_po_tussenrapportage.pdf

Oomens, M., Scholten, F., & Luyten, H. (2019). *Evaluatie wet eindtoetsing po.* [Evaluation report of the end of primary school test policy]. Oberon onderzoek/advies & Universiteit Twente. https://www.oberon.eu/media/plzfih4q/evaluatie-wet-eindtoetsing-po.pdf

Pekkala Kerr, S., Pekkarinen, T., & Uusitalo, R. (2013). School tracking and development of cognitive skills. J*ournal of Labor Economics, 31*(3), 577–602. https://doi.org/10.1086/669493

Pendergast, L. L., Nickens, L., Pham, S., Miliaresis, S., & Canivez, G. L. (2018). Race and gender differences in teacher perceptions of student homework performance: A preliminary examination. *Contemporary School Psychology, 22*(3), 294–302. https://doi.org/10.1007/s40688-017-0162-x

Peterson, E. R., Rubie-Davies, C. M., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction, 42*, 123–140. https://doi.org/10.1016/j.learninstruc.2016.01.010

Peugh, J. L., & Feldon, D. F. (2020). "How well does your structural equation model fit your data?": Is marcoulides and Yuan's equivalence test the answer? *CBE Life Sciences Education, 19*(3), 1–8. https://doi.org/10.1187/cbe.20-01-0016

Pianta, R. C. (2001). *Student-Teacher Relationship Scale: Short form* (P. A. Resources (Ed.)). Psychological Assessment Resources. https://doi.org/10.1037/t11872-000

Pianta, R. C., Steinberg, M. S., & Rollins, K. B. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology, 7*(2), 295–312. https://doi.org/10.1017/S0954579400006519

Pietsch, M., & Stubbe, T. C. (2007). Inequality in the transition from primary to secondary school: School choices and educational disparities in Germany. *European Educational Research Journal, 6*(4), 424–445. https://doi.org/10.2304/eerj.2007.6.4.424

Pit-ten Cate, I. M., Krolak-Schwerdt, S., & Glock, S. (2016). Accuracy of teachers' tracking decisions: Short- and long-term effects of accountability. *European Journal of Psychology of Education, 31*(2), 225–243. https://doi.org/10.1007/s10212-015-0259-4

Plante, I., de la Sablonnière, R., Aronson, J. M., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology, 38*(3), 225–235. https://doi.org/10.1016/j.cedpsych.2013.03.004

Plenair debat 56-4: Omzeilen van het schooladvies. [Debate about bypassing track recommendations]. (2015). In *Dossier 31289: Voortgezet onderwijs.* [Parliamentary debate]. Tweede Kamer der Staten-Generaal. Identifier: h-tk-20142015-56-4. https://www.tweedekamer.nl/downloads/document?id=2018D12774

PO-raad, & VO-raad. (2014). *Handreiking schooladvies.* [School track recommendation report]. ITS, Radboud Universiteit Nijmegen. https://www.vanponaarvo.nl/wp-content/uploads/2014/10/Handreiking-schooladvies.pdf

Poncelet, P., & Metis Associates. (2004). Restructuring schools in Cleveland for the social, emotional, and intellectual development of early adolescents. *Journal of Education for Students Placed At Risk, 9*(2), 81–96. https://doi.org/10.1207/s15327671espr0902

POVO. (2013). *POVO-procedure Utrecht 2013-2014: Docentenhandleiding.* [Teacher manual for the transition from primary to secondary education in Utrecht 2013].

POVO. (2015). *Naar het voortgezet onderwijs: Handleiding bij de POVO procedure voor leerkrachten en docenten 2015-2016.* [Teacher manual for the transition from primary to secondary education in Utrecht 2015].

Praetorius, A. K., Koch, T., Scheunpflug, A., Zeinz, H., & Dresel, M. (2017). Identifying determinants of teachers' judgment (in)accuracy regarding students' school-related motivations using a Bayesian cross-classified multi-level model. *Learning and Instruction, 52*, 148–160. https://doi.org/10.1016/j.learninstruc.2017.06.003

Raftery, A. E. (1993). Bayesian model selection in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 163–180). Sage.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163. https://www.jstor.org/stable/271063

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods, 5*(2), 199–213. https://doi.org/10.1037/1082-989X.5.2.199

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48*(2), 335–360. https://doi.org/10.3102/0002831210374874

Reed, H. C., Van Wesel, F., Ouwehand, C., & Jolles, J. (2015). Keeping on track: Performance profiles of low performers in academic educational tracks. *British Educational Research Journal, 41*(1), 48–71. https://doi.org/10.1002/berj.3132

Riley, T., & Ungerleider, C. (2008). Preservice teachers' discriminatory judgments. *Alberta Journal of Educational Research, 54*(4), 378–387. https://doi.org/10.11575/ajer.v54i4.55244

Riley, T., & Ungerleider, C. (2012). Self-fulfilling prophecy: How teachers' attributions, expectations, and stereotypes influence the learning opportunities afforded aboriginal students. *Canadian Journal of Education, 35*(2), 303–333. https://www.jstor.org/stable/10.2307/canajeducrevucan.35.2.303

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education, 5*, Article 589965. https://doi.org/10.3389/feduc.2020.589965

Rockwood, N. J. (2017a). *Advancing the formulation and testing of multilevel mediation and moderated mediation models.* Doctoral dissertation. The Ohio State University.

Rockwood, N. J. (2017b). *MLmed user guide.* The Ohio State University. http://www.obhrm.net/0resource/MLmed_User_Guide_5_17.pdf

Rockwood, N. J., & Hayes, A. F. (2017). *MLmed: An SPSS macro for multilevel mediation and conditional process analysis*. The Ohio State University. https://static1.squarespace.com/static/58d3d231893fc0bdd12db130/t/5964f41fd1758e9283b0913e/1499788320211/APS_Poster_MLmed.pdf

Rodrigues, R. G., Meeuwisse, M., Notten, T., & Severiens, S. E. (2018). Preparing to transition to secondary education: Perceptions of Dutch pupils with migrant backgrounds. *Educational Research, 60*(2), 222–240. https://doi.org/10.1080/00131881.2018.1463144

Roeleveld, J., Driessen, G., Ledoux, G., Cuppen, J., & Meijer, J. (2011). *Doelgroepleerlingen in het basisonderwijs: Historische ontwikkeling en actuele situatie.* [At risk students in Dutch primary education]. Kohnstamm Intituut, Universiteit van Amsterdam. https://www.nro.nl/sites/nro/files/migrate/Prestaties-en-loopbanen-van-doelgroepleerlingen-in-het-onderwijsachterstandenbeleid.pdf

Roeleveld, J., Mulder, L., & Paas, T. (2011). *De gevolgen van een latere afname van de Cito Eindtoets Basisonderwijs.* [The consequences of taking the Cito End of Primary School test later]. ITS, Radboud Universiteit & Kohnstamm Instituut, Universiteit van Amsterdam. https://archief.kohnstamminstituut.nl/rapporten/ki11-1.pdf

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*(3), 726–748. https://doi.org/10.1037/0033-2909.92.3.726

Roorda, D. L., Jak, S., Zee, M., Oort, F. J., & Koomen, H. M. Y. (2017). Affective teacher-student relationships and students' engagement and achievement: A meta-analytic update and test of the mediating role of engagement. *School Psychology Review, 46*(3), 239–261. https://doi.org/10.17105/SPR-2017-0035.V46-3

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review, 3*(1), 16–20. https://doi.org/10.1007/BF02322211

Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology, 80*(1), 121–135. https://doi.org/10.1348/000709909X466334

Salaschek, M., Zeuch, N., & Souvignier, E. (2014). Mathematics growth trajectories in first grade: Cumulative vs. compensatory patterns and the role of number sense. *Learning and Individual Differences, 35*, 103–112. https://doi.org/10.1016/j.lindif.2014.06.009

Sax, L. (2016). *Boys adrift: The five factors driving the growing epidemic of unmotivated boys and underachieving young men* (2nd ed.). Basic Books.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Scheerens, J., Luyten, H., & van Ravens, J. (2011). *Perspectives on educational quality: Illustrative outcomes on primary and secondary schooling in the Netherlands.* Springer.

Scheerens, J., Timmermans, A. C., & van der Werf, M. P. C. G. (2019). Socioeconomic inequality and student outcomes in the Netherlands. In L. Volante, S. V. Schnepf, J. Jerrim, & D. A. Klinger (Eds.), *Socioeconomic inequality and student outcomes* (Education Policy & Social Inequality; Vol. 4). Springer. https://doi.org/10.1007/978-981-13-9863-6_7

Schijf, G. M. (2009). *Lees- en spellingvaardigheden van brugklassers.* [Reading and spelling skills of first-graders]. [Doctoral thesis, Universiteit van Amsterdam].

Schneider, S. L., & Tieben, N. (2011). A healthy sorting machine? Social inequality in the transition to upper secondary education in Germany. *Oxford Review of Education, 37*(2), 139–166. https://doi.org/10.1080/03054985.2011.559349

Schnepf, S. V. (2002). *A sorting hat that fails? The transition from primary to secondary school in Germany* (No. 92; Innocenti Working Papers). UNICEF. https://www.unicef-irc.org/publications/341-a-sorting-hat-that-fails-the-transition-from-primary-to-secondary-school-in-germany.html

Schumacker, R. E., & Lomax, R. G. (2010). *Structural equation modeling* (3rd Ed). Routledge.

Schütz, G., Ursprung, H. W., & Wössmann, L. (2005). *Education policy and equality of opportunity* (No. 1906; IZA Discussion Papers). https://docs.iza.org/dp1906.pdf

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://www.jstor.org/stable/2958889

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly, 30*(4), 894–906. https://doi.org/10.2307/748203

Sierens, S., Van Avermaet, P., Van Houtte, M., & Ağirdağ, O. (2020). Does pre-schooling contribute to equity in education? Participation in universal pre-school and fourth-grade academic achievement. *European Educational Research Journal, 19*(6), 564–586. https://doi.org/10.1177/1474904120925981

Simpkins, S. D., Tulagan, N., Lee, G., Ma, T. L., Zarrett, N., & Vandell, D. L. (2020). Children's developing work habits from middle childhood to early adolescence: Cascading effects for academic outcomes in adolescence and adulthood. *Developmental Psychology, 56*(12), 1–12. https://doi.org/10.1037/dev0001113

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Smeets, E., van Kuijk, J., & Driessen, G. (2014). *Handreiking bij het opstellen van het basisschooladvies.* [Report for formulating school track recommendations]. ITS, Radboud Universiteit Nijmegen. https://doi.org/10.13140/2.1.1391.3929

Smeets, E., & van Langen, A. (2022). *De overgang van primair naar voortgezet onderwijs.* [The transition from primary to secondary education]. Onderwijskennis.nl (NRO). https://www.onderwijskennis.nl/kennisbank/de-overgang-van-primair-naar-voortgezet-onderwijs

Sneyers, E., Vanhoof, J., & Mahieu, P. (2018). Primary teachers' perceptions that impact upon track recommendations regarding pupils' enrolment in secondary education: A path analysis. *Social Psychology of Education, 21*(5), 1153–1173. https://doi.org/10.1007/s11218-018-9458-6

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage. https://doi.org/10.1017/CBO9781107415324.004

Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. *Journal of Educational Psychology, 105*(2), 465–477. https://doi.org/10.1037/a0031754

Speybroeck, S., Kuppens, S., Van Damme, J., Van Petegem, P., Lamote, C., Boonen, T., & de Bilde, J. (2012). The role of teachers' expectations in the association between children's SES and performance in kindergarten: A moderated mediation analysis. *PLoS ONE, 7*(4), Article e34502. https://doi.org/10.1371/journal.pone.0034502

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology, 64*(4), 583–616. https://doi.org/10.1111/1467-9868.00353

Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence, personality and motivation? *Educational Research, 56*(2), 230–243. https://doi.org/10.1080/00131881.2014.898917

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Stiggins, B. R. J., & Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 1–10. https://doi.org/10.1177/003172170208301010

Stoel, R. D., & Galindo-Garre, F. (2011). Growth curve analysis using multilevel regression and structural equation modeling. In J. J. Hox & J. K. Roberts (Ed.), *Handbook of advanced multilevel analysis* (pp. 97–111). Routledge.

Strand, G. M. (2020). Supporting the transition to secondary school: The voices of lower secondary leaders and teachers. *Educational Research, 62*(2), 129–145. https://doi.org/10.1080/00131881.2020.1750305

Strand, S. (2011). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal, 37*(2), 197–229. https://doi.org/10.1080/01411920903540664

Strand, S. (2012). The White British–Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal, 38*(1), 75–101. https://doi.org/10.1080/01411926.2010.526702

Strand, S. (2013). What accounts for ethnic achievement gaps in secondary schools in England? *BERA Insights, 4*, 1–4. https://doi.org/10.1080/01411920903540664

Strello, A., Strietholt, R., Steinmann, I., & Siepmann, C. (2021). Early tracking and different types of inequalities in achievement: Difference-in-differences evidence from 20 years of large-scale assessments. Educational Assessment, *Evaluation and Accountability, 33*(1), 139–167. https://doi.org/10.1007/s11092-020-09346-4

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–762. https://doi.org/10.1037/a0027627

Sui-Chu, E. H., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education, 69*(2), 126–141. https://doi.org/10.2307/2112802

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education, 5*(4), 541–542. https://doi.org/10.4300/jgme-5-4-18

Swart, L., van den Berge, W., & Visser, D. (2019). *De waarde van eindtoetsen in het primair onderwijs.* [The value of school leavers' tests in primary education]. Centraal Planbureau. https://www.cpb.nl/sites/default/files/omnidownload/CPB-policy-brief-2019-03-de-waarde-van-eindtoetsen.pdf

Terrin, É., & Triventi, M. (2023). The effect of school tracking on student achievement and inequality: A meta-analysis. *Review of Educational Research, 93*(2), 236–274. https://doi.org/10.3102/00346543221100850

Themelis, S. (2008). Meritocracy through education and social mobility in post-war Britain: A critical examination. *British Journal of Sociology of Education, 29*(5), 427–438. https://doi.org/10.1080/01425690802263601

Tieben, N., & Wolbers, M. (2010). Success and failure in secondary education: Socio-economic background effects on secondary school outcome in the Netherlands, 1927-1998. *British Journal of Sociology of Education, 31*(3), 277–290. https://doi.org/10.1080/03054985.2011.559349

Timmerman, G., & van Essen, M. (2004). De mythe van het 'vrouwengevaar': Een historiserende inventarisatie van (inter)nationaal onderzoek naar de relatie tussen feminisering en 'jongensproblemen' in het onderwijs. [The relationship between feminisation and 'boy problems' in education]. *Pedagogiek, 24*(1), 57–71. https://hdl.handle.net/11370/9c7f0927-91c0-4b1c-a57c-9ff2d2f42fca

Timmermans, A. C., de Boer, H., Amsing, H. T. A., & van der Werf, M. P. C. G. (2018). Track recommendation bias: Gender, migration background and SES bias over a 20-year period in the Dutch context. *British Educational Research Journal, 44*(5), 847–874. https://doi.org/10.1002/berj.3470

Timmermans, A. C., de Boer, H., & van der Werf, M. P. C. G. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education, 19*(2), 1–24. https://doi.org/10.1007/s11218-015-9326-6

Timmermans, A. C., de Wolf, I. F., Bosker, R. J., & Doolaard, S. (2015). Risk-based educational accountability in Dutch primary education. *Educational Assessment, Evaluation and Accountability, 27*(4), 323–346. https://doi.org/10.1007/s11092-015-9212-y

Timmermans, A. C., Dijks, M. A., & Warrens, M. J. (2023). Are track recommendations dependent on schools and school boards? A study of trends in the level of track recommendations, number of double recommendations and reconsiderations in Dutch urban and rural areas. *British Educational Research Journal, 49*(3), 613–647. https://doi.org/10.1002/berj.3859

Timmermans, A. C., Kuyper, H., & Van der Werf, M. P. C. C. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology, 85*(4), 459–478. https://doi.org/10.1111/bjep.12087

Timmermans, A. C., Kuyper, H., & van der Werf, M. P. C. G. (2013). *Schooladviezen en onderwijsloopbanen: Voorkomen, risicofactoren en gevolgen van onder- en overadvisering.* [Track recommendations and educational careers]. GION onderwijs/onderzoek, Universiteit Groningen. https://avs.nl/sites/default/files/documenten/artikelen/add/Rapport-schooladviezen-en-onderwijsloopbanen.pdf

Timmermans, A. C., Naaijer, H., Keuning, J., & Zijsling, D. (2017). *Cohortonderzoek COOL5-18 Basisrapport meting VO-3 in 2014.* [Report COOL5-18 data collection 9th grade 2014]. GION Onderwijs/Onderzoek, Rijksuniversiteit Groningen. https://hdl.handle.net/11370/29c0f6e5-7b28-4cb2-b860-949bc942d6a7

Timmermans, A. C., van der Werf, M. P. C. G., & Rubie-Davies, C. M. (2019). The interpersonal character of teacher expectations: The perceived teacher-student relationship as an antecedent of teachers' track recommendations. *Journal of School Psychology, 73*, 114–130. https://doi.org/10.1016/j.jsp.2019.02.004

Timmermans, A. C., & Zijsling, D. (2014). *Cohortonderzoek COOL5-18: Technisch rapport bij het COOL5-18 bestandenproject voor het voortgezet onderwijs.* [Technical report COOL5-18 data collection secondary education]. GION ondewijs/onderzoek, Rijksuniversiteit Groningen. https://hdl.handle.net/11370/2f677d2d-a9bb-4471-927e-5982b3b83b26

Tolsma, J., & Wolbers, M. H. J. (2010). Onderwijs als nieuwe sociale scheidslijn? De gevolgen van onderwijsexpansie voor sociale mobiliteit, de waarde van diploma's en het relatieve belang van opleiding in Nederland. [Education as a new social dividing line? The impact of educational expansion. *Tijdschrift voor Sociologie, 31*(3), 239–259. https://doi.org/10.21825/sociologos.86759

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrik, 38*(1), 1–10. https://doi.org/10.1007/BF02291170

Uerz, D., Dekkers, H., & Béguin, A. A. (2004). Mathematics and language skills and the choice of science subjects in secondary education. *International Journal of Phytoremediation, 21*(1), 163–182. https://doi.org/10.1076/edre.10.2.163.27908

UNESCO. (1985). *Monographs in higher education: Higher education in the Netherlands* (W. Volmann (Ed.)). Ministerie van Onderwijs, Cultuur en Wetenschap & UNESCO European Centre for Higher Education. https://doi.org/10.1080/0260293820070210

van Aarsen, E., Roeleveld, J., & Luyten, J. W. (2013). *Voorspellende waarde van LOVS toetsen voor schoolsucces.* [Predictive value of LOVS tests for school success]. Oberon. https://ris.utwente.nl/ws/portalfiles/portal/5141857/Voorspellende_waarde_van_LOVS_toetsen_voor_schoolsucces-dig_versie.pdf

van Boxtel, H. W., Engelen, R., & de Wijs, A. (2011). *Wetenschappelijke verantwoording van de eindtoets basisonderwijs 2010.* [Scientific justification of the end of primary school test 2010]. Cito. https://www.beteronderwijsnederland.nl/wp-content/uploads/2013/04/cito_wetenschappelijke_verantwoording_eindtoets.pdf

van de Ven, B. (2007). The education systems of Europe. In W. Hörner, H. Döbert, B. Von Kopp, & W. Mitter (Eds.), *The education systems of Europe* (pp. 555–572). Springer.

