



Inequality in Teacher Judgements, Expectations and Track Recommendations: A Review Study

Dr. Sara Geven, Dr. Anatolia Batruch & Prof. Dr. Herman van de Werfhorst

This report has been commissioned by the Dutch Ministry of Education, Culture and Sciences. The authors are part of the research group of the Vici project "Between institutions and mechanisms: education and inequality in comparative perspective" sponsored by the Netherlands' Organisation for Scientific Research.

Table of Contents

Executive Summary: Inequality in Teacher Expectations and Track Recommendations	4
1. Introduction.....	7
2. Methodological considerations.....	9
3. Student traits and teacher expectations	10
3.1 How ‘accurate’ are teachers?.....	10
3.2 Why are teachers inaccurate? Theoretical considerations	12
3.2.1. Discrimination in sociology: Statistical vs. taste-based discrimination	12
3.2.2. Evaluations based on individual students.....	13
3.2.3 Conclusion	15
3.3 Empirical Findings	16
3.3.1 Student demographic traits.....	16
3.3.2 Student behaviour and attitudes	23
3.3.3 Home situation	26
4. Traits of teachers	29
4.1. Stereotypes and attitudes	29
4.1.1 Theoretical considerations	29
4.1.2 Empirical findings	33
4.2 Congruence between a teacher’s and students’ social group	37
4.2.1 Theoretical considerations	37
4.2.2 Empirical findings	38
4.3 Teachers’ self-efficacy.....	41



5. Classroom composition	42
5.1. Theoretical considerations	42
5.2 Empirical findings	43
6. Institutional settings	46
6.1. Theoretical considerations	46
6.2. Empirical findings	50
6.2.1 School selection function.....	50
6.2.2 Accountability.....	53
6.2.3 Track allocation policies and inequality in educational outcomes	54
7. Interventions	56
8. Implications for the Dutch debate on the role of teachers	58
References.....	63

Executive Summary: Inequality in Teacher Expectations and Track Recommendations

❖ Student traits and teacher expectations

- *Students' school performance and demographic traits*
 - Teachers evaluate students from higher socioeconomic backgrounds more positively than students from lower socioeconomic backgrounds, even if their academic performance is the same.
 - Findings are mixed with respect to teacher biases against ethnic minority students. These heterogeneous findings may be due to the fact that teachers are generally more inaccurate about students from ethnic minority groups. This means that, compared to students from the native majority, students from ethnic minority groups are more often underestimated as well as overestimated by their teachers. A Dutch study also indicates that teachers differ in the extent to which their expectations are ethnically biased: some teachers hold the same expectations for students from different ethnic groups, other teachers hold higher expectations for students from ethnic minority groups, while another group of teachers hold lower expectations for them.
 - Research shows that Dutch teachers evaluate girls more positively than equally achieving boys. However, a new study suggests that this gap has recently disappeared. Moreover, findings in other countries show no consistent gender biases.
 - It should be noted that the influence of students' demographic traits is generally small, and that teacher judgments and expectations are considered to be relatively 'accurate'. This means that teachers primarily form them on the basis of students' performance even when students' demographic traits are influential.
- *Student behaviour and attitudes*
 - (Teachers' perception of) students' school behaviour and attitudes are related to teachers' expectations for students. However, they do not seem to explain socioeconomic inequality in teacher track recommendations. According to a Dutch study, gender differences in teacher track recommendations are partly explained by gender differences in teachers' perceptions of students' school attitudes and behaviours.
- *Home situation*
 - While some research suggests that parental school involvement is higher among parents from higher socio-economic backgrounds, other findings do not provide support for this, especially with respect to parental involvement that takes place in the home (e.g., discussing and showing an interest for school-related matters). Nevertheless, parents with a higher socioeconomic status seem to be more involved in activities that take place *at school* (e.g., volunteering at school; attendance at

parent-teacher meetings), and teachers also *perceive* their school involvement to be higher.

- Teachers evaluate students more positively when they perceive them to live in an educationally supportive home. However, there are no studies that explicitly examine whether teacher perceptions of parental support or involvement can explain socioeconomic inequality in teachers' evaluations of students.
- Parents from higher socioeconomic backgrounds may exert more pressure on teachers, and/or are more likely to request for higher tracks, leading to social inequality in track recommendations. However, findings on this subject are scarce and mixed.

❖ **Traits of teachers**

➤ *Stereotypes, attitudes and intergroup biases*

- Disadvantaged students are the target of more *implicit* negative attitudes from teachers as compared to advantaged students. It is particularly the case for students whose parents are immigrants or low-educated.
- Results are more mixed when it comes to teachers' *explicit* negative attitudes or stereotypes about gender or ethnicity. Yet, some teachers' (or pre-service teachers) do explicitly hold negative stereotypes or attitudes against low socioeconomic status students.
- Teachers' implicit prejudiced attitudes are associated with classroom consequences in a Dutch study. Negative implicit attitudes against ethnic minority students are related to differential teachers' expectations as well as to ethnic achievement gaps.