van de Werfhorst, H. G. (2018). Early tracking and socioeconomic inequality in academic achievement: Studying reforms in nine countries. *Research in Social Stratification and Mobility, 58*, 22–32. https://doi.org/10.1016/j.rssm.2018.09.002

van de Werfhorst, H. G. (2021). Sorting or mixing? Multi-track and single-track schools and social inequalities in a differentiated educational system. *British Educational Research Journal, 47*(5), 1209–1236. https://doi.org/10.1002/berj.3722

van de Werfhorst, H. G., & Heath, A. (2019). Selectivity of migration and the educational disadvantages of second-generation immigrants in ten host societies. *European Journal of Population, 35*(2), 347–378. https://doi.org/10.1007/s10680-018-9484-2

van den Berg, E., Prins, H., van den Broek, A., Lommertzen, J., Hendrix, M., van der Wel, J., & Bijman, D. (2022). *Implementatiemonitor NP onderwijs: 14 cases in het funderend onderwijs.* [Implementation monitor national program of education]. Regioplan. https://open.overheid.nl/repository/ronl-e8b5aa7b4b3a5cce13713342cc581c20ae8a270b/1/pdf/implementatiemonitor-np-onderwijs-14-cases-in-het-funderend-onderwijs.pdf

van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal, 47*(2), 497–527. https://doi.org/10.3102/0002831209353594

van der Aa, M. A., de Vries, E., Hoekstra, H. J., Coebergh, J. W. W., & Siesling, S. (2011). Socio-demographic factors and incidence of melanoma in the Netherlands, 1994-2005. *European Journal of Cancer, 47*(7), 1056–1060. https://doi.org/10.1016/j.ejca.2010.11.020

van der Linde, A. (2004). On the association between a random parameter and an observable. *Sociedad de Estadística e Investigación Operativa, 13*(1), 85–111. https://doi.org/10.1007/BF02603002

van der Ploeg, S. W. (1993). *The expansion of secondary and tertiary education in the Netherlands.* [Doctoral thesis, Tilburg University]. https://pure.uvt.nl/ws/portalfiles/portal/1208786/3955496.pdf

van der Veen, I. (2003). Parents' education and their encouragement of successful secondary school students from ethnic minorities in the Netherlands. *Social Psychology of Education, 6*(3), 233–250. https://doi.org/10.1023/A:1024767230056

van Elk, R., van der Steeg, M., & Webbink, D. (2009). *The effect of early tracking on participation in higher education* (CPB Document No. 182). Centraal Planbureau. https://www.cpb.nl/sites/default/files/publicaties/download/effect-early-tracking-participation-higher-education.pdf

van Grinsven, V., & van Rossum, V. (2022). *Sociale veiligheid van leerkrachten po.* [Social safety of teachers in primary education]. DUO Onderwijsonderzoek & Advies. https://www.duo-onderwijsonderzoek.nl/wp-content/uploads/2022/10/Rapportage-Sociale-veiligheid-van-leerkrachten-PO-2022-7-oktober-2022.pdf

van Hattem, S., Aarts, M. J., Louwman, W. J., Neumann, H. A. M., Coebergh, J. W. W., Looman, C. W. N., Nijsten, T., & de Vries, E. (2009). Increase in basal cell carcinoma incidence steepest in individuals with high socioeconomic status: Results of a cancer registry study in the Netherlands. *British Journal of Dermatology, 161*(4), 840–845. https://doi.org/10.1111/j.1365-2133.2009.09222.x

van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & van Tartwijk, J. (2023). Creative problem solving in primary school students. *Learning and Instruction, 88*, Article e101823. https://doi.org/10.1016/j.learninstruc.2023.101823

Van Houtte, M. (2017). Gender differences in context: The impact of track position on study involvement in Flemish secondary education. *Sociology of Education, 90*(4), 275–295. https://doi.org/10.1177/0038040717731604

van Huizen, T. (2018). The evolution of achievement gaps from early childhood to adolescence in the Netherlands. In G. Passaretta & J. Skopek (Eds.), *Roots and development of achievement gaps: A longitudinal assessment in selected European countries* (ISOTIS Rep, pp. 50–87). Trinity College. http://www.isotis.org/wp-content/uploads/2019/01/ISOTIS_D1.3-Roots-and-Development-of-Achievement-Gaps.pdf

van Huizen, T. (2021). *Teacher bias or measurement error bias? Evidence from track recommendations* (Utrecht School of Economics Working Paper Series No. 21-31). https://www.uu.nl/sites/default/files/LEG_USE_WP_21-13.pdf

van Huizen, T., & Plantenga, J. (2018). Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments. *Economics of Education Review, 66*, 206–222. https://doi.org/10.1016/j.econedurev.2018.08.001

van Leest, A., Hornstra, L., van Tartwijk, J., & van de Pol, J. (2024). Teachers taking perceptions of student attributes into consideration when formulating track recommendations? *British Educational Research Journal, 50*(3), 1127–1171. https://doi.org/10.1002/berj.3957

van Leest, A., Hornstra, L., van Tartwijk, J., & Van de Pol, J. (2020). Test- or judgement-based school track recommendations: Equal opportunities for students with different socio-economic backgrounds? *British Journal of Educational Psychology, 91*(1), 193–216. https://doi.org/10.1111/bjep.12356

van Leeuwen, N. (2019). *Statistische gegevens per vierkant en postcode 2018-2017-2016-2015*. [Statistical data per postal code of years 2018-2017-2016-2015]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/-/media/cbs/dossiers/nederland-regionaal/postcode/statistische-gegevens-per-vierkant-en-postcode-peiljaar-2015-en-2016.pdf

van Look, K., Broekhuizen, M., Damhuis, C., Henrichs, L., Hornstra, L., de Jong, B., de Kleijn, R., van Leest, A., van Tartwijk, J., van Aarsen, E., Bomhof, M., de Wit, W., & Bisschop, P. (2018). In K. van Look & J. van Tartwijk (Eds.), *Onderwijskansen in Utrecht.* [Educational opportunities in Utrecht]. Universiteit Utrecht, Oberon Onderzoek en Advies & SEO Economisch Onderzoek. https://www.oberon.eu/media/23dpkqsv/eindrapport-gelijke- onderwijskansen-in-utrecht-2018.pdf

van Lutsenburg Maas, J. (1964). The 'Mammoth Law' reforms in Dutch education. *Comparative Education Review, 7*(3), 279–285.

van Nuland, H. J. C. (2011). *Eliciting classroom motivation: Not a piece of cake* [Doctoral dissertation, Universiteit Leiden]. https://hdl.handle.net/1887/16693

van Rooijen, M., Korpershoek, H., Vugteveen, J., & Opdenakker, M. C. (2017). *De overgang van het basis- naar het voortgezet onderwijs en de verdere schoolloopbaan.* [Transition from primary to secondary education and future educational career]. Pedagogische Studiën, 94(2), 110–134. https://pure.rug.nl/ws/portalfiles/portal/54246950/De_overgang_van_het_basis_naar_het_voortgezet_onderwijs_en_de_verdere_schoolloopbaan.pdf

van Rooijen, M., Korpershoek, H., Vugteveen, J., Timmermans, A. C., & Opdenakker, M.-C. (2016). *Overgangen en aansluitingen in het onderwijs. Deelrapportage 2: Empirische studie naar de cognitieve en niet-cognitieve ontwikkeling van de leerlingen rondom de po-vo overgang*. [Transitions and interconnections in education]. GION Onderwijs/Onderzoek. https://www.rug.nl/gmw/research/pedagogical-and-educational-sciences/overgangen-en-aansluitingen-in-het-onderwijs/rapport-deelstudie-2.pdf

van Vuuren, D., & van der Wiel, K. (2015). *Zittenblijven in het primair en voortgezet onderwijs: Een inventarisatie van de voor- en nadelen.* [Grade repetition in primary and secondary education]. (CPB Polcy Brief No. 2015/1). Centraal Planbureau. https://www.cpb.nl/sites/default/files/publicaties/download/cpb-policy-brief-2015-01-zittenblijven-het-primair-en-voortgezet-onderwijs.pdf

Vanlommel, K., & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal, 56*(3), 792–821. https://doi.org/10.1080/1369183X.2017.1409177

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2020). Sorting pupils into their next educational track: How strongly do teachers rely on data-based or intuitive processes when they make the transition decision? *Studies in Educational Evaluation, 69*, Article e100865. https://doi.org/10.1016/j.stueduc.2020.100865

Veen, A., & Leseman, P. P. (2015). *Pre-COOL cohortonderzoek: Resultaten over de voorschoolse periode.* [Report results PreCOOL data collection preschool]. https://kohnstamminstituut.nl/wp-content/uploads/2019/08/932-Pre-COOL-cohortonderzoek.-Resultaten-over-de-voorschoolse-periode.pdf

Veenstra, R. (1999). *Leerlingen – klassen – scholen: Prestaties en vorderingen van leerlingen in het voortgezet onderwijs.* [Achievement and progress of secondary school students]. [Doctoral Dissertation, Rijksuniversiteit Groningen]. https://hdl.handle.net/11370/0d553b26-d14c-4760-89bc-8b34e13ece82

Venkatakrishnan, H., & Wiliam, D. (2003). Tracking and mixed-ability grouping in secondary school mathematics classrooms: A case study. *British Educational Research Journal, 29*(2), 189–204. https://doi.org/10.1080/0141192032000060939

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology, 22*(3), 407–423. https://doi.org/10.1002/acp.1414

Visser, D., Lemmens, A., Magnée, C., & Dillingh, R. (2022). *Stapelen in het voortgezet onderwijs.* [Stacking in secondary education]. Centraal Planbureau. https://www.cpb.nl/sites/default/files/omnidownload/CPB-Notitie-Stapelen-in-het-voortgezet-onderwijs.pdf

Volman, M. (1999). Girl power en boys problems: Het onderwijssuccess van meisjes in de jaren negentig. [Educational success of girls in the 1990s]. *Tijdschrift voor Genderstudies, 2*(2), 49–59. https://hdl.handle.net/11245/1.163725

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation, 24*(3–5), 124–179. https://doi.org/10.1080/13803611.2018.1548798

Watkins, C. S., & Howard, M. O. (2015). Educational success among elementary school children from low socioeconomic status families: A systematic review of research assessing parenting factors. *Journal of Children and Poverty, 21*(1), 17–46. https://doi.org/10.1080/10796126.2015.1031728

Yaluma, C. B., & Tyner, A. (2021). Are U.S. schools closing the "gifted gap"? Analyzing elementary and middle schools' gifted participation and representation trends (2012–2016). *Journal of Advanced Academics, 32*(1), 28–53. https://doi.org/10.1177/1932202X20937633

Zhu, C., & Urhahne, D. (2020). Temporal stability of teachers' judgment accuracy of students' motivation, emotion, and achievement. *European Journal of Psychology of Education, 36*, 319–337. https://doi.org/10.1007/s10212-020-00480-7

Zijsling, D., Keuning, J., Keizer-Mittelhaëuser, M.-A., Naayer, H., & Timmermans, A. C. (2017). *Cohortonderzoek COOL5-18: Technisch rapport meting VO-3 in 2014.* [Technical report COOL5-18 data collection 9th grade in 2014]. https://research.rug.nl/files/41740853/Cool_afn1314_vo3_techrapport.pdf

Zijsling, D., Keuning, J., Naayer, H., & Kuyper, H. (2014). *Cohortonderzoek COOL 5-18: Technisch rapport meting vo-3 in 2011* (tweede versie). (pp. 1–104). [Technical report COOL5-18: data collection 9th grade in 2011]. GION Onderwijs/Onderzoek.

Zimmer, R. (2003). A new twist in the educational tracking debate. *Economics of Education Review, 22*(3), 307–315. https://doi.org/10.1016/S0272-7757(02)00055-9

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology, 34*(4), 390–400. https://doi.org/10.1037/h0078865

Zumbuehl, M., & Dillingh, R. (2020). *Ongelijkheid van het jonge kind.* (CPB Notitie). [Inequality of the young child]. Centraal Planbureau. https://www.cpb.nl/sites/default/files/omnidownload/CPB-Notitie-Ongelijkheid-van-het-jonge-kind.pdf

Appendix

Dutch literature examples *(in Dutch)*

Murat Isik – Wees Onzichtbaar (2017)

"Ik wist dat er veel zou afhangen van de vervloekte Cito-toets. We hadden het maandenlang nergens anders over gehad in de klas. Op momenten dat ik wakker lag van de spanning, probeerde ik mezelf moed in te praten door aan de woorden van meester Gregory te denken, die me meermaals openlijk had geprezen om mijn schrijf- en leesvaardigheid.

[…]

Ik wist dat een slecht resultaat bij de Cito-toets mijn ouders zou teleurstellen. Mijn vader zou bevestigd worden in het beeld dat hij van mij had: een weifelende jongen die in niets op hem leek en nog geen vleugje bezat van de bravoure en rebellie waarmee hij als tiener door het leven was gegaan. Als ik weleens iets zei wat hem onzinnig in de oren klonk, zei hij dat ik sprekend op de vader van mijn moeder leek. Het was niet bepaald als een loftuiting.

[…]

Dat stond er dus voor mij op het spel, samen met de angst om voor altijd in middelmatigheid te leven. Ik wist dat ik boven mezelf moest uitstijgen en vooral de schade bij het onderdeel rekenen moest beperken om niet door het ijs te zakken. Sinds de derde klas had ik een moeizame relatie met rekenen. Ondanks de bijlessen van meneer Rolf wist ik me nog steeds geen raad met staartdelingen en breuken.

[…]

Meester Ronald weet mijn onvermogen om de rekensommen op te lossen aan luiheid en desinteresse. Regelmatig liet hij me nablijven, net zo lang tot ik huilend willekeurige antwoorden invulde in mijn schrift en met een waarschuwing op zak naar huis werd gestuurd. Zo kwam het dat ik sindsdien al sidderde wanneer een overhoring voor rekenen werd aangekondigd. Daarom vreesde ik de Cito-toets. Ik wist dat ik op het onderdeel taal meer dan voortreffelijk moest scoren om de rampspoed die mij met rekenen wachtte niet fataal te laten worden voor mijn eindresultaat.

Op de eerste grote dag van mijn leven zat ik ontspannen aan het tafeltje waar ik al veel complimenten van meester Gregory had ontvangen. Ik concentreerde me hard toen ik me richtte op de rekenvragen, gokte hier en daar wanneer ik een antwoord niet wist. De

taalvragen kwamen me zo eenvoudig voor dat ik ervan overtuigd was dat ik iets over het hoofd zag. Daarom las ik sommige vragen wel drie keer, twijfelde ik even aan het antwoord dat ik had gegeven, maar ging ik daarna verder omdat de tijd begon te dringen.

Na afloop wist ik zeker dat ik de Cito-toets uitstekend had gemaakt en zag ik mijn ouders in gedachten al trots glimlachen. Maar in de dagen die volgden, brokkelde dat goede gevoel langzaam af toen ik me afvroeg of ik niet te veel had gegokt bij het onderdeel rekenen en of ik de taalvragen niet had onderschat.

Zo kwam het dat ik de ochtend van de uitslag wakker werd met een steek in mijn maag die zich langzaam verspreidde naar de rest van mijn lichaam. Het was misschien wel de belangrijkste dag van mijn leven tot dan toe, want vandaag zou duidelijk worden of ik in de voetsporen van mijn zus zou treden, vandaag zou ik te horen krijgen of ik net als zij naar het vwo mocht.

Meester Gregory deelde de Cito-uitslagen uit alsof hij informatiefolders verspreidde, nonchalant en zonder de formaliteit die daar voor mijn gevoel bij hoorde. Ik vroeg me af of dat misschien een aankondiging van naderend onheil was, maar misschien deed hij het bewust om de druk bij ons te verlagen.

Er werden al snel scores voorgelezen door mijn klasgenoten. '520,' riep Ishana als eerste. Orlando volgde: 'Shit, ik heb 517.' Niet veel later riep Esmeralda dat ze 533 punten had. Toen Saleem zijn briefje kreeg, keek hij er een tijdje in stilte naar. Net als een paar anderen weigerde hij zijn score hardop met de klas te delen. Fenuku begon heen en weer te wippen op zijn stoel nadat hij zijn briefje in ontvangst had genomen. Hij keek alsof hij zojuist te horen had gekregen dat hij terug moest naar Ghana.

Toen meester Gregory eindelijk aan mijn tafel stond, zei hij: 'Goed gedaan, Metin.' Hoewel ik voorvoelde dat ik rekenen had verprutst, kreeg ik weer vertrouwen door zijn woorden en hoopte ik heimelijk op een score van ver boven de 540. Maar wat ik had gevreesd, was uitkomen: rekenen had mij genekt. Ik had weliswaar uitmuntend gescoord op het onderdeel taal en hoorde met mijn score bij de drie besten van de klas, maar toch was ik teleurgesteld.

'Wat heb je gehaald?' vroeg Saleem.

'531,' antwoordde ik.

'Jezus, waarom kijk je dan zo teleurgesteld?' Saleem stompte me op mijn bovenarm. 'Wees blij, man.' Op gedempte toon voegde hij eraan toe: 'Mijn score is 520.'

Ik kon Saleem niet vertellen dat hij mijn referentiekader niet was. Ik had me al die tijd gericht op de score van mijn zus. Zij had 540 punten gehaald en ik zat daar ver onder en wist dat het nu nog maar de vraag was of ik naar het vwo kon.

Meester Gregory weigerde te vertellen op welk niveau hij ons zou laten instromen op de middelbare school. 'Dat ga ik op de ouderavond met jullie ouders bespreken, maar voor de meesten van jullie zal het geen verrassing zijn, gelet op jullie score.' Ik kon aan hem zien dat hij teleurgesteld was over de algehele uitslag in onze klas.

'Wat zijn we goed', zei Esmeralda toen we die middag weer samen naar huis liepen. Ik zei niets over mijn teleurstelling.

[…]

‘We hebben gepraat’, zei mijn vader. ‘Nou ja, in het begin heeft hij vooral gepraat.’ Hij nam een diepe trek van zijn sigaret en zei: ‘Je zei dat je leraar zo onder de indruk is van jou. Nou, hij wil je naar een mavo/havo-klas sturen.’

Dat moest mijn vader verkeerd begrepen hebben. Ik hoorde hem nog zeggen dat ik het goed had gedaan toen hij me de uitslagbrief had overhandigd. En ik behoorde, samen met Esmeralda, tot de drie besten van de klas met mijn score.

Mijn moeder kwam erbij staan. ‘Hij zei dat je het een kind niet te moeilijk moet maken en dat de havo ook veel mogelijkheden voor de toekomst biedt.’

Rekenen was me dus fataal geworden. Ik had me beter moeten voorbereiden, ik had meer moeten oefenen met meneer Rolf.

‘Maar we hebben het er niet bij laten zitten,’ zei mijn vader ineens. ‘We hebben benadrukt dat je op het vwo thuishoort.’

Ik vroeg me af hoe mijn ouders dat zo zeker wisten. Misschien was het wel te hoog gegrepen en zou wiskunde me op dat niveau de pas afsnijden, zoals meester Gregory vreesde.

‘En uiteindelijk is hij akkoord gegaan’, zei mijn moeder terwijl ze in haar handen klapte.

‘Ik heb hem zover gekregen: je mag naar het vwo,’ vulde mijn vader aan met een tevreden blik alsof ik het allemaal aan hem te danken had.

En hoewel een gevoel van opluchting zich meester van mij maakte, stak me dat mijn ouders zo hun best hadden moeten doen om meester Gregory te overtuigen. Ik dacht aan alle keren dat hij me in de klas had gecomplimenteerd met mijn prestaties. Maar op het beslissende moment had hij aan me getwijfeld. Op het moment dat de verdere koers van mijn leven zou bepalen, had hij me laten vallen.

[…]

Net als mijn zus was ik voorbestemd om naar het vwo te gaan om daarna aan de universiteit te studeren. Mijn ouders verwachtten niets minder van ons, ieder op hun eigen specifieke manier. Mijn moeder gaf ons vertrouwen, moedigde ons aan en informeerde dagelijks of we ons huiswerk gemaakt hadden. Mijn vader op zijn beurt vond het niet meer dan vanzelfsprekend dat mijn zus en ik in dit welvarende land met al zijn mogelijkheden zouden uitblinken en een universitaire studie afronden. Wij boften volgens hem enorm, want we hadden kansen die hij nooit had gehad als zoon van een boerin een onbeduidend dorp

in Oost-Turkije. Hoewel hij dat niet hardop uitsprak, was het duidelijk dat hij vond dat je haastachterlijk moest zijn om niet een universitaire titel te behalen in dit Europese land. Zeker nu hij zijn bijdrage had geleverd: hij had strijd gevoerd met meester Gregory en hem ervan overtuigd zijn mavo/havo-advies te herzien.

Op mijn eerste schooldag brachten mijn ouders me met de auto.

[…]

In de aula bestudeerden we de klassenlijsten die op een groot bord waren opgehangen. Mijn naam stond niet tussen de havo/vwo-klassen van de eerste brugklas. Toen mijn blik afdwaalde naar de overige klassen, zag ik mijn naam staan bij klas 1F.

'Hier; riep ik. 'Kijk, hier staat mijn naam.'

Mijn vader keek hoofdschuddend naar het bord. 'Dat kan niet, dat is een mavo/havo-klas. Er is een fout gemaakt. Hij keek om zich heen en trok een passerende leraar aan zijn jasje. 'Hallo, er is een fout gemaakt,' ze hij naar het bord wijzend. De leraar verwees hem door naar de administratie waar een lange rij stond. Een gezette Surinaamse vrouw zat achter een tafeltje en beantwoordde met een lang gezicht de vragen van de ouders en leerlingen terwijl de telefoon constant rinkelde. Toen wij aan de beurt waren, zei mijn vader voor de derde keer dat er een fout was gemaakt.

'Wat voor fout?'

'Mijn zoon moet naar het vwo!'

De vrouw keek mij onderzoekend aan, haar hoofd daarbij op een vreemde manier schuin houdend, alsof ze wilde zeggen: 'Dat joch daar?' Ze zette een bril op, bladerde met zichtbare tegenzin door een map en schudde toen gedecideerd haar hoofd.

'Nee, hoor. Ze overhandigde mijn vader een blaadje. 'Als u het er niet mee eens bent, verzoek ik u naar de rector te gaan.' Ze keek naar de almaar groeiende rij achter ons. 'Ondertussen kunt u uw zoon naar klas1F sturen'

En terwijl we wegliepen, vroeg ik me af of meester Gregory zich na een lange vakantie soms had bedacht, en vond dat ik beter eerst rustig kon instromen op een lager niveau. Misschien kwam het door het avontuur dat me te wachten stond dat ik mijn schouders ophaalde: 'Dan ga ik toch naar klas1F?'

Mijn vader keek me aan alsof ik had gezegd dat ik vuilnisman wilde worden. 'Praat geen onzin, jongen. We hebben ons er niet al die jaren voor ingezet om je naar de mavo te laten gaan.'

Ik dwaalde door de gangen van het gebouw op zoek naar het juiste lokaal, terwijl mijn ouders verhaal gingen halen bij de rector.

[…]

'Ik heb een hele leuke klas,' zei ik terwijl ik een glimlach niet kon onderdrukken. 'Er zit een meisje in dat ik al kende'.

'Dat is jouw klas niet', riep mijn vader verontwaardigd. 'Die prutsers hebben een fout gemaakt, al ontkent de rector dat, alsof ik een leugenaar ben.'

'Je vader heeft gelijk,' sprak mijn moeder op kalmere toon. 'De rector heeft toegezegd dat hij meester Gregory gaat bellen. Als hij het ermee eens is, ga je alsnog naar een havo/vwo-klas en komt het allemaal goed.'

Mijn ouders hadden misschien gelijk. Maar waarom vond ik het dan niet erg om in een mavo/havo-klas geplaatst orden?

[…]

'Hij heeft tegen ons gelogen,' zei mijn vader hoofdschuddend. 'Gregory en ik hebben elkaar nog de hand geschud om de afspraak te bezegelen. Wat een waardeloze vent is die meester van jou.'

'Ja, hij is onbetrouwbaar gebleken,' vulde mijn moeder nu zelfs aan en hoewel zij andere woorden koos en op een mildere toon sprak, klonk het uit haar mond scherper, want zij had zich niet eerder zo over hem uitgelaten.

Ik vond het vervelend dat mijn ouders zo over meester Gregory spraken. Plotseling moest ik denken aan de woorden die hij in onze laatste weken op de Bijlmerhorst had uitgesproken. Hij had gezegd dat het niet goed was kinderen op de toppen van hun kunnen te laten opereren om ze maar op het hoogste niveau onderwijs te laten volgen. Een kind moest niet aan te veel druk worden blootgesteld, want anders kon het averechts werken. En bovendien was er volgens hem niets mis met de mavo, of havo. Niet ieder kind kon het niveau van het vwo aan.

De volgende dag bezochten mijn ouders de rector opnieuw. Ik ging deze keer mee. De rector bleek een indrukwekkende man te zijn die net zo goed sergeant bij de landmacht had kunnen zijn, met zijn kortgeschoren kop en militaristische snor als van een Latijns-Amerikaanse juntaleider. Hij vertelde dat hij meester Gregory nog niet had kunnen bereiken, maar zodra hij meer wist, zou hij mijn ouders daarvan op de hoogte stellen. 'In de tussentijd kan uw zoon gewoon les volgen in klas 1F.'

'Dat is niet nodig,' zei mijn vader. 'Straks gaat hij toch naar het vwo.'

'Dat is nog helemaal niet zeker, meneer," sprak de rector rustig. 'Daarom stel ik voor dat hij ondertussen toch naar 1F gaat.' Het klonk niet als een verzoek, maar als een dienstbevel. Misschien verzette mijn vader zich daarom.

'Mijn zoon gaat niet naar de mavo.'

'Meneer Mutlu, als u uw zoon niet naar school stuurt, ben ik verplicht dat te melden bij de leerplichtambtenaar.'

'Doe maar!' riep mijn vader, 'dan vertel ik hem dat er een fout is gemaakt.'

Even later in de Lada keek mijn vader me dreigend aan via de achteruitkijkspiegel. 'Je gaat niet naar een mavo-klas! Je gaat pas terug naar die school als ze je toelaten tot de klas waar je thuishoort.'

[…]

Pas een paar dagen later, op donderdagmiddag, rond een uur of vier, kregen we een telefoontje van de rector. Meester Gregory had er alsnog mee ingestemd dat ik naar een havo/vwo-klas ging. Mijn moeder juichte en mijn vader keek voldaan. Hij had het weer geregeld. Ik dacht alleen maar aan Nicole. Daarna probeerde ik me voor te stellen hoe het telefoontje was verlopen. Wat had meester Gregory gezegd? Hij had zijn twijfels over mij geuit toen de rector hem de situatie had voorgelegd, of was het daadwerkelijk een vergissing geweest? Misschien had hij me wel het voordeel van de twijfel gegeven toen hij aan de mooie zinnen dacht die ik in zijn klas had geschreven.

[…]

'Goed. Toen ik je leraar inlichtte over je situatie, ze hij dat hij er met jouw ouders uitgebreid over had gesproken.' Hij nam traag een slok van zijn koffie. 'Hij is er geen voorstander van de druk op leerlingen onnodig op te voeren, maar hij stemde er toch mee in dat jij alsnog naar een havo/vwo-klas gaat. Hij wenste je succes toe, en ik doe dat hierbij ook, want het is bepaald niet makkelijk om met jouw Cito-score naar een havo/vwo-klas te gaan. Het is weliswaar nog maar de eerste brugklas, maar het niveau en tempo liggen erg hoog, veel hoger dan in een mavo/havo-klas. Je zult dus goed je best moeten doen, Mettie. Maar

als dat niet lukt, is dat niet erg. Er is namelijk helemaal niets mis met de mavo of de havo. Begrijp je dat?'

[…]

'Wil je zelf ook graag naar het vwo?'

Hier moest ik natuurlijk volmondig 'ja' op antwoorden zonder een greintje twijfel uit te stralen. En snel ook. 'Ja, meneer."

Karin Amatmoekrim – Het gym (2011)

"De kinderen verzamelden handtekeningen van de leraren voor later, en de ouders hadden het over naar welke school hun kind zou gaan. Dat was makkelijk, want iedereen ging ofwel naar het Valeriuscollege ofwel naar de lts, die naast het Valerius stond en wat dus eigenlijk op hetzelfde neerkwam. Iemand vroeg: 'Is het waar dat Sandra naar het gennasium gaat?' Haar moeder knikte trots: 'Ze heeft er veel zin in.'

Dat was een leugen.

De moeder van Anouk, die net zo'n rotkop had als d'r dochter, zei: 'Weet je wel zeker of ze dat wel ken?', en de andere ouders keken bezorgd van Sandra naar haar moeder en weer terug.

'Natuurlijk kán ze dat,' antwoordde haar moeder, met de nadruk op 'kan'.

'Ik haat die school', zei Sandra tegen Tanya. Ze waren van het muurtje afgegleden en liepen in de richting van Tanya's flat.

'Jij hebt makkelijk praten. Jedereen gaat naat het Valerius.'

Zijn er op dat gennasium dan helemáál geen leuke lui?'

Ze schudde haar hoofd.

'Nou ja, als het niks is ga je er volgende jaar vanaf. Kom je gewoon bij ons in de klas.'

'M'n moeder ziet me aankomen,' antwoordde Sandra.

[…]

'Wat ik wil zeggen', zei haar moeder nu op een toon die Sandra niet van haar kenden, 'is dat het een hele dure school is. Je moet je best doen. Begrijp je dat? Je gaat me niet teleurstellen. Oké?'

Ze schudde haar hoofd.

'Als je een beter leven wilt, moet je daar zelf voor zorgen.'

'Ja, ma.'

'Niemand anders gaat dat voor je doen, hoor je?'

'Jaha.'

[…]

Achter lange tafels vol met keurig gestapelde boeken, hielpen oudere leerlingen en een handvol leraren met uitdelen, en van al die mensen keek zo ongeveer iedereen net iets te lang naar Sandra en haar moeder. Misschien wisten ze dat zij niet zoals de anderen op tijd het geld hadden overgemaakt. Zij kwamen als een stel arme sloebers contant betalen, en daarom werd er nu zo naar hen gekeken.

[…]

En jij bent zeker Sandra,' zei hij met een glimlach. 'We zijn heel, heel erg blij om jou hier op school te hebben."

Robert Vuijsje – Alleen maar nette mensen (2008)

"Het was bijna twee jaar geleden dat ik van het gymnasium kwam. Ik was nog niet begonnen aan een studie.

Sinds de kleuterschool wist ik dat het niet een vraag was of ik de universiteit zou afmaken. Het was een gegeven, een opdracht waar ik op geen enkele manier onderuit kon.

[…]

Op de lagere school begon ik er nachtmerries over te krijgen. Dat ik van school werd gestuurd omdat ik geen goeie cijfers haalde. Als ik wakker werd, kwam mijn moeder bij me op bed zitten. Mijn vader bleef beneden, of op zijn werkkamer.

Ik vroeg: 'Bestaat er ook een vuilnismannenschool?'

Mijn moeder zei dat dat niet bestond.

'Wat als ze me van school sturen?' vroeg ik. 'En ik moet naar de vuilnismannenschool?'

Mijn moeder zei dat het niet zou gebeuren.

[…]

Het was de eerste dag van de middelbare school. Daan zat naast me bij de eerste les, Latijn. Hij vroeg: 'Die Naomi, is dat familie van jou?' Ze leek op mij. Hij vond dat ze er lekker uitzag. Niet dat hij daarmee wilde zeggen dat ik er ook lekker uitzag.

Naomi droeg dezelfde soort kleren als ik, ze praatte op dezelfde Oud-Zuid-manier als ik, en ze had dezelfde donkere krulletjes als ik. Die van haar waren donkerbruin."

Mano Bouzamour – De belofte van Pisa (2013)

"Moeder Maria vol genade, ik ben aangenomen op het Hervormd Lyceum Zuid, godverdomme hé. Mijn broer en ik liepen vrolijk tussen rokende jongelui het bordes af van de school waar ik net een kennismakingsgesprek had gehad. Omdat ik van mijn zwakzinnige basisschooljuf een vmbo-advies had gekregen terwijl de Cito-uitslag mij tot het vwo verhief, moest ik met een van mijn ouders op gesprek komen. Aan de hand daarvan zou er worden gekeken of ik wel of niet geschikt was voor het lyceum. Zoals gewoonlijk bij schoolgesprekken ging mijn broer mee.

Ik werd door een docente Nederlands aan een kruisverhoor onderworpen. Ze bond me nog net niet vast aan de armleuningen van de stoel. Drie kwartier later rondde ze af met: 'Beloof je mij dat je heel hard je best zult doen als ik je aanneem op deze school?'

'Zonder twijfel, mevrouw.'

Ze keek naar mijn broer, die naast mij zat.

'U heeft dat ook gehoord? Mooi. Dan heb ik in ieder geval een getuige. Gaat u erop toezien dat hij dat zal doen?'

‘Als een gevangenisbewaarder, maakt u zich daar maar geen zorgen over. De komende zes jaar is-ie van mij.

[…]

Mijn broer vertelde toen Soesi weg was dat ik het me niet kon permitteren om het te verprutsen. Niemand van de familie had een vwo-diploma gehaald. Soesi ook niet.

‘Waarom heb jij het eigenlijk niet gehaald?’

Het ijs in zijn bakje was gesmolten, hij slurpte het op en keek naar de ABN AMRO aan de overkant waar klanten in en uit liepen.

‘Ik kon het makkelijk halen.’

[…]

‘Weet je wat het was, Sam, ik had geen begeleiding. Thuis vroeg niemand me naar huiswerk. Of ik iets niet snapte. Logisch, vader en moeder kunnen niet eens lezen en schrijven. Ze zijn zelf nooit naar school geweest. Maar goed, ik zat op het Montessori Lyceum, vijfde klas vwo. Geblokt alsof de duivel me op de hielen zat. Ik kan eigenlijk nog naar de avondschool om alsnog mijn diploma te halen, maar daar heb ik echt geen zin in. Zie je mij al braaf achter een schooltafel naar een lompe leraar luisteren? Ik zat dus in de een-na-laatste klas vwo en kon mij overdag thuis niet op mijn huiswerk concentreren, dus bleef ik de nachten op. Ik zat soms tot vier uur ’s nachts te leren. Tot ik helemaal lijp werd en dacht: waarvoor doe ik het eigenlijk? Het kan me allemaal wat. Al die leraren, al die puistige pubers, vader en moeder, de hele wereld kon me wat. Als ik er nu op terugkijk, is het zo stom. Al dat werk voor niks. Maar de grap is, op dat moment heb je jezelf nog niet door. Ik wist niet waar ik mee bezig was. Ik was toen zestien, ik ben nu vierentwintig. Waarom neem ik jou overal mee naartoe? Waarom denk je? Omdat niemand ene flikker om mij gaf en niemand mij ergens mee naar- toe nam. Wat ik probeer te zeggen, zonder de juiste begeleiding laat je de boel makkelijker vallen. Maar jij hoeft je niet druk te maken, ik heb je rug, broertje. We gaan het samen doen.’

[…]

‘Ik wil dat je me wat belooft, Sam.’

Aan zijn stem hoorde ik dat het menens was. Ik keek hem vragend aan. Hij legde zijn hand weer in mijn nek. Ik wachtte tot hij iets zei.

‘Beloof je me dat je verder zult gaan waar ik ben blijven haken? Dat je over een paar jaar het Hervormd Lyceum Zuid uit loopt en dat fucking vwo-diploma in je handen hebt?’

Hij stak vastberaden zijn hand uit, zijn armbandjes rinkelden als ketens heen en weer. Ik dacht even na terwijl ik naar de neonverlichting van de ijssalon blikte waar met rode sierletters PISA IJS op stond. Ik veegde met het servetje mijn plakkerige vingers schoon en bezegelde de belofte met een ferme handdruk.

Mijn broer keek naar de ijssalon, toen naar mij en zei: ‘De belofte van Pisa.”

Speech Loes Ypma, House of Parliament 2015 *(in Dutch)*

“Mevrouw Ypma (PvdA): Voorzitter. Voor 200.000 kinderen is een spannend en belangrijk moment aangebroken: de overstap van de basisschool naar de middelbare school. Gisteren was ik bij de Landelijke Ouderraad, waar veel telefoontjes van bezorgde ouders binnenkomen. Het is natuurlijk maar het topje van de ijsberg, maar bijvoorbeeld de moeder van Johan vertelde dat de middelbare school hem wil weigeren omdat zijn scores in groep 6 en 7 zo laag waren. In die tijd lag zijn vader echter op sterven. In groep 8 heeft hij zich hersteld en dat herstel was aanleiding voor het havoadvies van de leerkracht. Of neem het meisje dat een flinke taalachterstand had, maar deze heeft weggewerkt in de laatste twee jaar van de basisschool. Zij wordt echter afgerekend op toetsgegevens van groep 6 en 7. Dat maakt mij boos. Kinderen die opkrabbelen, laatbloeiers, de vechters, verdienen kansen.

De juffen en meesters van de basisschool geven een schooladvies omdat zij het kind in acht jaar goed hebben leren kennen. Ze maken gebruik van het leerlingvolgsysteem, van observatie en toetsgegevens voor alle vakken, maar ook van gegevens over werkhouding, samenwerken et cetera. Natuurlijk gaat er weleens iets mis bij dat schooladvies. Daarom hebben we ook geregeld dat leerlingen die hoger scoren op de eindtoets, de Cito-toets, dan het schooladvies, het voordeel van de twijfel krijgen. Zo is de Cito-toets een goede second opinion geworden om onderadvisering te ontdekken in het geval van kinderen die bijvoorbeeld vanwege hun achtergrond te laag zijn ingeschat door de leerkracht. Zo geven we kinderen maximale kansen en zorgen we ervoor dat onderwijs de motor van de emancipatie blijft.

Basisscholen zijn erbij gebaat om informatie te krijgen over de schoolloopbaan van oud-leerlingen, zodat ze met behulp daarvan hun schooladviezen kunnen verbeteren en op kwaliteit kunnen houden. Mijn eerste vraag aan de staatssecretaris is dan ook of hij bereid is om ervoor te zorgen dat deze informatie, die al wordt verzameld, ook naar de basisschool wordt gestuurd.

Het schooladvies geeft jongeren dus recht op toelating, maar er zijn ook middelbare scholen die hieraan maling hebben en extra toelatingseisen stellen. Juist door niet te kijken naar waar iemands wieg staat maar door kinderen een kans te geven op basis van hun progressie en het oordeel van de basisschool leidend te laten zijn, zorgen we ervoor dat kinderen kansen krijgen om door te groeien. Mijn tweede vraag aan de staatssecretaris is dan ook hoe hij mijn met brede steun aangenomen motie uitvoert waarin wordt gesteld dat scholen die aanvullende eisen stellen, moeten worden aangepakt. De Inspectie voor het Onderwijs moet deze scholen sanctioneren en beboeten. Immers, zij ontnemen met

name de laatbloeiers de kans om het beste uit zichzelf te halen. Welke sancties heeft de staatssecretaris inmiddels getroffen?

Zou het niet beter zijn om het voortgezet onderwijs pas gegevens te laten opvragen na de toelating, zo is mijn derde vraag. Ouders die klem zitten tussen de middelbare school die heel veel opvraagt en de basisschool die zegt dat dit helemaal niet mag meetellen voor de toelating, kennen de regels niet. Mijn vierde vraag is dan ook hoe de staatssecretaris hierover heeft gecommuniceerd tot nu toe. Is hij bereid om bij de inspectie een meldpunt in te richten waar ouders terechtkunnen als blijkt dat scholen drempels opwerpen en kinderen weigeren, ondanks het schooladvies? Een meldpunt met doorzettingsmacht. Een meldpunt dat direct ingrijpt en ook dwingend kan optreden als de toelatingsprocedure in de desbetreffende regio moet worden aangepast voor de hele regio.

Onder ons zijn er misschien ook wel wat laatbloeiers die gelukkig toch kansen hebben gekregen dankzij een gecombineerd advies, bijvoorbeeld havo/vwo. Ik vind het wenselijk dat basisscholen gecombineerde adviezen kunnen geven. Ik ben van mening dat we kinderen altijd de kans moeten bieden om het beste uit zichzelf te halen. Daarom moet bij een gecombineerd advies ook altijd het hoogste advies leidend zijn bij de toelating van het kind. Is de staatssecretaris dit met mij eens, zo is mijn vijfde vraag. Kan hij toezeggen, dit duidelijk te gaan communiceren aan basisscholen en ouders?”

Bron: Tweede Kamer (25 februari 2015). Omzeilen van het schooladvies. Kamerstuk 2018D12774 (TK56-4). https://www.tweedekamer.nl/downloads/document?id=2018D12774

Original Dutch & Translated Quotes

Chapter 1

"Er komt eigenlijk zoveel bij kijken voordat je zegt: dit is het advies dat bij dit kind past."

"There is actually so much to consider before you say: this is the track recommendation that is appropriate for this child."

"De harde gegevens gebruiken we om te bekijken of een kind in potentie iets zou kunnen, maar een schooladvies gaat ook altijd samen met zicht op de 'zachte' gegevens."

"We use the hard data to see whether a child has the potential, but a track recommendation is always accompanied by insight into the 'soft' data."

"We voelen ons wel tekort gedaan als er in het nieuws komt dat opleidingsniveau leidt tot lagere/hogere adviezen, omdat we naar zoveel meer kijken."

"We do feel shortchanged when the news reports that education level of parents leads to lower or higher track recommendations, because we look at so much more."

Chapter 2

"Vroeger werd er te strikt naar de Cito gekeken en dat is nu minder."