➤ *Congruence between a teacher's and student's social group*

- Teachers tend to evaluate students of the same race (slightly) more positively than students of a different race in the U.S. Teacher evaluations of black students, but not those of white students, vary by teacher's race. The findings of studies focusing on teacher-student gender similarity are inconsistent.

❖ **Classroom composition**

- The relationship between class composition and teacher evaluations of students is equivocal. Research shows positive and negative relationships between the average ability of a class and teacher expectations or track recommendations, and positive and no relationship between the average socioeconomic status of a class and teacher expectations or track recommendations. These discrepant findings may be explained by country-level factors and/or the operationalization of class composition variables.

❖ **Institutional settings**

- Research on institutions focuses mostly on the relationship between educational systems' (track recommendation) policies and inequality in educational outcomes. A handful of studies have more specifically examined the impact of institutions on teachers' behaviour.
 - Research on educational systems and inequality in educational outcomes shows that early tracking (rather than late tracking) and non-binding teacher track recommendations (rather than binding ones) are related to higher levels of socioeconomic inequality in educational outcomes. However, ethnic inequalities in

educational outcomes appear smaller when teacher track recommendations are not binding instead of binding. The relationship between standardized examinations and educational inequality is complex. One study shows that they may reduce social inequality in educational outcomes in highly tracked systems, but not in more comprehensive ones.

- The use of educational practices aiming at selecting students instead of helping students improve encourages biases against low socioeconomic status students.
- Holding teachers' accountable for their decisions could reduce ethnically biased decisions.

❖ **Interventions**

- Unpublished work suggests that training teachers to base their decisions on predefined rules and giving them feedback on the extent to which they follow these rules, reduces ethnic biases in teacher track recommendations.
- Interventions designed to reduce educational inequalities are most often geared towards students instead of teachers. They find that changing the psychological meanings of educational settings can be effective to help minority students feel more included, be more resilient in the face of academic difficulties, and to perform better. These studies shed light on how teachers could help foster minority students' performance and reduce inequality in educational outcomes.

❖ **Implications for the Dutch context**

- Teachers should become more aware of inequalities in education, and of their potential role in creating (or reducing) them. Teacher training programmes would be natural environments to do so for new teachers.
- The current obligation of schools to reconsider their recommendation if the standardized test would lead to a higher-level recommendation, requires more guidance for schools how to take up the reconsideration.
- While there is a lot of standardized information available to schools across the primary school years, more research would be needed on the question whether using standardized information assessed at a younger age is associated to weaker or stronger inequalities by socioeconomic background, migration background, and gender.
- Newly developed data sources to inform schools about the future school career of their former students can be used to systematically evaluate a school's accuracy of track recommendations.