"In the past, the Cito test was looked at too strictly, but that is less the case now."

"Verder hebben ouders een mbo-opleiding afgerond: dat is niet relevant voor mij."

"Furthermore, parents have completed a secondary vocational education: that is not relevant to me."

"Begrijpend lezen en rekenen zijn de belangrijkste factoren en daarin volgen we ook de adviesprocedure die er is."

"Reading comprehension and arithmetic are the most important subjects to consider and we also follow the existing track recommendation procedure."

Chapter 3

"Mijn leerlingen zitten op taalgebied vaak op een lager niveau doordat ze bijvoorbeeld thuis geen Nederlands spreken of minder Nederlands. Hierdoor is het soms moeilijker in te schatten op welk niveau je de leerling moet plaatsen. Spelling lukt meestal met goed oefenen wel, maar vooral woordenschat en begrijpend lezen is een probleem."

"My students are often at a lower level in language because, for example, they do not speak Dutch at home or speak less Dutch. This sometimes makes it more difficult to estimate at which level you should place the student. Spelling can usually be done with good practice, but vocabulary and reading comprehension are particularly problematic."

Chapter 4

"We kijken dan ook naar de ontwikkeling in prestaties, vooral als we twijfels hebben."

"We also look at the development in achievement, especially if we have doubts."

"We kijken naar de Cito-gegevens 6, 7, en 8 en met name begrijpend lezen en rekenen, omdat de middelbare scholen daar ook vooral om vragen. Het kan zijn dat dat fluctueert en daarom vind ik het belangrijk om naar de lijn te kijken. Het kan zijn dat er een stijgende lijn in zit. Dan geeft mij dat hoop voor de toekomst. Zijn er veel pieken en dalen, dan kijken we naar de methodengebonden toetsen."

"We look at the Cito data 6, 7 and 8 and in particular reading comprehension and mathematics, because secondary schools also mainly ask for this. This may fluctuate, which is why I think it is important to look at the line. It may be that there is an upward trend. Then that gives me hope for the future. If there are many peaks and troughs, we look at the teaching method-related tests."

"Bij kinderen die veel groei laten zien, is het niet altijd gegarandeerd dat het doorzet in het vo. Bij een aantal wel, maar niet bij allemaal. We zijn daarom voorzichtig met adviseren. We krijgen rapporten terug van het vo en we zien dan vaak dat bij leerlingen bij wie we twijfelden en die we naar een hoger niveau hebben laten gaan, bijna altijd weer terug zakken naar een lager niveau."

"With children who show a lot of growth, it is not always guaranteed that it will continue in secondary education. With some, it is, but not with all. We are therefore cautious when formulating track recommendations. We receive reports from secondary education, and we often see that with students for whom we had doubts and who we allowed to go to a higher level, almost always drop back to a lower level."

“Een leerling was een laatbloeier: hij begon met een 4 voor begrijpend lezen in groep 5, toen naar 3 in groep 6, toen 2, naar 1. Hij doet nu vwo. Dan zie je die lijn die gaat omhoog: die gaat het gewoon redden.”

“One student was a late bloomer: he started with a 4 (low score) for reading comprehension in group 5, then to 3 in group 6, then 2, to 1 (high score). He is now doing pre-university education. Then you see that line going up: he is just going to make it.”

Chapter 5

“Ik zou moeten zien hoe ze aan haar opdrachten werkt. Hoe ze er mee bezig is. Je kunt aan haar werkhouding ook aflezen of iets veel te gemakkelijk is.”

“I would have to see how she works on her assignments. How she is doing. You can also tell from her work habits if something is too easy.”

“Als je een leerling hebt die heel gemotiveerd is en je hebt ouders die het ondersteunen dan is het een win-win situatie.”

“If you have a student who is very motivated and you have supportive parents, it is a win-win situation.”

“Soms denk ik ook weleens dat ik niet te erg naar werkhouding moet kijken, omdat sommige kinderen het onderwijs op de basisschool minder leuk lijken te vinden en dat ze misschien op een vo-school bijvoorbeeld in een paardenklas tot hun recht komen. Dat zie ik dan weleens bij bepaalde opdrachten die ik geef. Ik zie dan dat ze daar heel goed in zijn. Hun houding kan dan veranderen als ze naar een vo-school gaan die past bij hun interesses, omdat ze op die school een andere manier onderwijs krijgen. Soms is dat wel een lastige overweging en dan moet je het echt inschatten.”

“Sometimes I also think that I should not look too much at work habits, because some children do not seem to like primary school as much. They might thrive at secondary school, for example in a horse class. I sometimes see that with certain assignments I give. I then see that they are very good at that. Their attitude may then change if they go to a secondary school that suits their interests, because they are educated in a different way at that school. Sometimes that is a difficult consideration and then you really have to assess it.”

“We hebben wel een groep leerlingen die hoog scoren en daar weinig voor hoeven te doen. De werkhouding is dan niet heel goed. Wij hebben wel het beeld op school dat kinderen die ervoor moeten werken verder komen in het vo dan kinderen die nu niet zoveel hoeven te doen (achterover hangen en even snel snel). Bij deze laatste groep is het cognitief allemaal

goed, maar ze lijken toch sneller af te stromen omdat ze geen goede werkhouding hebben. Het dilemma is dan: cognitie zegt dit, werkhouding zegt dat. Wat gaan we daarmee doen? Welk advies geef je? Ik denk dat het uiteindelijke advies 50/50 is: de ene keer naar boven, de andere keer naar beneden."

"We do have a group of students who score high and do not have to do much to achieve it. Their work habits are not very good. We do have the impression at school that children who have to work for it get further in secondary education than children who do not have to do much now (hang back and just do everything quickly). In this last group, everything is fine cognitively, but they still seem to drop out faster because they do not have good work habits. The dilemma then is cognition says this, work habits says that. What are we going to do with that? What advice do you give? I think eventually the track recommendations is 50/50: sometimes higher, sometimes lower."

"De ene 2 is niet de andere 2. Hoe je een leerling beleeft in de klas: stukje algemene ontwikkeling, aanwezigheid, motivatie, interesse, samenwerken. Kijken naar de vaardigheden: iemand die in zichzelf is, niet durft, dat is een heel ander kind dan een die initiatief neemt, communicatief vaardig is, een grote algemene ontwikkeling heeft, en naar je toekomt om te zeggen dat hij dingen nog niet begrijpt, moeilijk vindt en vraagt of je extra oefenstof hebt: die wil ervoor gaan. Het is niet erg als een kind dat niet doet, maar het heeft wel invloed op het advies: het kan soms net het verschil zijn tussen havo en vwo."

"One 2 (test score) is not the other 2. How you experience a student in class: general development, presence, motivation, interest, collaboration. Looking at the skills: someone who is introverted, does not dare, that is a completely different child than one who takes initiative, has good communication skills, has a great general development, and comes to you to say that he does not understand things yet, finds it difficult and asks if you have extra practice material: he or she is willing to go for it. It is not a problem if a child does not do this, but it does influence the track recommendation: it can sometimes be the difference between havo and vwo."

Chapter 6

"Wij hadden vroeger wel de neiging om te hoge adviezen te geven: te hoog in te zetten. Bij twijfel denk ik: laat ze het maar proberen. Tot we bij een aantal leerlingen, waarbij we echt twijfelden welke kant het op zou gaan, zagen dat het niet goed ging, die hadden wat meer onvoldoendes op de rapporten in het vo. Ik denk dat het komt sinds wij de grootste invloed hebben, sinds ons advies bindend is. Voorheen was het nog even spannend, want dan konden vo-scholen ook nog naar de scores en het hele dossier kijken. Tot het moment van plaatsing konden leerlingen nog op niveau afgewezen worden. Dat kan nu niet meer."

“We used to have the tendency to give track recommendations that were too high. When in doubt, I think: let them try. Until we saw that things were not going well with a number of students, for whom we really had doubts about which way things would go, and they had a few more failing grades on their reports in secondary education. I think it is because we have the greatest influence, since our track recommendation is binding. Previously, it was still a bit exciting, because secondary schools could also look at the scores and the entire student file. Until the moment of placement, students could still be rejected. That is no longer possible.”

Anne van Leest

OP HET *Juiste* SPOOR?

De Determinanten van het Basisschooladvies Ontrafeld

Nederlandstalige Samenvatting

De Overgang van Basisonderwijs Naar Voortgezet Onderwijs

In hiërarchisch georganiseerde (tracked) onderwijsstelsels zitten leerlingen van alle niveaus in het basisonderwijs in principe bij elkaar, maar worden zij op basis van hun capaciteiten onderverdeeld in verschillende vormen of niveaus (*tracks*) van voortgezet onderwijs. Hierbij worden de theoretisch meer uitdagendere niveaus over het algemeen als 'hoger' aangeduid.[19] De vorm van voortgezet onderwijs die leerlingen vervolgens doorlopen is sterk bepalend voor hun toekomstige (school)loopbaan (Boone & Demanet, 2020; Glock et al., 2012; Korpershoek et al., 2016). Ook Nederland kent een dergelijk hiërarchisch georganiseerd systeem. Daarin is bovendien sprake van een sterke padafhankelijkheid. Dit houdt in dat wanneer leerlingen eenmaal toegewezen zijn aan een bepaald niveau in het voortgezet onderwijs, het moeilijk is om te wisselen naar een ander niveau (Driessen et al., 2005; Inspectie van het Onderwijs, 2014). Vooral 'opstromen' naar theoretisch uitdagendere niveaus blijkt in de praktijk lastig (Timmermans et al., 2013; van Rooijen et al., 2017).

De procedure voor het verwijzen van leerlingen naar een specifieke vorm van voortgezet onderwijs aan het einde van de basisschool is dus erg belangrijk voor de onderwijskansen van leerlingen. In Nederland heeft het basisschooladvies, dat aan het einde van het basisonderwijs door de leraar van groep 8 opgesteld wordt, een centrale plaats in die procedure. Gezien de bepalende rol van dit basisschooladvies is het belangrijk dat dit advies past bij de capaciteiten en de daarmee samenhangende verwachtingen over de toekomstige prestaties van een leerling in het voortgezet onderwijs (Tieben & Wolbers, 2010).

Al vanaf de jaren '60 is de overgang van basisonderwijs naar voortgezet onderwijs onderwerp van politieke discussies met als inzet het bevorderen van gelijke onderwijskansen. Beleid dat daaruit volgt, richt zich ten eerste op de verandering van de structuur van het onderwijsstelsel. Daarin passen concepten als de middenschool, de (verlengde) brugperiode in het voortgezet onderwijs en 10-14-scholen. Ten tweede staan de basisschooladviezen ter discussie. Daarbij gaat het dan bijvoorbeeld om de mate waarin deze gebaseerd zijn op gestandaardiseerde toetsresultaten of op leerkrachtverwachtingen ten aanzien van de capaciteiten van leerlingen. In het onderzoek wat in dit proefschrift is samengebracht, heeft het basisschooladvies een centrale plaats.

---

19 Woorden als 'hoger' en 'lager' kunnen in de context van onderwijs de connotatie hebben van meer of minder waardevol. In dit proefschrift is het gebruik van hoger en lager gekozen omwille van de leesbaarheid, maar het is niet de bedoeling om een oordeel te vellen over de waarde van verschillende onderwijstrajecten.

Toetsgebaseerde Basisschooladviezen

Basisschooladviezen die voornamelijk gebaseerd zijn op gestandaardiseerde toetsresultaten, ook wel 'toetsgebaseerde basisschooladviezen', impliceren een meritocratische perspectief op kansengelijkheid (OECD, 2016b): het streven is dat leerlingen met vergelijkbare capaciteiten (hun 'merits'), die blijken uit hun prestaties op objectieve toetsen, vergelijkbare basisschooladviezen krijgen. Om de beschikbaarheid van dergelijke gestandaardiseerde toetsen te organiseren werd in Nederland het Centraal Instituut voor Toetsontwikkeling (Cito) opgericht. Dit instituut ontwikkelde onder meer de Cito eindtoets (1968) en later het leerlingvolgsysteem (LVS; Cito, n.d.-a; Faasse et al., 1987; Luijkx & de Heus, 2008). De Cito eindtoets werd in 1968 ingevoerd. De eindtoets bood basisscholen een objectief criterium voor het vaststellen van het cognitieve niveau van leerlingen om te bepalen welke vorm van voortgezet onderwijs het meest 'geschikt'[20] zou zijn voor een leerling. De rol van achtergrondkenmerken van leerlingen, zoals geslacht, sociaaleconomische status (SES) en migratieachtergrond, zou dan worden geminimaliseerd, waarmee gelijke kansen op een succesvolle onderwijsloopbaan zouden worden bevorderd (de Rooy, 2018). Dit beleid heeft echter niet kunnen voorkomen dat er nog steeds grote verschillen zijn tussen de basisschooladviezen van leerlingen met verschillende achtergrondkenmerken (Chzhen et al., 2018; Inspectie van het Onderwijs, 2014; OECD, 2012), wat met name veroorzaakt wordt door de grote verschillen in prestaties van leerlingen met verschillende achtergronden (de Boer et al., 2010; Timmermans, Kuyper, et al., 2015; van Rooijen et al., 2016).

In eerder onderzoek naar de rol van prestaties in basisschooladviezen worden meestal de meest recente toetsscores van leerlingen meegenomen. Vaak worden gemiddelden van recente resultaten over schoolvakken heen of één eindtoetsscore gebruikt (de Boer et al., 2010; Glock et al., 2012; Luyten & Bosker, 2004). Echter, bij deze wijze van werken wordt uitgegaan van het feit dat prestaties naar één bepaald onderwijsniveau wijzen. Dat levert problemen op wanneer leerlingen uiteenlopende prestaties in meerdere afzonderlijke vakken (er kan dan inconsistentie in prestaties tussen vakken zijn) of prestatiegroei over meerdere jaren (snelle of langzame ontwikkeling wordt dan niet meegewogen) laten zien. Het is echter onbekend in hoeverre leraren deze typen prestatiefactoren meewegen bij het opstellen van het advies – wanneer ze daartoe de mogelijkheid hebben – en wat voor effect dit heeft op het basisschooladvies.

Leraargebaseerde Basisschooladviezen

Een andere reden waarom het gebruik van gestandaardiseerde toetsen voor het bepalen van het basisschooladvies wordt bekritiseerd is dat gestandaardiseerde toetsen niet alle

20 Met het meest 'geschikte' of 'passende' basisschooladvies wordt het advies bedoeld dat de meest optimale reflectie is van de (toekomstige) potentie van een leerling in het voortgezet onderwijs (Boone & Van Houtte, 2013; Klapproth et al., 2012; Korpershoek et al., 2016).

relevante capaciteiten van leerlingen omvatten die voorspellend zijn voor (toekomstig) schoolsucces (Geven et al., 2018; OECD, 2016b). Leraren zouden ook andere, meer 'subjectieve', leerlingkenmerken, waarvan zij denken dat ze voorspellend zijn voor schoolsucces in het voortgezet onderwijs, mee (moeten) kunnen wegen bij het opstellen van het basisschooladvies. Hierbij kan gedacht worden aan werkhouding of gedrag in de klas (Feron et al., 2013; Geven et al., 2018), maar ook aan creatief probleemoplossend vermogen, kritisch denken (van Hooijdonk et al., 2023) of ouderbetrokkenheid (Timmermans et al., 2016). Deze factoren kunnen bovenop prestaties worden meegenomen door leraren, naast het feit dat een aantal van deze factoren ook in prestaties doorwerken. Wanneer leraren de kans krijgen om zelf afwegingen te maken welke factoren ze meewegen in het advies en de mate waarin ze dat doen, is er sprake van 'leraargebaseerde basisschooladviezen'.

Verschillen Tussen Groepen Leerlingen

Er zijn ook aanwijzingen dat de verschillen in basisschooladviezen tussen leerlingen met verschillende achtergronden verklaard kunnen worden doordat leraren (onbewust) achtergrondkenmerken van leerlingen meewegen in de adviezen (bias), zoals geslacht, SES en migratieachtergrond (Boone & Van Houtte, 2013; Klapproth et al., 2013; Timmermans et al., 2018). Met name leerlingen met een lagere SES lijken aanhoudend kansen mis te lopen (Batruch et al., 2023; Pit-ten Cate et al., 2016).

Het is echter onduidelijk in welke mate de bias in basisschooladviezen op basis van achtergrondkenmerken van leerlingen zou verminderen als gestandaardiseerde toetsresultaten zwaarder zouden wegen in de basisschooladviezen. Een beleidsverandering ten aanzien van de basisschooladviezen in 2014 in Nederland heeft ons de unieke kans geboden om dit te onderzoeken. Voor deze beleidsverandering waren er toetsgebaseerde basisschooladviezen, waarbij leraren meer gestandaardiseerde toetsresultaten tot hun beschikking hadden dan na de beleidsverandering, toen er leraargebaseerde adviezen werden opgesteld (Ministerie van Onderwijs Cultuur en Wetenschap, 2014).

Bovendien is er, naast de unieke bijdrage van de verschillende factoren aan het basisschooladvies, mogelijk een complexe wisselwerking tussen eerdere prestaties, leerlingkenmerken en achtergrondkenmerken die nog niet onderzocht is in eerder onderzoek naar basisschooladviezen. Zo kunnen leraren hun perceptie van bepaalde leerlingkenmerken mogelijk zwaarder laten meewegen in hun adviezen voor bepaalde groepen leerlingen. Ter illustratie: als ouders van leerlingen met een lagere SES als minder betrokken worden ervaren (Bakker et al., 2007), en als leraren dit bovendien sterker meewegen in het advies bij deze leerlingen (Sneyers et al., 2018; Timmermans et al., 2016), dan zouden deze effecten samen kunnen leiden tot nog lagere basisschooladviezen voor deze leerlingen. Ditzelfde geldt bijvoorbeeld ook voor inconsistentie in prestaties. Het is echter onbekend in hoeverre deze wisselwerking tussen prestatie-, leerling- en achtergrondkenmerken invloed heeft op het basisschooladvies.

Verschillen Tussen Leraren

Tot slot heeft het meeste eerdere onderzoek naar basisschooladviezen de verschillen tussen leraren buiten beschouwing gelaten (voor een uitzondering zie Timmermans et al. (2016, 2019)). Er kunnen verschillen zijn tussen leraren in de mate of de richting (positief of negatief) waarin zij verschillende factoren meewegen. Zo baseren sommige leraren zich mogelijk sterker op 'subjectieve' leerlingkenmerken en andere sterker op prestaties. Ter illustratie: mogelijk geven sommige leraren een lager advies aan leerlingen van wie de ouders als minder betrokken gezien worden, omdat ze veronderstellen dat de ouders niet in staat zullen zijn om hun kind te helpen met het schoolwerk en dit negatief zal uitpakken voor de leerling in het voortgezet onderwijs. Andere leraren geven deze leerlingen mogelijk een hoger advies, omdat de leerlingen gedurende het basisonderwijs zelf in staat waren het specifieke onderwijsniveau te bereiken zonder hulp van hun ouders en ze mogelijk nog meer in hun mars hebben. In dit voorbeeld, waar de invloed van ouderbetrokkenheid op het basisschooladvies varieert van negatief tot positief, is het totale effect van ouderbetrokkenheid mogelijk niet zichtbaar in analyses als er niet gekeken wordt naar verschillen tussen leraren, omdat de tegengestelde effecten elkaar opheffen. Dit kan leiden tot een onderschatting van de voorspellende waarde van deze factoren voor het basisschooladvies.

Onderzoeksvragen

Al met al zou een beter begrip van de wisselwerking tussen de verschillende factoren die worden meegenomen in het basisschooladvies en de mogelijke verschillen tussen leraren hierin kunnen bijdragen aan het lopende debat over hoe het plaatsingsproces op basis van basisschooladviezen het beste kan worden vormgegeven. Echter, dit is slechts zeer beperkt onderzocht. Daarom luidt de eerste vraag die richting gaf aan het in dit proefschrift gepresenteerde onderzoek als volgt:

Welke aspecten van leerlingprestaties, leraarperceptie van 'subjectieve' leerlingkenmerken en achtergrondkenmerken van leerlingen nemen leraren in overweging bij het formuleren van basisschooladviezen en hoe voorspellen deze factoren samen de basisschooladviezen?

Omdat het basisschooladvies bedoeld is als een optimale weergave van de potentiële onderwijsprestaties van leerlingen in het voortgezet onderwijs (Boone & Van Houtte, 2013; Glock et al., 2012; Timmermans, Kuyper, et al., 2015), is het bovendien belangrijk om te weten welke factoren die worden meegenomen in het basisschooladvies voorspellend zijn voor de prestaties van leerlingen in het voortgezet onderwijs. Hoewel onderzoek wijst op een sterke correlatie tussen prestaties in het basisonderwijs en in het voortgezet onderwijs (de Boer et al., 2010; van Rooijen et al., 2016), is het niet duidelijk in hoeverre de andere factoren ook voorspellend zijn voor succes in het voortgezet onderwijs. Vandaar dat de tweede onderzoeksvraag van dit proefschrift luidt:

In hoeverre zijn aspecten van leerlingprestaties en leraarperceptie van 'subjectieve' leerlingkenmerken uit de basisschoolperiode en achtergrondkenmerken van leerlingen voorspellend voor de prestaties van leerlingen in het voortgezet onderwijs?

Om deze vragen te beantwoorden bevat dit proefschrift vier empirische studies die beschreven zijn in de Hoofdstukken 2 tot en met 5. Deze studies zijn gebaseerd op twee longitudinale datasets met gegevens over de onderwijsontwikkeling van Nederlandse leerlingen in het basisonderwijs (groep 5 tot en met 8) en voortgezet onderwijs (leerjaar 1 tot en met 3). In de Hoofdstukken 2, 3 en 4 is gebruikgemaakt van een dataset die informatie bevat over leerlingen in de stad Utrecht in vier opeenvolgende schooljaren (2012-2016), en in Hoofdstuk 5 is gebruikgemaakt van een nationale dataset verkregen uit de COOL[5-18] cohortstudie in twee schooljaren (2007-2008 en 2010-2011).

Onderzoeksresultaten

Factoren die Leraren Meenemen in het Basisschooladvies

De eerste onderzoekvraag richtte zich op de factoren die een rol spelen bij de totstandkoming van het basisschooladvies.

De Rol van Prestaties

In lijn met eerder onderzoek (Feron et al., 2013; Timmermans, Kuyper, et al., 2015) suggereren de bevindingen in de Hoofdstukken 2 tot en met 5 dat leraren hun basisschooladviezen voornamelijk baseerden op de meest recente prestaties van leerlingen die zij op dat moment tot hun beschikking hadden. Het kan daarbij gaan om de resultaten van de eindtoets of, wanneer deze resultaten niet beschikbaar zijn, de meest recente prestaties van leerlingen voor begrijpend lezen en rekenen-wiskunde uit een leerlingvolgsysteem. Over het algemeen kregen leerlingen die aan het eind van de basisschool beter presteerden op deze toetsen hogere basisschooladviezen. De meest recente prestaties van leerlingen verklaarden 75 tot 83% van de verschillen (variantie) in basisschooladviezen. De beschikbaarheid van een eindtoets versterkte de invloed van prestaties op het advies. Het creëren van gelijke onderwijskansen volgens het meritocratische ideaal lijkt hiermee voor een groot deel gerealiseerd te worden: leerlingen met vergelijkbare prestaties krijgen meestal vergelijkbare basisschooladviezen.

Hoewel de meeste leerlingen redelijk vergelijkbaar presteren op de verschillende vakken, zijn er ook leerlingen van wie de prestaties substantieel verschillen, bijvoorbeeld als zij beter presteren in begrijpend lezen dan in rekenen-wiskunde of andersom (Luyten, 1998; Timmermans, Kuyper, et al., 2015). De resultaten in Hoofdstuk 3 lieten zien dat dit voor ongeveer 20% van de leerlingen het geval was (>1 SD verschil). Het algehele effect

van inconsistentie in prestaties op de basisschooladviezen was beperkt, maar bij zeer grote inconsistenties (>2 *SD* verschil) bleken leraren wel voorzichtiger te adviseren met als resultaat dat de laagste prestatie een relatief groter gewicht kreeg. Deze bevindingen laten zien dat het geven van een passend basisschooladvies in het geval van inconsistenties lastig is, omdat prestaties dan niet naar één onderwijsniveau wijzen.

Prestatiegroei, oftewel hoe de ontwikkeling in prestaties gedurende het basisonderwijs verloopt, kan ook verschillen tussen leerlingen en kan relevant zijn om in overweging te nemen bij het opstellen van het basisschooladvies. Immers, wanneer leerlingen de basisschool zijn begonnen met een relatief grote achterstand in prestaties, maar ze die achterstand hebben ingelopen tegen het einde van het basisonderwijs, dan kan dat wijzen op meer ontwikkelpotentieel dan hun prestaties aan het einde van het basisonderwijs doen vermoeden. In tegenstelling tot eerder onderzoek (Caro, Lenkeit, et al., 2009; Klapproth & Fischer, 2019) wijzen de bevindingen in Hoofdstuk 4 erop dat leraren in hun basisschooladviezen alleen rekening hielden met de meest recente prestaties van leerlingen. Prestatiegroei speelde dus geen rol. Dit kan liggen aan de grote schommelingen in prestaties van leerlingen tijdens het basisonderwijs en aan de aanzienlijke variatie in groeipatronen tussen leerlingen. Prestatiegroei meenemen in het advies is dus, net als bij inconsistenties het geval is, veel lastiger dan het bepalen van het advies op basis van één enkele toetsscore of een gemiddelde van recente toetsscores.

De Rol van Andere Leerlingkenmerken

Deze bevindingen ten aanzien van prestaties lijken te bevestigen dat door het gebruik van gestandaardiseerde toetsscores bij het opstellen van het basisschooladvies de rol van achtergrondkenmerken van leerlingen beperkt wordt (Au, 2013; Faasse et al., 1987; Luijkx & de Heus, 2008). Echter, nog steeds is 17 tot 25% van de verschillen (variantie) in de basisschooladviezen 'onverklaard'. Dit suggereert dat er mogelijk andere factoren een rol spelen bij het opstellen van het basisschooladvies. Dit is een reden om nog preciezer te kijken naar welke kenmerken van leerlingen een rol spelen in het advies.

De Rol van Geslacht

Voor geslacht gaven de resultaten van twee van de drie studies in dit proefschrift (Hoofdstuk 3 en 4, waarin gebruikgemaakt werd van data uit Utrecht van de schooljaren 2014-2015 en 2015-2016) aan dat er geen verschillen in basisschooladviezen waren tussen jongens en meisjes bovenop prestaties. Jongens en meisjes met vergelijkbare prestaties, inconsistentie in prestaties of vergelijkbare prestatiegroei bleken dus vergelijkbare adviezen te krijgen. Echter, in Hoofdstuk 5 (waarin gebruikgemaakt werd van de nationale COOL$^{5\text{-}18}$-dataset van de schooljaren 2007-2008 en 2010-2011) waren er wel kleine verschillen in basisschooladviezen tussen jongens en meisjes. Bij gelijke prestaties kregen meisjes hogere adviezen dan jongens. Eerder onderzoek liet dezelfde wisselende resultaten met betrekking tot verschillen in adviezen ten aanzien van geslacht zien (Jürges & Schneider, 2011; Krolak-Schwerdt et al., 2017; Timmermans et al., 2018). Mogelijke verklaringen voor de wisselende

bevindingen kunnen de verschillen in gebruikte data zijn of het gegeven dat het verschil in basisschooladviezen tussen jongens en meisjes de afgelopen decennia is afgenomen doordat leraren zich bewuster zijn van de prestatieverschillen tussen jongens en meisjes en daarnaar gehandeld hebben (Timmermans et al., 2018).

De Rol van SES

Met betrekking tot de SES van leerlingen bleken, zoals verwacht op basis van eerder onderzoek (Batruch et al., 2023; Feron et al., 2016; Timmermans et al., 2018), basisschooladviezen lager te zijn voor leerlingen met een lagere SES (Hoofdstuk 2 tot en met 5). De verschillen in basisschooladviezen tussen leerlingen met een lagere en hogere SES werden voornamelijk verklaard door verschillen in prestaties van leerlingen: leerlingen met een lagere SES hadden lagere prestaties dan hun leeftijdsgenoten met een hogere SES, en zij kregen als gevolg daarvan lagere basisschooladviezen. Echter, ook bij gelijke prestaties bleken leerlingen met een lagere SES iets lagere schooladviezen te krijgen.

Verder suggereren de bevindingen in Hoofdstuk 2 dat de aard van de gestandaardiseerde toetsresultaten van leerlingen waartoe leraren toegang hadden niet zo belangrijk was voor de invloed van SES op het advies, zolang ze maar toegang hadden tot dergelijke resultaten. In dit hoofdstuk werden de situaties vergeleken (a) waarin leraren de beschikking hadden over de resultaten van een eindtoets en (b) waarin dat niet het geval was en leraren zich baseerden op toetsgegevens uit het leerlingvolgsysteem. De invloed van SES bovenop prestaties was in beide situaties ongeveer even groot (1%). Deze bevindingen komen niet overeen met de bevindingen van Luyten en Bosker (2004) die een versterkte invloed van SES vonden wanneer leraren eindtoetsresultaten niet meenamen in hun adviezen.

Daarnaast liet het onderzoek in Hoofdstuk 3 zien dat inconsistentie in prestaties een kleine rol speelde in verschillen in basisschooladviezen op basis van SES. Leraren leken iets voorzichtiger te adviseren voor leerlingen met inconsistentie in prestaties met een hogere dan met een lagere SES. Het zou kunnen zijn dat leraren bij twijfel leerlingen met een hogere SES niet het hoogste niveau adviseren, maar voorzichtig zijn en het op één na hoogste niveau adviseren. Bij leerlingen met een lagere SES zijn de adviezen sowieso al lager vanwege algeheel lagere prestaties, dus heeft inconsistentie in prestaties waarschijnlijk minder invloed op de hoogte van het advies. In het onderzoek dat is beschreven in Hoofdstuk 4 werden wel SES-gerelateerde verschillen in prestatiegroei gevonden, maar dit vertaalde zich niet door naar verschillen in basisschooladviezen, omdat prestatiegroei daarin niet meegenomen werd.

De Rol van Migratieachtergrond

In het onderzoek beschreven in Hoofdstuk 5 werden voor migratieachtergrond wisselende resultaten gevonden. Er was namelijk een verschil tussen de twee cohorten die zijn onderzocht. De data van de schooljaren 2007-2008 lieten bij gelijke prestaties hogere basisschooladviezen voor leerlingen met een migratieachtergrond dan voor leerlingen zonder migratieachtergrond zien. Dit effect van migratieachtergrond was echter niet aanwezig

in de data van de schooljaren 2010-2011. Dit komt overeen met de bevindingen van Timmermans et al. (2018) die tussen 1995 en 2014 een daling in de hoogte van adviezen voor leerlingen met een migratieachtergrond lieten zien. De adviezen van leerlingen met en zonder migratieachtergrond kwamen daardoor dichterbij elkaar te liggen. Mogelijke verklaringen hiervoor zouden kunnen zijn dat de positieve discriminatie ten aanzien van deze groep leerlingen is afgenomen of dat de SES en migratieachtergrond van leerlingen vaak overlappen: leerlingen met een migratieachtergrond hebben vaak een lagere SES (Luyten & Bosker, 2004; Van der Veen, 2003), waardoor er mogelijk geen uniek effect van migratieachtergrond is (Driessen, 2012).

Overige kenmerken

Naast achtergrondkenmerken nemen leraren vrijwel geen andere leerlingkenmerken, die in principe losstaan van hun achtergrondkenmerken, mee in hun basisschooladviezen, zo bleek uit de bevindingen in Hoofdstuk 5. Het gaat dan om kenmerken, waaronder werkhouding, gedrag, onderpresteren, populariteit, leraar-leerlingrelatie en ouderbetrokkenheid, zoals de leerkracht ze waarneemt in de klas. Samen verklaarden deze kenmerken minder dan 1% van de verschillen (variantie) in de basisschooladviezen. Deze bevindingen zijn in overeenstemming met resultaten uit eerder onderzoek (Timmermans et al., 2016). Dit betekent echter niet dat deze factoren er niet toe doen voor basisschooladviezen. In plaats daarvan lijkt het aannemelijk dat deze factoren gerelateerd zijn aan prestaties (Kpolovie et al., 2014; Malecki & Elliott, 2002), en als zodanig geen unieke voorspellende waarde hebben voor de basisschooladviezen van leerlingen wanneer er ook naar de prestatie van leerlingen wordt gekeken.

Cumulatieve Effecten

Tenslotte is het belangrijk om op te merken dat, hoewel de gevonden invloed van achtergrondkenmerken van leerlingen op basisschooladviezen over het algemeen vrij klein is, er ook cumulatieve effecten kunnen zijn waardoor factoren gezamenlijk een grotere invloed kunnen hebben op het basisschooladvies dan dat ze afzonderlijk hebben. Zo laat dit proefschrift al deels cumulatieve effecten zien, bijvoorbeeld van SES en ouderbetrokkenheid, maar zijn nog niet alle onderzochte factoren tegelijkertijd meegenomen. Zo zou het kunnen dat een leerling met een lagere SES een slechte werkhouding en inconsistentie in prestaties heeft, gezien dit vaker voorkomt bij leerlingen met een lagere SES. Deze factoren samen resulteren dan in een nog lager advies dan alleen op basis van lagere prestaties verwacht wordt.

Verschillen Tussen Leraren

Naast de verschillen tussen leerlingen met betrekking tot het basisschooladvies zijn in dit proefschrift ook de verschillen tussen leraren in kaart gebracht. Er waren kleine verschillen tussen leraren in de mate waarin zij in hun basisschooladviezen rekening hielden met

verschillen in prestaties (Hoofdstuk 3), hun perceptie van leerlingkenmerken (Hoofdstuk 5), en in het bijzonder de SES van leerlingen (Hoofdstuk 2). Dat betekent dat de kans dat twee leerlingen met vergelijkbare kenmerken verschillende basisschooladviezen kunnen krijgen afhankelijk van de leraar die ze hebben klein, maar niet afwezig is.

Wanneer bijvoorbeeld de leraarperceptie van ouderbetrokkenheid laag was, gaven sommige leraren een hoger basisschooladvies (misschien omdat de leerling zo ver gekomen is ondanks een gebrek aan ondersteuning van ouders), terwijl andere een lager advies gaven (misschien juist omdat ze verwachten dat ouders geen steun kunnen bieden in het voortgezet onderwijs, terwijl een leerling dat misschien wel nodig heeft). Hoewel er landelijke en lokale richtlijnen bestaan voor het opstellen van het basisschooladvies (Inspectie van het Onderwijs, 2014; Oomens et al., 2019; Smeets et al., 2014) zijn deze niet bindend en kunnen er verschillen tussen leraren en scholen ontstaan. Zelfs als basisscholen wel de richtlijnen hanteren, worden ze in de praktijk niet altijd toegepast. Met name in situaties waarin de prestaties van leerlingen niet duidelijk wijzen op één specifiek onderwijsniveau, blijken de richtlijnen niet voldoende houvast te bieden en ontstaan er verschillen tussen leraren.

De Voorspellende Waarde Voor Prestaties in het Voortgezet Onderwijs

De tweede onderzoeksvraag richtte zich op de voorspellende waarde van factoren die meegenomen (kunnen) worden in het basischooladvies voor schoolsucces van leerlingen in het voortgezet onderwijs. De bevindingen beschreven in de Hoofdstukken 4 en 5 suggereren dat de factoren die leraren in overweging namen bij het opstellen van hun basisschooladviezen in grote mate overeenkwamen met factoren die de prestaties van leerlingen in het voortgezet onderwijs voorspelden. Het gaat dan voornamelijk om de meest recente prestaties in rekenen-wiskunde en begrijpend lezen en/of eindtoetsscores van leerlingen in het basisonderwijs die voorspellend bleken te zijn voor de prestaties van leerlingen in het voortgezet onderwijs (40-55% verklaarde variantie). Dit is in overeenstemming met bevindingen uit eerder onderzoek (Feron et al., 2013; Timmermans, Kuyper, et al., 2015; van Rooijen et al., 2016). Uit dit proefschrift blijkt dat leerlingen met hogere toetsresultaten op het gebied van begrijpend lezen en rekenen-wiskunde aan het einde van de basisschool ook hogere prestaties op het gebied van begrijpend lezen en wiskunde in het voortgezet onderwijs hadden en een hoger niveau volgden, zelfs nadat er rekening was gehouden met verschillen in plaatsing in de brugklas.

De verwachting was dat prestatiegroei in het basisonderwijs ook voorspellend zou zijn voor schoolsucces in het voortgezet onderwijs, omdat prestatiegroei een indicatie kan zijn voor het potentieel van een leerling. De bevindingen van Hoofdstuk 4 toonden echter aan dat de prestatiegroei in het basisonderwijs geen aanvullende voorspellende waarde had voor de schoolprestaties van leerlingen in het voortgezet onderwijs boven op de toetsresultaten van leerlingen in groep 8 van de basisschool. Dit zou het gevolg kunnen zijn van

(vroege) selectie en differentiatie in het voortgezet onderwijs (Borghans et al., 2019; Dockx et al., 2019; Hanushek et al., 2006). Het lijkt aannemelijk dat leerlingen presteren op basis van het onderwijsniveau waarop ze zijn ingedeeld, zelfs als ze meer potentie hebben.

De resultaten in Hoofdstuk 5 lieten zien dat de leraarperceptie van leerlingkenmerken die niet gerelateerd is aan hun achtergrond of geslacht, niet of slechts minimaal voorspellend waren voor de prestaties van leerlingen in het voortgezet onderwijs wanneer rekening wordt gehouden met (eerdere) prestaties. Tenslotte toonden de bevindingen in Hoofdstuk 4 en 5 aan dat achtergrondkenmerken van leerlingen invloed hadden op hun schoolsucces in het voortgezet onderwijs. Zo presteerden meisjes en leerlingen met een hogere SES meestal, maar niet in alle situaties, beter dan jongens en leerlingen met een lagere SES.

Implicaties Voor Onderwijspraktijk en -beleid

De bevindingen van het in dit proefschrift beschreven onderzoek hebben mogelijke implicaties voor de onderwijspraktijk en het onderwijsbeleid gericht op de manier waarop basisschooladviezen worden opgesteld. Daarnaast suggereren de bevindingen van dit proefschrift dat onderwijsongelijkheid niet zozeer ontstaat bij de overgang tussen het basisonderwijs en voortgezet onderwijs, maar al eerder in de schoolloopbaan van leerlingen aanwezig is en bovendien versterkt lijkt te worden door de inrichting van het voortgezet onderwijs.

Bij de Overgang van Basis- Naar Voortgezet Onderwijs

Het Basisschooladvies Naar Boven Bijstellen

Ten eerste lieten de bevindingen in dit proefschrift zien dat leerlingen met een lagere SES ook lagere basisschooladviezen kregen (Hoofdstukken 2 tot en met 5). Na de beleidsverandering van 2014 werden 'bijstellingen' van basisschooladviezen ingevoerd: als de resultaten op de eindtoets wijzen op een hoger onderwijsniveau dan het initiële advies, kan het advies naar boven worden bijgesteld. Deze bijstellingen waren echter optioneel, en met name het advies voor leerlingen met een lagere SES werd vaak niet bijgesteld (Swart et al., 2019). De reden hiervan is dat met name ouders/verzorgers met een hogere SES mogelijk eerder vraagtekens zetten bij het basisschooladvies en ook meer druk uitoefenen op leraren om het advies aan te passen dan dat ouders van leerlingen met een lagere SES zouden doen (Batruch et al., 2023; Timmermans et al., 2018). Na de meest recente beleidsverandering van 2024 is het bijstellen van het advies de standaard en kan daar alleen van worden afgezien als daarvoor duidelijke en dwingende redenen zijn (Ministerie van Onderwijs Cultuur en Wetenschap, 2023). Dat lijkt vanuit het perspectief van kansengelijkheid voor leerlingen met een lagere SES een goede zaak.

Kansrijk Adviseren

Daarnaast laten bevindingen in dit proefschrift zien dat leraren voorzichtiger adviseerden wanneer prestaties niet duidelijk één onderwijsniveau aanduidden. Dat resulteerde in lagere basisschooladviezen dan op basis van gemiddelde prestaties verwacht zou mogen worden (Hoofdstuk 3). Met name bij leerlingen met een lagere SES speelde dit een rol. Kansrijk adviseren, oftewel het voordeel van de twijfel geven door hoog of hoger te adviseren, zou de verschillen in basisschooladviezen op basis van SES kunnen verkleinen (Inspectie van het Onderwijs, 2021a; Onderwijsraad, 2021). Het geven van kansrijke basisschooladviezen blijkt namelijk over het algemeen gunstig te zijn voor leerlingen, omdat hogere adviezen en de daaropvolgende plaatsing in de brugklas leiden tot hogere prestaties in het voortgezet onderwijs (Hebbink et al., 2022; Lenhard & Schröppel, 2014; Tolsma & Wolbers, 2010). Uit onderzoek blijkt dat leraren, in ieder geval het eerste jaar na de invoering van dit beleid, steeds vaker kansrijke basisschooladviezen formuleerden (Onderwijsinspectie, 2021b). Hierdoor wordt het aanbevolen dat beleidsmakers kansrijke basisschooladviezen blijven aanmoedigen en dat leraren deze blijven geven.

Kansrijk adviseren kan bovendien de verschillen tussen leraren, die ook in dit proefschrift beschreven zijn in de Hoofdstukken 2, 3 en 5, verkleinen: doordat leraren verschillend omgaan met de mate en richting (positief of negatief) waarin bepaalde factoren mee worden genomen in het basisschooladvies, kan kansrijk adviseren ervoor zorgen dat leraren allemáál het hogere advies geven wanneer ze twijfelen. Ook kan dit beleid de invloed van achtergrondkenmerken op het basisschooladvies beperken, bijvoorbeeld doordat leerlingen met inconsistentie in prestaties (wat vaker voorkomt bij leerlingen met een lagere SES, zoals de resultaten gepresenteerd in Hoofdstuk 3 lieten zien) daardoor hogere adviezen zouden krijgen.

Duidelijkere Richtlijnen Opstellen

Dat het soms moeilijk is voor leraren om basisschooladviezen op te stellen die het beste passen bij de (toekomstige) capaciteiten van leerlingen kwam ook naar voren uit de bevindingen beschreven in de Hoofdstukken 2, 3 en 5. Het bleek dat leerlingen met vergelijkbare prestaties, leerlingkenmerken of achtergronden bij verschillende leraren soms verschillende basisschooladviezen kregen. Het maakt voor de kansen van leerlingen dus soms uit wie hun leraar is of op welke school ze zitten. Om de verschillen in basisschooladviezen tussen leraren (en scholen) verder te verkleinen, zijn er duidelijkere richtlijnen voor het opstellen van basisschooladviezen nodig. Hoewel er landelijke en lokale richtlijnen en procedures bestaan (Inspectie van het Onderwijs, 2014; Oomens et al., 2019; Smeets et al., 2014), zijn basisscholen hier niet aan gebonden en mogen ze hun eigen richtlijnen opstellen. Bovendien lijkt een aanzienlijk aantal scholen (37%) geen duidelijk protocol te hebben (Inspectie van het Onderwijs, 2014). Zelfs als een school een protocol heeft, zijn de richtlijnen meestal niet duidelijk over hoe leraren verschillende factoren kunnen meewegen als ze twijfelen, bijvoorbeeld in het geval van inconsistentie in prestaties of wisselende prestatiegroei, maar

ook bij meer 'subjectieve' factoren. Het is dus belangrijk dat deze richtlijnen handvaten bieden voor dergelijke situaties en vervolgens daadwerkelijk worden geïmplementeerd en gebruikt.

Samen Basisschooladviezen Opstellen

De verschillen tussen leraren verkleinen kan ook bewerkstelligd worden wanneer onderwijsprofessionals (nog meer) samen basisschooladviezen opstellen. Uit onderzoek (Inspectie van het Onderwijs, 2014; Smeets et al., 2014) bleek al dat basisschooladviezen in de meerderheid van de basisscholen door meerdere onderwijsprofessionals samen worden geformuleerd. Samenwerking tussen scholen of schoolbesturen kan helpen om verschillen tussen leraren en scholen verder te verkleinen, aangezien zij het beleid en de procedures vaststellen voor het opstellen van basisschooladviezen (Timmermans et al., 2023). Scholen en schoolbesturen zouden hun basisschooladviezen kunnen bespreken en vergelijken om op die manier te kunnen vergelijken of ze soortgelijke adviezen geven. Daarnaast zou het ook nuttig kunnen zijn als leraren van basisscholen worden bijgestaan door een leraar van het voortgezet onderwijs bij het opstellen van advies, zoals dat in Luxemburg gebeurt (Klapproth et al., 2012). Op de meeste plaatsen is er al enige overdracht van leerlinginformatie tussen de leraar van de basisschool en het voortgezet onderwijs, de 'warme overdracht', maar dit is niet verplicht.

Voorafgaand aan de Overgang van Basis- Naar Voortgezet Onderwijs

De bevindingen van dit proefschrift geven aan dat basisschooladviezen zeker niet de belangrijkste oorzaak zijn voor ongelijke kansen in het onderwijs. De belangrijkste reden voor verschillen in basisschooladviezen tussen leerlingen van verschillende achtergronden is het feit dat er substantiële en hardnekkige prestatieverschillen zijn tussen groepen leerlingen, meestal nadelig uitpakkend voor onder andere leerlingen met een lagere SES of met een migratieachtergrond (Hoofdstukken 2 tot en met 5). Deze verschillen in prestaties tussen leerlingen met verschillende achtergronden zijn al zichtbaar aan het begin van het basisonderwijs (Magnuson & Duncan, 2016; OECD, 2017). Deze hardnekkige prestatieverschillen kunnen erop wijzen dat – met het oog op het versterken van kansengelijkheid – beleid zich niet zozeer zou moeten richten op het basisschooladvies, maar juist op het nog meer inzetten van interventies dan nu al het geval is om verschillen voorafgaand aan of tijdens de basisschoolperiode te verkleinen. Voorbeelden hiervan tijdens de basisschoolperiode zijn tutoring, extra lestijd, intensieve ondersteuning (Dietrichson et al., 2017) of het faciliteren of aanmoedigen van ouderbetrokkenheid bij onderwijsactiviteiten (Watkins & Howard, 2015). Dit gebeurt al veel, maar gezien de toenemende ongelijkheid (Inspectie van het Onderwijs, 2019, 2020) is het belangrijk dat er nog meer extra middelen uitgetrokken worden voor leerlingen uit kansarme milieus. Bovendien zouden interventies die zich richten op begrijpend lezen en taalvaardigheid nuttig zijn, aangezien leerlingen met een lagere SES doorgaans slechter presteren op deze gebieden (Hoofdstuk 3).

Na de Overgang van Basis- Naar Voortgezet Onderwijs

Hoewel het basisschooladvies er uiteindelijk voor zorgt dat de leerling op een bepaald niveau in het voortgezet onderwijs terechtkomt, zorgt de padafhankelijkheid van het systeem ervoor dat de toewijzing aan een niveau in zeer sterke mate bepalend is voor de gehele verdere onderwijsloopbaan van leerlingen. Uit het onderzoek in dit proefschrift is gebleken dat lager presterende leerlingen, waaronder leerlingen met een lagere SES, nadelen ondervonden van dit selectiemoment doordat ze lagere basisschooladviezen kregen (Hoofdstukken 2 tot en met 5). Een oplossing hiervoor is het uitstellen van het selectiemoment en het introduceren van een heterogene onderbouw in het voortgezet onderwijs. Dit is vooral gunstig voor leerlingen met lage prestatieniveaus of met een lagere SES (Brunello & Checchi, 2007; Hanushek et al., 2006; Hopwood et al., 2016). Het uitstellen van het selectiemoment zorgt ervoor dat deze leerlingen meer tijd hebben om hun volledige potentieel te ontwikkelen voordat ze worden ingedeeld in verschillende onderwijsniveaus (van de Werfhorst & Heath, 2019), en het kan ertoe leiden dat lager presterende leerlingen beter gaan presteren door de invloed van hun hoger presterende medeleerlingen (Borghans et al., 2012; Geven et al., 2021; Grift et al., 2010). In 2021 werd dit ook al aanbevolen door de Onderwijsraad (2021), die adviseerde om leerlingen pas na drie jaar voortgezet onderwijs in onderwijsniveaus in te delen.

Verder tonen de bevindingen van dit proefschrift aan dat een substantieel deel van de leerlingen aanzienlijke prestatieverschillen had tussen verschillende vakken (Hoofdstuk 3). Toch volgen leerlingen alle vakken op hetzelfde onderwijsniveau in het voortgezet onderwijs, wat zou kunnen betekenen dat ze moeite hebben om sommige vakken te halen en andere vakken op een te makkelijk niveau volgen. Door leerlingen te groeperen op basis van hun vaardigheden of niveau per vak zouden leerlingen lessen kunnen volgen op een niveau dat beter aansluit bij hun capaciteiten. In verschillende andere landen, bijvoorbeeld in de Verenigde Staten, wordt al gewerkt volgens dit systeem (Irizarry, 2021; Yaluma & Tyner, 2021).

Beperkingen van het Onderzoek en Suggesties Voor Vervolgonderzoek

De context van dit proefschrift was het Nederlandse onderwijs, en soms ging het specifiek om een stedelijke context. Daarom zijn resultaten van de studies van dit proefschrift mogelijk niet zonder meer generaliseerbaar naar Nederland als geheel en zeker niet naar andere landen waar basisschooladviezen (net) anders geformuleerd worden. Bovendien hebben we ons gericht op initiële basisschooladviezen die leraren formuleerden en niet op het definitieve basisschooladvies of de plaatsing in het voortgezet onderwijs. Deze initiële basisschooladviezen zijn betekenisvol voor het proces dat leraren doorlopen bij het formuleren van hun basisschooladvies. Door ons te richten op deze initiële adviezen konden we

bestuderen hoe leraren verschillende factoren wegen en konden we in Hoofdstuk 2 twee type basisschooladviezen met elkaar vergelijken. Bij de definitieve basisschooladviezen kunnen echter ook nog verschillende andere factoren een rol spelen, zoals druk vanuit ouders op leraren (Inspectie van het Onderwijs, 2014; van Grinsven & van Rossum, 2022).

Daarnaast is de relatie tussen het initiële basisschooladvies en plaatsing in de brugklas gecompliceerd, omdat het voortgezet onderwijs niet verplicht is om leerlingen te plaatsen op een niveau dat exact overeenkomt met het basisschooladvies (Inspectie van het Onderwijs, 2023). Leerlingen met hetzelfde basisschooladvies kunnen daardoor terechtkomen in heel verschillende type brugklassen, wat effect heeft op hun verdere schoolloopbaan vanwege de padafhankelijkheid van het systeem. Het zou interessant zijn om in vervolgonderzoek zowel het definitieve advies als plaatsing mee te nemen als er wordt gekeken naar de voorspellende waarde van basisschooladviezen op de schoolloopbaan van leerlingen.

De studies in dit proefschrift zijn verder allemaal gebaseerd op grote kwantitatieve databestanden. Hoewel dit belangrijke inzichten kan geven in patronen en groepsverschillen, geeft het geen inzicht in de besluitvormingsprocessen zelf, redenen achter bepaalde keuzes van leraren of hun (impliciete) houdingen en stereotypen ten opzichte van leerlingen met verschillende achtergronden. Kwalitatief onderzoek kan bijvoorbeeld inzicht verschaffen in de persoonlijke overwegingen van leraren bij het formuleren van basisschooladviezen.

Zitten we op het Juiste Spoor?

De bevindingen van dit proefschrift wijzen er op dat ongelijkheid in het onderwijs niet op de eerste plaats te herleiden is naar de basisschooladviezen en ook niet naar de overgang van het basisonderwijs naar het voortgezet onderwijs. De ongelijkheid is al aanwezig voorafgaand aan het basisonderwijs, vermindert niet tijdens het basisonderwijs, en wordt daarna versterkt door de manier waarop het voortgezet onderwijs in Nederland is georganiseerd. Door de vroege selectie en differentiatie, en de hiërarchische en rigide indeling biedt het Nederlandse onderwijssysteem weinig mogelijkheden om verschillen tussen leerlingen met verschillende achtergronden te verkleinen. Hoewel er de afgelopen decennia verschillende beleidsveranderingen zijn doorgevoerd, blijft de prestatiekloof tussen leerlingen met verschillende achtergronden hardnekkig aanwezig. De meeste beleidsveranderingen waren weinig structureel, zoals veranderingen in het tijdschema van de eindtoets of de mogelijkheid om basisschooladviezen bij te stellen naar aanleiding van eindtoetsresultaten. Er zijn echter meer substantiële interventies nodig om de onderwijsongelijkheid voor en na de overgang van het basisonderwijs naar het voortgezet onderwijs drastisch te verminderen.

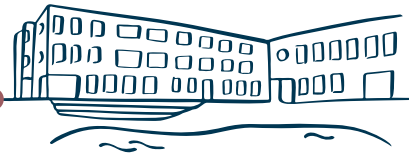
About the Author

Anne van Leest was born on September 26, 1986 in Utrecht, the Netherlands. In 2004, she completed secondary education at VeenLanden College in Mijdrecht. She studied Communication Sciences (2004) and General Linguistics (2005) at the University of Amsterdam. She completed both bachelor's degrees in 2008, and with some work experience in between, the master's degrees in 2016 and 2013 respectively. After her first master's degree, she enrolled in the traineeship Eerst de Klas tranche V in 2013, in which she combined education (obtaining a teacher degree at Utrecht University and working as a secondary school teacher) with a leadership programme and internship in sustainability. During the small research project at the end of her teacher degree, she developed her love for scientific educational research which led to becoming a PromoDoc in 2015. As a PromoDoc she worked 3 days a week as a teacher of Dutch language and culture in secondary education and 2 days a week on her PhD project at Utrecht University. During these years she worked in three different secondary schools (VeenLanden College in Mijdrecht, Minkema College in Woerden and Erfgooiers College in Huizen) and the university of applied science Aeres in Almere. As a tutor, member of the participation council (MR), leader of the debate club and De Inktaap (literary club), member of the national formative learning network of SLO, developer of reading comprehension test materials for Cito, Lesson Study teacher, website and social media administrator of the school, and entrance coordinator, she has developed in different roles. As a researcher, her PhD project resulted in a dissertation about equal educational opportunities at the transition from primary to secondary education in the the Dutch educational system, with a focus on teachers' track recommendations. In September 2022, she started working as an assistant professor at Utrecht University, being a teacher educator, and continuiing her research on equal educational opportunities, track recommendations, students' school careers and determination processes in (lower) secondary education.

1986
Born in Utrecht on
26 September 1986

1990-1998
Primary school (basisschool):
De Wegwijzer in Vianen
& St. Jozefschool in Vinkeveen

1998-2004
Secondary school (middelbare school):
VeenLanden College in Mijdrecht



2004-2008
University
Bachelor Communicatie Science,
University of Amsterdam

2005-2008
Bachelor General Linguistics,
University of Amsterdam

2008-2010 & 2013
Master General Linguistics,
University of Amsterdam

2009-2011 & 2016
Master Communication Science,
University of Amsterdam

2013-2015
Trainee Eerste de Klas
tranche V

2013-2015
Teacher degree
Utrecht University

2013-2015
Teacher at VeenLanden
College in Mijdrecht

2015-2024
PromoDoc

2016
Teacher at Minkema
College in Woerden

2016-2020
Teacher at Erfgooiers
College in Huizen

2020-2022
Lecturer at Aeres University of Applied
Sciences in Almere

2022-now
Assistant professor at
Utrecht University

List of Publications and Professional Contributions

Peer Reviewed Publications

van Leest, A. M. C., Hornstra, T. E., van Tartwijk, J. W. F., & van de Pol, J. E. (2020). Test- or judgement-based school track recommendations: Equal opportunities for students with different socio-economic backgrounds? *British Journal of Educational Psychology, 91*(1), 193-216. https://doi.org/10.1111/bjep.12356

van Leest, A. M. C., Hornstra, T. E., van Tartwijk, J. W. F., & van de Pol, J. E. (2024). Teachers taking perceptions of student attributes into consideration when formulating track recommendations? *British Educational Research Journal, 50*(3),1127-1171. https://doi.org/10.1002/berj.3957

van Leest, A. M. C., van de Pol, J. E., van Tartwijk, J. W. F., & Hornstra, T. E. (2024). Inconsistency in student achievement across subject domains: Examination of associations with students' gender, socioeconomic status, and teachers' track recommendations. *Educational Research and Evaluation, 29*(5), 274-298. https://doi.org/10.1080/13803611.2024.2345197

Bronkhorst, L.H., Renes, M.J., Bon, D.L., Rookmaaker, N.S., & van Leest, A.M.C. (2024). Learning to change the world: Dis/continuity in learning across climate activism and life-wide contexts. *Frontline Learning Research,* 12(3), 1–19. https://doi.org/10.14786/flr.v12i3.1381

Papers in Progress

van Leest, A. M. C. Hornstra, T. E., van Tartwijk, J. W. F., & van de Pol, J. E. (2024). Teachers' track recommendations and students' secondary school performance: The role of achievement growth. [Manuscript under review].

Conference Presentations as Presenting Author

van Leest, A. M. C., de Kleijn, R. A. M., Hornstra, T. E., & van Tartwijk, J. W. F. (2016, November). *The effects of teacher expectations on students' educational careers.* Round table at the ICO International Fall School (ICO IFS2016), Bad Schussenried, Germany.

van Leest, A. M. C., de Kleijn, R. A. M., Hornstra, T. E., & van Tartwijk, J. W. F. (2017, April). *Teacher to test.* Paper presented at the ICO National Spring School (ICO NSS2017), Utrecht, the Netherlands.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2019, June). *De po-vo-overgang: De invloed van SES op het basisschooladvies.* Paper presented at the Onderwijs Research Dagen (ORD), Heerlen, the Netherlands.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2021, August). *Students' track recommendation, and performance development in primary and secondary education.* Conference of the Junior Researchers of EARLI (EARLI), Online.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2021, August). *Students' (in)consistent performance and the relation with SES, gender, and track recommendations.* Paper presented at the European Association for Research on Learning and Instruction Conference (EARLI), Online.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2021, July). Basisschooladviezen bij leerlingen met inconsistentie in prestaties (taal- versus reken-scores). In J. E. van de Pol, J. E. (Chair), *Basisschooladviezen bij de po-vo overgang: Totstandkoming, overwegingen en bijstellingen op leerkracht-, school- en bestuursniveau.* Symposium conducted at the Onderwijs Research Dagen (ORD), Online.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2022, July). Zouden basisschoolleerkrachten hun perceptie van leerlingkenmerken moeten meewegen bij het opstellen van basisschooladviezen? In M. Dijks (Chair), *Basisschooladviezen bij de po-vo overgang: Totstandkoming, overwegingen en bijstellingen op leerkracht-, school- en bestuursniveau.* Symposium conducted at the Onderwijs Research Dagen (ORD), Hasselt, Belgium.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2023, July). *Basiss-chooladviezen en prestatieontwikkeling van leerlingen in het basis- en voortgezet onderwijs.* Paper presented at the Onderwijs Research Dagen (ORD), Amsterdam, the Netherlands.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2023, July). *Basisschooladviezen en kansengelijkheid: Wat weten we al en wat nog niet?* Round table at the Onderwijs Research Dagen (ORD), Amsterdam, the Netherlands.

van Leest, A. M. C., Hornstra, T. E., van de Pol, J. E., & van Tartwijk, J. W. F. (2024, July). *Op het juiste (onderwijs)pad?* Spotlight session at the Onderwijs Research Dagen (ORD), Tilburg, the Netherlands.

Professional Contributions

van Look, C. J., Broekhuizen, M. L., Damhuis, C. M. P., Henrichs, L. F., Hornstra, T. E., de Jong, E. M., de Kleijn, R. A. M., van Leest, A. M. C., van Tartwijk, J. W. F., van Aarsen, E., Bomhof, M., de Wit, W., & Bisschop, P. (2018). Onderwijskansen in Utrecht (K. van Look & J. van Tartwijk (eds.)). [Educational opportunities in Utrecht]. Universiteit Utrecht, Oberon Onderzoek en Advies & SEO Economisch Onderzoek. https://www.oberon.eu/media/23dpkqsv/eindrapport-gelijke-onderwijskansen-in-utrecht-2018.pdf

van Leest, A. M. C., Hornstra, T. E., van Tartwijk, J. W. F., & van de Pol, J. E.. (2020). *Eindtoets en gelijke kansen. Een vergelijking voor en na.* Web publication/site, Didactief. https://didactiefonline.nl/artikel/eindtoets-en-gelijke-kansen-een-vergelijking-voor-en-na

ICO Dissertation Series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, Radboud University Nijmegen, University of Amsterdam, University of Antwerp, University of Ghent, KU Leuven, Université Catholique de Louvain, University of Groningen, University of Twente, Utrecht University, Vrije Universiteit Amsterdam, and Wageningen University, and formerly Tilburg University (until 2002). The University of Groningen, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series.

List update 23 August, 2024 (the list will be updated every year in January)

503. Vulperhorst, J.P. (08-04-2022) Students' interest development prior, during and after the transition to a higher education programme. Utrecht: Utrecht University
504. Aben, J.E.J. (21-04-2022) Rectifying errors: A reconceptualization of the role of errors in peer-feedback provision and processing. Groningen: University of Groningen
505. Kooloos, C.C. (19-05-2022) Eye on Variety: The teacher's work in developing mathematical whole-class discussions. Nijmegen: Radboud University
506. El Majidi, A. (20-05-2022) Debate as a Tool for L2 Learning. Investigating the Potential of In-Class Debates for Second Language Learning and Argumentation Skills. Utrecht: Utrecht University
507. Van Halem, N. (24-05-2022) Accommodating Agency-Supportive Learning Environments in Formal Education. Amsterdam: Vrije Universiteit Amsterdam
508. Lee, Y.J. (8-06-2022) The medical pause in simulation training. Maastricht: Maastricht University
509. Loopers, J.H. (14-06-2022) Unravelling the dynamics of intrinsic motivation of students with and without special educational needs. Groningen: University of Groningen
510. Neroni, J. (17-06-2022) The Adult Learning Open University Determinants (ALOUD) study: psychosocial factors predicting academic success in adult distance education. Heerlen: Open University of the Netherlands
511. Schlatter, E. (17-06-2022) Individual differences in children's scientific reasoning. Nijmegen: Radboud University

512. Bransen, D. (22-06-2022) Beyond the self: A network perspective on regulation of workplace learning. Maastricht: Maastricht University
513. Valero Haro, A. (27-06-2022) Fostering Argumentation with Online Learning Systems in Higher Education. Wageningen: Wageningen University
514. Schat, E. (05-07-2022) Integrating intercultural literary competence: An intervention study in foreign language education. Utrecht University
515. Soppe, K.F.B. (07-07-2022) To Match or not to Match? Improving Student-Program Fit in Dutch Higher Education. Utrecht: Utrecht University
516. Biwer, F. (08-07-2022) Supporting students to study smart – a learning sciences perspective. Maastricht: Maastricht University
517. Weijers, R.J. (15-09-2022) Nudging Towards Autonomy: The effect of nudging on autonomous learning behavior in tertiary education. Rotterdam: Erasmus University Rotterdam
518. Van der Linden, S. (14-10-2022) Supporting teacher reflection in video-coaching settings. Enschede: University of Twente
519. Wijnen, F. (28-10-2022) Using new technology and stimulating students' higher-order thinking: a study on Primary school teachers' attitudes. Enschede: University of Twente
520. De Jong, W.A. (02-11-2022) Leading collaborative innovation in schools. Utrecht: Utrecht University
521. Bron, R. (04-11-2022) Collaborative course design in higher education – a team learning perspective. Enschede: University of Twente
522. De Vries, J.A. (18-11-2022) Supporting teachers in formative assessment in the classroom. Enschede: University of Twente
523. Le, T.I.N.H. (29-11-22) Towards a democratic school. Leiden: Leiden University
524. Wildeman, E. (30-11-2022) Vocational teachers' integrated language teaching. On the role of language awareness and related teaching behaviour. Eindhoven: Eindhoven University of Technology
525. Wolterinck, C.H.D. (2-12-2022), Teacher professional development in assessment for learning. Enschede: University of Twente
526. Kuang, X. (07-12-2022) Hypothesis generation, how to put students into motion. Enschede: University of Twente
527. Emhardt, S.N. (15-12-2022) You see? Investigating the effects of different types of guidance in eye movement modeling examples. Heerlen: Open University of the Netherlands
528. Mordang, S.B.R. (07-02-2023) Challenges in high-value cost-conscious care training of residents: Exploring the various stakehold
529. Alhanachi, S. (09-02-2023) Creating Space for Culturally Responsive Teaching: A multi-method study in Dutch secondary education Rotterdam: Erasmus University Rotterdam
530. Sleenhof, J.P.W. (12-05-2023) Secondary school teachers' collective allocation decisions: On the role of structure and interaction, objectivity issues, and teacher support Eindhoven: Eindhoven University of Technology

531. Oudman, S. (15-05-2023) Improving Students' and Teachers' Judgments of Student Performance in Primary School Mathematics Problem Solving. Utrecht: Utrecht University
532. Van Hooijdonk, M. (26-05-2023) Assessing Creative Problem Solving in Primary School Students, Utrecht: Utrecht University
533. Van Dijk, E.E. (02-06-2023) Teacher Expertise of Academics in Research-Intensive Universities – A Task-Based Approach Utrecht: Utrecht University
534. Eegdeman, I.M. (02-06-2023) Enhancing Study Success in Dutch Vocational Education Amsterdam: Vrije Universiteit Amsterdam
535. Bogaerds-Hazenberg, S.T.M. (05-06-2023) Text structure instruction in Dutch primary education: Building bridges between research and practice. Utrecht: Utrecht University
536. Zhang, S. (29-06-2023) Finger Pointing as a Cognitive Load Self-Management Strategy for Learning From Split-Attention Examples Rotterdam: Erasmus University Rotterdam
537. Wammes, D.S. (01-09-2023) Diagnosing Engineering Skills in Primary Classrooms Utrecht: Utrecht University
538. Draijer, J.M. (08-09-2023) Interest in Development: Understanding the Situated, Idiosyncratic, and Multiple Nature of Adolescent
539. Adams, T. (15-09-2023) Learning to navigate classroom dynamics: Studying student teachers' classroom management learning during their teacher education internship Wageningen: Wageningen University
540. Boels, L.B.M.M. (20-09-2023) Interpreting statistical graphs and making conjectures from these graphs Utrecht: Utrecht University
541. Baas, M.A.A., (5-10-2023) Open to all, not known to all. Sustaining practices with open educational resources in higher education Leiden: Leiden University
542. Veugen, M.J. (18-10-2023) Formative assessment practice in secondary education Wageningen: Wageningen University
543. Veugen, M.J. (18-10-2023) Formative assessment practice in secondary education Wageningen: Wageningen University
544. Megawanti, M (13-11-2023) Access and equity in higher education in Indonesia: A Case Study into the Role of Social Factors for Low SES Students Wageningen: Wageningen university
545. Dijks, M.A. (16-11-2023) What's next? A study on subject choices in Dutch secondary education Groningen: University of Groningen
546. Bouchée, T. (24-11-2023) Teaching and learning of quantum physics using digital materials Eindhoven: Eindhoven University of Technology
547. Rivera, D. (01-12-2023) Social learning in MOOC discussion forums. Pedagogical design of MOOC forums to facilitate meaningful interactions Louvain: Université Catholique de Louvain

Dankwoord (Acknowledgements)

De titel van dit proefschrift ‘On the right track’ kun je op verschillende manieren vertalen, zoals ‘Op het juiste spoor’ of ‘Op het juiste pad’. De eerste vertaling ‘Op het juiste spoor’ past goed bij dit proefschrift, omdat het niet alleen iets zegt over of leerlingen op het voor hen ‘juiste’ spoor (niveau) zijn beland, maar ook of leraren en het gemaakte onderwijsbeleid op het ‘juiste’ spoor zitten wat betreft advisering. In dit proefschrift gaat het dan vooral over de vraag of de meest voorspellende factoren worden meegenomen in het basisschooladvies. De tweede vertaling, ‘Op het juiste pad’, richt zich in mijn ogen alleen op mensen en niet zozeer op onderwijsbeleid. In de onderwijscontext gaat het voor leerlingen dan vooral om het bewandelen van je eigen onderwijspad, waarbij verschillende routes en opties tot een diploma mogelijk zijn. Ik zie hierbij een vergelijking met wandelen: als je in je vrije tijd gaat wandelen, zijn er ook verschillende paden die je kunt nemen naar een bepaalde eindbestemming. In beide gevallen geldt: welke route je neemt, hangt af van jezelf, maar ook van andere factoren waar je misschien geen of minder invloed op hebt. Ook ik heb een bepaald pad tijdens dit promotietraject bewandeld, zoals ik ook in het voorwoord al even aanhaal. Een deel van dit pad is visueel weergegeven door de fantastische illustraties van Karin van der Vegt. De illustraties bij elke hoofdstuk laten zien welke dingen ik op mijn PhD-pad ben tegengekomen, misschien herkennen jullie ze wel? De antwoorden zijn te vinden aan het einde van dit dankwoord. Een ander deel van dit pad vertel ik hieronder. Wandel je met me mee?

Overigens, het kan een lange wandeling worden, want zoals altijd ben ik lang van stof (ik hoef gelukkig niet te verdedigen waarom mijn dankwoord zo lang is), dus ga lekker zitten met een borrel (0.0 mag ook) of een bakje koffie erbij.

Mijn pad[1]

"Life is like a road that you travel on"
Tom Cochrane – Life is a highway

Naast Nederlandse literatuur – maar dat zit al verwerkt in het voorwoord – is muziek mijn grote liefde. Dat kan natuurlijk niet ontbreken in dit dankwoord. Deze songtekst is niet alleen een verwijzing naar het (onderwijs)pad die elk van ons bewandelt, maar ook naar de letterlijke wegen waarop een deel van dit proefschrift is geschreven. Zo is het begin van dit PromoDoctraject, de – toen nog – pilot van startende leraren die hun onderwijsbaan combineren met promoveren, dat in 2015 startte met een voorstel dat ingediend moest worden bij OCW. Dit voorstel heb ik voornamelijk geschreven in een bus op weg naar een skivakantie in Oostenrijk met de middelbare school waar ik toen werkte. Toen was nog niet duidelijk waar de weg naartoe zou leiden, want eerst moest het voorstel nog goedgekeurd worden. Hoelang de weg zou duren ook niet. Het onbekende, het avontuur, de uitdaging, dat is voor mij wel weggelegd. "Ik heb het nog nooit gedaan, dus ik denk dat ik het wel kan."

Hoewel ik dit dankwoord nu aan de eettafel van mijn huis in Hilversum aan het schrijven ben, heb ik de laatste hand aan de 'echte' inhoud van dit proefschrift in de auto in de VS geschreven, terwijl Jan mij en meestal slapende Nola door het Amerikaanse landschap reed op weg naar ons volgende hotel. En overigens ging de laptop ook geregeld tussendoor mee op vakantie 'om toch nog even wat af te maken'.

Maar taadaaa, hier ligt hij, mijn proefschrift. 9 jaar, 3 middelbare scholen, 1 hbo en 1 universiteit verder (of eigenlijk 2 als je de masterscriptie aan de UvA meetelt die ik tussendoor ook nog 'even' – over een heel ander onderwerp dan mijn proefschrift – schreef). In die negen jaar met verschillende werkplekken ben ik natuurlijk veel mensen tegengekomen die ik graag wil bedanken.

Regenboogpad[2]

"And you know,
We're on each other's team"
Lorde – Team

---

1 Wil je ervaren hoeveel bloed, zweet, tranen en (figuurlijke) kilometers ik heb afgelegd tijdens mijn PhD-traject? Misschien is deze wandelroute dan wat voor jou: Het Takke End, een wandelroute door de Botanische Tuinen in Utrecht: https://www.uu.nl/sites/default/files/Het_Klaverpad_wandelroutes.pdf

2 Dit was – destijds in ieder geval – het langste regenboogpad ter wereld: https://www.uu.nl/organisatie/equality-diversity-inclusion/regenboogfietspad

Geen pot goud, maar een proefschrift aan het einde van het regenboogpad, dat mooi voor ons Langeveldgebouw ligt. Om dat te bereiken hebben wij, Jan, Lisette, Janneke en ik, als (begeleidings)team met elkaar samengewerkt. En ik ben blij dat ik dit zo kan stellen, want dat samenwerken was vooral in het begin niet zo vanzelfsprekend, omdat ik veel moeite had met de transitie naar de (soms voor mij hiërarchisch voelende) PhD-context. Het voelde dan vaak ook niet aan zoals het verharde regenboogpad, maar eerder als een modderig zandpad[3], waarbij je geregeld wegzakt, vast blijft zitten en dan niet weet hoe je verder moet. Team up, is mijn advies!

Jan, wij ontmoetten elkaar aan het einde van de lerarenopleiding en jij zag wel wat in het onderzoeksthema gelijke kansen rondom de overgang van po naar vo. Een – in jouw woorden – 'lollig projectje' was je aan het opzetten, PromoDoc, of dat niet iets voor mij was? Eenmaal met een goedgekeurd voorstel regelde jij het wel – zoals altijd – en binnen no time zaten we met de gemeente Utrecht en POVO om tafel. Van presentatie bij de minister van OCW in Den Haag tot de sollicitatie voor mijn huidige UD-plek: jij herinnert me gelukkig regelmatig aan de eerdergenoemde 'niet van Pipi Langkous afkomstige'-quote. Wat je me echter nog wel even moet leren is het maken van die flitsende PowerPoints, want daarmee steel jij altijd (expres) de show. Naast deze hobby heb je ook nog de doorgewinterde klusdienst Van Tartwijk en co. Overigens, voor mij geldt de Pippi-uitspraak niet zo met betrekking tot klussen, dus mocht je je ooit eens vervelen, ik heb nog wel wat klusjes in de aanbieding thuis...

"Whatever it takes I will stay here with you
Take it to the good times
See it through the bad times"
Starship – Nothing's gonna stop us now

Lisette, ik wil je zo ontzettend bedanken voor jouw geloof en vertrouwen in mij, want mij begeleiden was niet altijd makkelijk. Dat je het traject desondanks verder met me aan durfde te gaan, was voor mij genoeg om in mezelf te gaan geloven (misschien hoor ik hier toch wél thuis?). Je had het bij het rechte eind blijkt nu, want ik heb mijn droombaan gevonden, en nog wel met jou als kamergenoot! F3.38: we're here to stay (tot we naar de overkant verhuizen). Als de koffie maar goed is!

3 Om het modderige te ervaren raad ik – uit eigen ervaring – deze wandelroute aan, met name in een natte periode van het jaar: https://www.staatsbosbeheer.nl/uit-in-de-natuur/wandelroute-wolvenspoor

Janneke, ik zocht versterking voor mijn begeleidingsteam, en daar was jij! Van pilots van jouw onderzoeksproject uitvoeren in mijn vo-klassen tot het samen schrijven van een onderzoeksvoorstel: inmiddels zijn we goed op elkaar ingespeeld en vullen we elkaar aan. Tot in Brabant!

Renske, hoewel wij het traject niet samen hebben afgemaakt, wil ik je bedanken voor je positiviteit en je eerlijkheid.

Catwalk (aka Poezenpad)[4]

我們一起學貓叫 一起喵喵喵喵

"Whatever you do I do meow meow meow"

Learning to meow – Xiao Fengfeng & Xiao Panpan

Leren miauwen gaat misschien wat ver, maar de liefde voor katten (en voor dit hilarische nummer: if you know, you know) was zeker aanwezig in F3.01, onder andere te zien aan de verschillende kattenfoto's die we op onze lockers geplakt hadden. Dankjulliewel, mijn ex-F3.01-collega's, waaronder David, Marloes, Angela, Katrijn, Jonne, Sophie, Mei, Xiaojing, Mare, Christa, Karin, Gesa, Ellen, Susan, Lucia, Desirée, Pierre, Dannie, Larissa, Sophia, Michaela, Yuanyuan, Anouk, Linda, Simone, Selma en alle anderen die daar op een blauwe maandag een (deel van een) bureau hebben gehad of in de zomer van de airco kwamen genieten. Onze vaste middaglunch, samen proosten op successen en huilen als het wat tegenzat, het strak ingedeelde werkrooster vanwege het delen van de bureaus, de verschillende competities die we hielden, het winnen van de LV3-quiz, het bezetten van de dansvloer onder leiding van Jonne, samen musea bezoeken en naar de Efteling gaan, online ontbijten, theedrinken en borrelen tijdens de coronalockdowns, maar bovenal de dansjes die we samen hebben gedaan op mijn bruiloft: wat was het fijn met jullie! En natuurlijk zal ik de avontuurlijke reis in het busje naar Bad Schussenried voor ICO International Fall School met een aantal van jullie nooit vergeten: in het donker, met weinig slaap en 'ik heb nog nooit in een busje gereden, dus ik denk dat ik het wel kan'. Afscheid nemen was vaak moeilijk, maar gelukkig zijn veel van jullie nog steeds collega's of komen we elkaar toch in het werkveld weer tegen.

Mei, Xiaojing & Yuanyuan, unfortunately, our trip to China was cancelled due to Covid, but hopefully we will meet again in person! Sophia & Michaela, being the voice of both of your research projects was an honour. Sophia, I will miss your kindness and generosity, but most of all your pantoffels. Michaela, let's replace our hugs with digital ones? Sophie, dank voor de datapackagecheck en onze fijne gesprekken over de kind-/werkstruggles in

4 Wandelen langs katten van graffiti kun je doen in Wijnaldum: https://www.harlingenwelkomaanzee.nl/nl/bezoeken/routes/2649598042/kattenroute-wijnaldum

het leven, maar nog meer voor het feit dat je Janneke met me wilde delen. Ontzettend tof dat we nu samen een project gaan doen! Jonne, de leukste buurman van LV3 met de beste moves die je maar kunt hebben. Angela, Esther en Jael, samen doorliepen we de eindfase van onze PhD's – alhoewel op mijn verdediging nog even gewacht moest worden – en de daarbij horende onzekerheden en toekomstvragen. Esther, de Annejurken (thanks for that) staan je geweldig! Jael, bedankt dat je jouw artistiek talent ook weer in wil zetten tijdens mijn verdediging door het maken van foto's. Angela, elkaars paranimf zijn is toch wel een prachtige afsluiting van onze PhD-trajecten. De laatste jaren met jou in F3.01 waren zó fijn. Samen de weg naar een nieuwe baan, waarbij we allebei voor iets anders kozen (en jij inmiddels al doctor bent), maar daarin toch samen de verbinding zoeken in combinatie met het moederschap. Een ritje naar de Zeeman om babykleertjes te shoppen tijdens de ORD2022 kon dan natuurlijk ook niet ontbreken (al is het echte verhaal dat we bij de verkeerde school voor een rondleiding stonden). David en Marloes, zoals gezegd ging ons tripje naar China niet door. Wat hadden we daar ZIN in! Hopelijk ooit, want dromen moet je hebben toch? Wat hebben we veel gelachen samen. Marloes, we kunnen eindelijk zeggen: we made it, we saved the best for last! Dank voor je support het laatste jaar waarin we het toch allebei alleen moesten doen. David, dankjewel voor jouw deur die altijd openstaat. Niet alleen om werkdingen te bespreken, maar tegenwoordig kunnen we ook samen miepen over ons tekort aan slaap. Last, but not least, Katrijn, voor jou heb ik eigenlijk maar twee tekens nodig: <3. Mijn steun en toeverlaat tijdens de beginjaren van mijn PhD, niet alleen op het werk, privé, maar ook in het vinden van ons duurzame pad.[5] Wát ben ik blij met jou als paranimf!

Mijn verdere onderzoekspad[6]

Mijn verdere onderzoekspad heb ik niet alleen met mijn begeleidingsteam of kamergenoten aan mijn zij bewandeld, maar ook zeker met andere collega's. Mede-PromoDocs Mare, David, Marloes, Sophie, Esther, Floor, Anne, Paulien, Sophie en Saro, alhoewel ons grote Promodocsoverleg op nogal onregelmatige momenten plaatsvond, waren de inhoudelijke sparsessies, maar nog meer het uitwisselen van ervaringen over hoe we het allemaal gingen combineren ('eh Jan?'), heel fijn. Ook de mede-promovendi van de andere kamer Renée, Monika, Eva, Brechje, Steven, Alex, Joris en Minke, bedankt, net als de fantastisch 'nieuwe' (zo nieuw is het niet meer voor iedereen) PhD-lichting. Hoewel onze PhD-paden elkaar net kruisen, vind ik het geweldig om jullie op mijn onderzoekspad tegen te komen, voor nu met name Jonne, Femke en Teuntje.

---

5 Ons duurzame pad is niet uit te schrijven in een wandeling, maar gelukkig kun je voor een wandeling langs duurzame initiatieven ook gewoon terecht in Amsterdam: https://www.iamsterdam.com/zien-en-doen/natuur-en-actief/routes/wandelroute-voor-een-betere-planeet

6 Het Onderzoekspad is ook een wandelroute door de Botanische Tuinen in Utrecht: https://www.uu.nl/sites/default/files/Het_Klaverpad_wandelroutes.pdf

Daarnaast wil ik alle andere GST-/ALPO-/O&T-/LV3-collega's en studenten bedanken, in het bijzonder collega's Martine, Ineke, Larike, Lotte, Anouk, Joke, en mijn lesgeefbuddy's Hanneke, Anna, Nienke, Marijke en Peter. Ik heb zoveel geleerd en gelachen tijdens de voorbereiding en uitvoering van onze werkgroepen, waarin we het doen van rollenspellen soms wel heel letterlijk namen. Oud-student Loes Ypma, bedankt dat jouw missie onderdeel mocht zijn van mijn pad. Monique, expect the unexpected (en voor al je andere geweldige titels weet je nu bij wie je moet zijn), daar gaan we voor, júíst wanneer we niet helemaal zo goed voorbereid zijn als dat we zouden willen. Pascale, zo leuk om je na Eerst de Klas en de landelijke debatwedstrijden weer tegen te komen op LV3. Luce, dank voor al je support en wijze lessen. Monika, LV3-overbuurvrouw, dank voor je luisterende oor als we allebei weer eens met wallen tot op onze tenen, verward haar en koffie iets te laat op kantoor aan kwamen zetten. We doen het toch allemaal maar 'even'!

Voor de dataverzameling een shoutout naar Harriët Smit, Yasmina Daoudi en Iris Kensenhuis (van POVO destijds), en voor de nauwe samenwerking voor het 'gemeenteproject' naar collega Karin van Look en Marjolein Bomhof. Karin, dank voor je creatieve onderzoeksideeën tijdens de sparsessies met leerkrachten en schoolleiders.

Verder ben ik in mijn huidige functie als UD op ontdekkingstocht voor nieuwe onderzoekspaden. Samenwerken met andere onderzoekers en (leraren in de) scholen is wat ik het liefst in mijn onderzoekstijd doe. One thing is for sure: er komt veel moois aan! Zó fijn jullie ontmoet te hebben, Pomme en Antoinette.

Als laatste kruiste mijn onderzoekspad die van jou, Monique Dijks. Ook al zijn we geen directe collega's, we kwamen elkaar regelmatig tegen tijdens onze PhD's en nu zitten we samen in het bestuur van de VOR, divisie Onderwijs & Samenleving. Het voelt altijd goed en vertrouwd, alsof we elkaar al jaren kennen, en letterlijk gezien is dat ook zo, maar we zien elkaar natuurlijk niet wekelijks. Tijd weer voor een etentje!

Mijn onderwijspad[7]

Zo lijkt het bijna alsof ik het PromoDoctraject alleen op de universiteit heb afgelegd, maar niets is minder waar. In dit traject was ik juist voor het grootste gedeelte werkzaam in het onderwijsveld, namelijk 3 (maar in de praktijk werden dat al snel 4 of 5) dagen voor de klas en slechts 2 op de uni.

7 Een leuke onderwijswandelroute: https://www.wandel.nl/routes/wandelen-langs-bijzondere-scholen/

"I looked at the Rubens and Rembrandts
I liked the John Singer Sargents"
The art teacher – Rufus Wainwright

Mijn onderwijspad als docent Nederlands begon op het VeenLanden College in Mijdrecht, mijn eigen oude middelbare school. Dit was mijn start. Hier kwam ik erachter dat lesgeven misschien toch wel mijn ding was in plaats van 'ik weet niet wat ik wil, dus ik probeer gewoon wat', en wat heb ik van die reis genoten. Naast lesgeven heb ik hier ook de overgang po-vo in de praktijk gezien en heb ik mogen proeven aan andere taken, zoals mentor en MR-lid. Ik was wel een vreemde eend in de bijt, iets waar de schoolleiding niet helemaal (of beter gezegd helemaal niet) mee om kon gaan. De keuze om iets anders te gaan zoeken was niet makkelijk. Dank voor je altijd aanwezige steun, Joke Kok. Door jou voelde ik me echt welkom. Wát een geweldige mentorklas hadden we samen, Réka. En Yacintha, inmiddels kunnen we lachen om het feit dat ik (ja, echt, je gelooft het niet) zelfs een (onofficieel) verbod kreeg om de school te betreden. Dank voor het zijn van mijn chaperonne, als ik toch weer eens de school in wilde. Dank voor alle gezellige (en hopelijk nog meer toekomstige) kunstanalyses En Museumclubleden Yacintha, Réka, Marieke, Tamara, Rianne en Karlijne. Wanneer gaan we weer Rubens en Rembrandts bekijken?

"Ik kom niet alleen
Want ik heb chips en cola"
Chips & Cola – Lil' Kleine & Ronnie Flex

Na een korte, fijne tussenstop op het Minkema College in Woerden (waar ik het uitstekend kon vinden met de beruchte 5 vwo, tot uiterste verbazing van de rector) belandde ik op het Erfgooiers College in Huizen. Vanaf het moment dat ik daar binnenliep, had ik een goed gevoel en dat bleek ook te kloppen. Wat een heerlijke school, waar ik ook zeker weer uitvoerig allerlei activiteiten en taken kon oppakken: mentor, sectieleider, Lesson study-begeleider, deelnemen aan het landelijk netwerk Formatief Evalueren van SLO, leesvaardigheidstoetsen maken voor Cito, MR-lid, debatteam- en leesclubleider, en PR- en websitebeheer. Dankjewel Hans en Ineke voor alle kansen. Ik hoop dat jullie het me inmiddels hebben vergeven dat ik letterlijk de voorlaatste dag voor de zomervakantie mijn baan opzegde. Chips & cola, oftewel Marissa en Pauline, ik mis jullie tot op de dag van vandaag. Wat waren wij een geweldig onderwijsteam. Elk jaar naar de HSN-conferentie, in de laatste week bij de Hema schoolspullen voor het nieuwe jaar scoren (alsof je dat als docent echt nodig hebt) en een ijsje met spikkels bij de Jamin halen, en knakworsten eten als jullie jarig waren. De wijsheid liever lui dan moe, daarin konden wij ons wel vinden. Op de vrijdagmiddag komt nog vaak de (chips en) cola tevoorschijn, maar dat haalt het toch niet bij die uit de automaat, en vooral niet met mijn laptop als vervanging voor jullie gezellige aanwezigheid.

Bij deze ervaringen wil ik ook niet mijn ontzettend leuke oud-leerlingen vergeten. Zonder jullie was deze start nooit zo succesvol geweest. Recent ben ik meerdere van jullie tegengekomen, en wat blijft het bijzonder dat we elkaar na zoveel jaar nog kennen en zulke fijne herinneringen hebben aan deze tijd. Dat is voor mij een teken dat ik mijn juiste pad bewandeld heb.

"We are not what you think we are
We are golden, we are golden"
We are golden – Mika

Na het vo-avontuur kwam ik in het hbo terecht, bij Aeres Hogeschool in Almere, via Anca. Wat een speeltuin om een curriculum Nederlands te mogen opzetten binnen een hbo dat hoog staat aangeschreven om hun duurzaamheidsgedachte, perfect passend bij mij. Wij als onverslaanbaar gouden duo, elkaar nog kennende van Eerst de Klas. Wat mis ik jou aan mij zij in mijn onderwijsbaan, want jij haalde het beste in mij naar boven. Gelukkig hebben we onze uitstapjes naar de kringloop en de Ancast (voicememo's die hele podcastafleveringen zijn) nog. Ik kijk uit naar onze eerste conferentieworkshop samen, of gaan we toch echt dat boek samenstellen?

Vrienden langs mijn pad[8]

Sommige (ex-)collega's die hierboven genoemd zijn, zijn ook zeker tot mijn vriendenkring gaan horen, maar er zijn ook vrienden die geen collega's zijn. Voor jullie allemaal geldt: hoewel ik het persoonlijk de afgelopen twee jaar 'wat' moeilijk had, kan (en vooral wil) ik niet zonder jullie, en jullie hopelijk ook niet zonder mij. Tegen een aantal van jullie wil ik graag nog iets persoonlijks zeggen hieronder (in alfabetische volgorde), maar ik wil ook noemen: de Alivegang van RTV Ronde Venen – al zijn we niet zo vaak meer alive op de radio tegenwoordig – Roel, love you so much, Tijmen, Elmar, Elroy, Joris, wat is het altijd fijn om jullie te zien; de twee moederclubs waarmee ik alles rondom het moederschap kan delen: de Loedermoeders (jaja, ze bestaan echt) en de BAEBS (de Bevallen Als Een Baas-tribe met extra liefde voor Graciëlla); en Judith, Ruud en verjaardagsgenoot GJ.

"Een van mijn grootste liefdes, ja m'n beste vriend
Dwars door alle stormen bleef je staan
Soms mis ik zelf de kracht
Maar met jou kan ik het aan"
Met jou kan ik het aan – Anouk

8 Vrienden samen bedachten het Vriendenpad, een mooie wandelroute in de omgeving van Vlaardingen: https://vlaardingendoen.nl/broekpolder-vriendenpad-route

Frederike (Fre), van Anoukfans tot getuigen bij elkaars bruiloft: wie had daar vooraan bij het podium van Anouk kunnen bedenken dat we beste vriendinnen zouden worden, en bovenal nog steeds zo fan van Anouk zijn als toen. Het hele land door crossen, en uren en uren in de wachtrij staan voordat een Anoukshow begint, bepakt met pakjes appelsap, stroopwafels en grote paperclips (om aan de barriers maken om onze jas op te hangen), inmiddels gaan we ook altijd naar de concerten van Beth Hart, lunchen bij At seven, een ijsje halen bij Pisa en samen met onze gezinnen naar Disneyland, een dierentuin of de Efteling. Ook alle liefs voor Stefan, en mijn grote vriendinnen Lora en Minte. Wat is het een eer om hen zo van dichtbij groot te mogen zien worden.

"You are good, all the time
And all the time, you are good"
You are good – Israel Houghton

Marcella (Mars), al op de basisschool werden wij beste vriendinnen en dat bezegelden we door in groep 8 samen naar een optreden van de Spice Girls te gaan. Hoewel we op de middelbare school niet alle jaren bij elkaar in de klas zaten, belden we elkaar zo ongeveer elke middag op om te vragen of we konden afspreken. Lekker theedrinken met chocola, genieten van je moeders zelfgemaakte kippensoep, Mariakoekjes bij koor eten, Maria en Jozef (of ook weleens de ezel en Maria haha) spelen in het kerststukje, optreden als Slimme Schemer, een jongerenvakantie naar Terschelling, met World Servants naar Egypte en Zambia om daar vrijwilligerswerk te doen, en samen op naailes (jij ging vaker dan ik), waar we toch maar mooi samen ons eerste kledingsstuk voor de kinderen naaiden: ik had al die avonturen met niemand anders aan mijn zij willen beleven. En toen even wat minder met mij ging, kwam jij onaangekondigd toch even langs om te kijken hoe het met me is. You're the best!

"Als ik je vergeet geloof jij nog steeds in mij
Ook als ik weer verpest om iets goeds te zijn
De helft van wat je doet
Is voor mij al meer dan goed"
De helft van wat je doet – Suzan & Freek

Stephanie, als Loedermoeders leerden we elkaar in een nogal hectische appgroep kennen. En hoewel we elkaar pas twee jaar privé spreken, zijn wij, maar ook onze kinderen, verknocht aan elkaar. Dat kan ook niet anders, want elkaar een dag niet gebeld hebben, is een dag niet

9 Een leuke tip voor een wandeling met kleine kinderen: een Kabouterpad. Met heuse kaboutermutsen en een echt diploma aan het einde: https://www.staatsbosbeheer.nl/uit-in-de-natuur/kabouterpad

geleefd (06-mijnietbellen). Wij zijn twee totaal verschillende types, maar over veel dingen denken wij exact hetzelfde. En zo niet, dan is dat ook goed. Jij bent degene die altijd in mij gelooft en voor me klaarstaat. Mijn liefde voor de DM en de Kik heb ik van jou (lees: de halve winkel leegkopen), en als koopjesjagers met gierende banden naar de Hema sale of de tweedehands kinderkledingsbeurs, terwijl we eigenlijk niks nodig hebben. Volgens jouw eigen woorden ben ik je vriendin, echtgenoot en huisgenoot in één. Nou andersom is dat volgens mij ook zo, vraag maar aan Jan. Je bent inmiddels onderdeel van ons meubilair. Kusjes aan Tommy, ook van Nola. Laten we snel weer een Kabouterpad[9] doen, aka op boevenpad[10] gaan in de ogen van onze kinderen.

"Make it last forever
Friendship never ends"
Wannabe – Spice Girls

Lieve Anne, Laura, Lottie, Marjolein, Marlou en Nathasja, ook wel bekend als de Spekjes. Sommige van jullie waren studiegenoten, sommige huisgenoten, sommige allebei, maar bovenal mijn Amsterdamse vriendinnen die nu bijna allemaal niet meer in Amsterdam wonen. Wij zijn gek op Martini's (in een vorig leven dan), stapelfoto's, plakbh's, lipdubs (en vooral de bloopers daarvan), vriendinnenweekend Chicks on Tour, verstoppertje spelen, last minute fotocollageverjaardagskalenders maken wanneer je 30 wordt (dat is alweer even geleden…), WIDM, kledingruil organiseren (tegenwoordig ook met kinderkleding), Lays Sensations Mexican peppers & cream eten, een spelletje doen (niet altijd zonder valsspelen), maar op zijn tijd ook een goed gesprek. Op nog vele jaren, want met sommige vriendschappen weet je het gewoon: friendship never ends!

Familie langs mijn pad[11]

Naast mijn lieve schoonfamilie, wil ik mijn eigen familie ook bedanken. Jullie vroegen altijd zonder oordeel hoe het ervoor stond en via papa of mama volgden dan de updates. In het bijzonder wil ik mijn oma bedanken, die volgens haar eigen zeggen van bovenaf een oogje in het zeil zou houden.

"Love conquers all
This one will last a lifetime"
Love conquers all – Deep Purple

---

10 Het Boevenpad wandelen kan ook echt: https://www.parool.nl/ps/wandelen-in-amsterdam-langs-het-boevenpad-in-de-jordaan~b2e65f82/

11 Vlonderpaden zijn top voor alle familieleden, van groot tot klein: https://www.waanzinnigewereld.nl/dit-zijn-alle-vlonderpaden-in-drenthe/ Hieronder valt ook het eerdergenoemde Wolvenspoor, aka modderpad.

Houdoe en bedankt, papa, mama, Moniek, Jan, Kevin en Wesley. Vanaf een afstandje hebben jullie mijn pad gevolgd. Soms voelde ik me een beetje een vreemde eend in de bijt, maar altijd probeerden jullie te vragen wat ik dan precies deed, en er altijd vol liefde en met volledige steun voor me te zijn. En anders was er altijd nog Google, die hielp met het vertalen van een gepubliceerd artikel en volgens jou, mam, was dat best goed te volgen. Ondanks dat er waarschijnlijk momenten waren dat beslissingen wat minder handig leken, bijvoorbeeld wanneer ik weer eens plotseling van baan ging wisselen, lieten jullie me volledig in mijn waarde en vertrouwden jullie op mijn keuze. Samen een modderpad bewandelen op witte sneakers, een lipdub opnemen met geschminkte snorren om op papa te lijken, valsspelen met Monopoly of een rondje achter op de motor: kom maar op met dat volgende weekendje weg!

"But words they only go so far
Can't describe what's in my heart"
In love – DeWolff

"I will be here with you
Just like I told you I would
I'd love to always love you"
Rooting for you – London Grammar

En dan nu 'mijn' Jan, zoals je hier op de afdeling ook wel bekend staat na, want het is best ingewikkeld met al die Jannen. Ik heb het wel steeds over mijn pad, maar eigenlijk is dit het pad[12] dat wij samen bewandeld hebben, want zonder jou had ik het nooit zo ver geschopt. Van een kopje verse muntthee maken als ik 's avonds toch echt nog even moest werken tot de volle lading over je heen krijgen omdat die ellendige laptop mijn ingewikkelde analyses weer eens niet wilde draaien: jij was en bleef gelukkig altijd aan mij zij. Inmiddels met iets meer rimpels, wallen en grijze haren dan tijdens ontmoeting bij 'de radio', maar dankzij onze lijfspreuk 'love for the haters' niet minder verliefd (reken namelijk maar eens uit hoe groot die liefde vandaag de dag wel niet is). We kunnen elkaar geregeld achter het behang plakken, maar gelukkig hebben we geen behang. Samen op avontuur in Zuid-Italië of de VS, onze gedeelde liefde voor muziek, een goede havercappu, een lekkere IPA, sushi (bij ons gaat liefde inderdaad ook door de maag), grote supermarkten en winkels, Moordzaken – de podcast luisteren, de DM en Pokémon Go (zonder jou was ik bij het vangen van Pikachu blijven steken). Het maakt me niet uit waar ons pad heengaat, zolang het maar samen is.

12 Alsof hij voor jou bedacht is, de Ome Janroute langs Zonnestraal, het gebouw waar wij getrouwd zijn: https://www.wandel.nl/routes/zonnestraal-en-hoorneboegse-heide/

“And there she stands
Throwing both her arms around the world
The world that doesn’t even know
How much it needs this little girl
It’s all gonna be magnificent”
Magnificient – Elbow

Deze songtekst stond op jouw geboortekaartje en ik had vooraf niet kunnen bedenken hoe toepasselijk hij is. Niet zozeer dat de wereld jou nodig heeft, lieve Nola, maar vooral ik (okay en je vader misschien nog iets meer). Van 24/7 aan het werk naar het liefst alleen maar bij jou willen zijn. Jij hebt me laten voelen dat er belangrijkere dingen zijn dan werk alleen. ‘Mama hoeft niet meer te werken’ is niet alleen jouw favoriete uitspraak, maar inmiddels ook die van mij.

“Always sitting by your side
Always by your side”
Lucifer Sam – Pink Floyd

Dino*, Diva en Frits, mijn harige vrienden. Thuiswerken was nog nooit zo fijn met een van jullie naast (maar eigenlijk hadden jullie liever op) mijn laptop. Ik denk dat jullie mijn halve proefschrift inmiddels wel gelezen hebben.

“Although we’ve come to the end of the road
Still, I can't let go”
End of the road – Boyz II Men

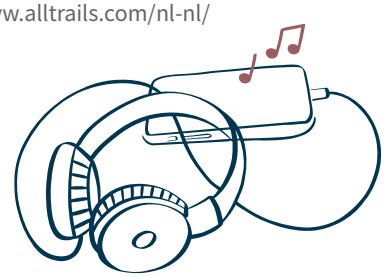
Het einde van dit pad is hier, het einde van onderzoek doen naar dit thema nog niet. Volgende paden[13] liggen klaar om met evenveel plezier bewandeld te worden.

PS: hou je, net als ik, van muziek? Scan de QR-code[14] hiernaast om alle nummers uit dit dankwoord te luisteren.

---

13 Zoals de Never End Trail in Zwitserland: https://www.alltrails.com/nl-nl/wandelpad/switzerland/grisons/the-never-end

14 Of ga naar: https://bit.ly/ProefschriftAnne



Landschappen

- **Omslag =** Herken je mijn basisschool en middelbare school?
- **Hoofdstuk 1 =** Zonnestraal in Hilversum, de locatie waar Jan en ik op 21 december 2018 trouwden.
- **Hoofdstuk 2 =** Matuku, Zambia, waar ik in augustus 2018 vrijwilligerswerk heb gedaan met World Servants.
- **Hoofdstuk 3 =** Jan en ik (en Nola) hebben verschillende reizen door de VS gemaakt. Hier zie je een mix van San Francisco en Monument Valley (2014), Washington DC en NYC (2024).
- **Hoofdstuk 4, 5 en 6 =** een mix van de verschillende landschappen die we hebben gezien tijdens onze reizen naar de VS in 2014, 2017, 2019, 2022 en 2024, onze zomers in het zuiden van Italië en de creatieve vrijheid van Karin.

Anne van Leest combined working as a secondary school teacher with doing her PhD research on equal educational opportunities at the transition from primary to secondary education. This dissertation's findings indicate that teachers' track recommendations are not the primary source of inequality during the transition from primary to secondary education. Rather, these recommendations amplify and formalise existing achievement gaps, further consolidating disparities due to the structure of the Dutch secondary education system.

Let's go on and reach
for what you want in life

Anouk – It's a new day