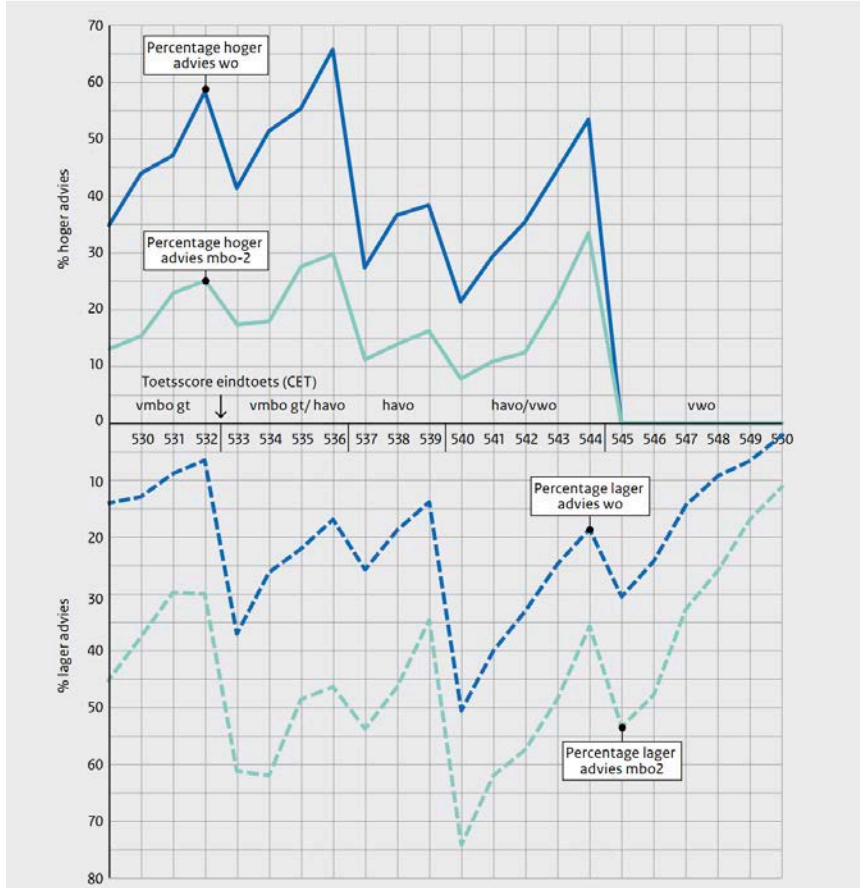
1. Introduction

In 2016, the Dutch Inspectorate of Education presented an alarming report about social inequality in teacher track recommendations (Inspectie van het Onderwijs, 2016). Pupils from high socioeconomic backgrounds were found to be placed into higher ability tracks than their peers from lower socioeconomic background with the same score on the school leaver's test (i.e., Eindtoets Basisonderwijs). Moreover, the findings suggested that this inequality had risen since 2015; the year in which teachers started to formulate their track recommendations *before* students obtained their score on the school leaver's test, instead of *after*. The report led to heated debates in the Netherlands, and a call for more research on the sources of this inequality.

Two years later, teacher track recommendations seem more in line with students' score on the school leaver's test, yet inequality in teacher track recommendations is still considerable (see figure 1) (Inspectie van het Onderwijs, 2018a). Moreover, there are large variations between schools in the extent to which teacher track recommendations match students' score on the school leaver's test.

In this review study, we present an overview of theories and empirical findings on the sources of inequality in teacher track recommendations, expectations, and judgements. As this inequality is not confined to the Netherlands, we review both national and international research. We primarily focus on quantitative work and start with a section on different quantitative methods. Subsequently, we discuss student (chapter 3) and teacher (chapter 4) traits that may affect teacher evaluation. In the fifth and sixth chapter, we elaborate on how the larger social context (i.e., composition of the school class and the institutional setting) may influence teacher expectations and judgements. In chapter 7, preliminary findings on interventions are reviewed. Finally, in chapter 8, we reflect on implications for the Netherlands.

Figure 1: Percentage of students who obtain either a higher or a lower track recommendation than expected on the basis of their score on the school leaver’s test, for students whose parents obtained a college degree (i.e., HBO master or University degree), and students whose parents obtained an upper secondary degree (i.e., havo, vwo or MBO-2).



Source: Inspectie van het Onderwijs, 2018a

In this review study, we distinguish between teacher expectations, judgements, and track recommendations. While teacher expectations refer to a teacher’s predictions about a student’s future performance and possible academic progression, judgements refer to a teacher’s estimate of a student’s current performance (Rubie-Davies, Peterson, Sibley, & Rosenthal, 2015). Track recommendations refer to a teacher’s placement of students into ability groups or special programs. We use the generic term ‘teacher evaluations’ (of students) when referring to a combination of these perceptions and predictions.

2. Methodological considerations

Most quantitative studies fall into one of two methodological categories: experimental or observational studies. Each study should be evaluated with the strengths and weaknesses of their methodological design in mind. Experiments present the advantage of identifying causal effects. By restraining the influence of external factors and randomly assigning participants to different experimental conditions, researchers can isolate the factor of interest and determine whether it directly leads to an outcome. For instance, in the case of teachers' biases, students' performance can be kept identical across experimental conditions to ensure that students' social background is the only variable factor. When the design permits a causal inference, as is often the case with well-conducted experiments, it is considered to possess high *internal validity*.

Results of observational studies often suffer from low internal validity, because many factors intersect in real-life settings to produce an outcome. For instance, a student's background may correlate with his or her school performance, making it difficult to disentangle their separate influence on teacher evaluations. Moreover, teachers may observe student traits that are not (accurately) observed by the researcher, but that do influence the judgements of the teacher. For example, teachers may give higher track recommendations to diligent students, yet researchers may be unable to observe this student trait in observational data. If diligence in school is correlated with certain demographic student traits, researchers may erroneously conclude that teachers discriminate. In experimental studies, student information that is observed by the teacher is also observed by the researcher.

Experimental research on teacher evaluations is most common in psychology. Most of these experimental studies rely on vignette experiments (case vignettes) in which teachers are asked to evaluate a hypothetical student whose characteristics are experimentally manipulated (e.g., Glock, Krolak-Schwerdt, & Pit-ten Cate, 2015; Klapproth, Kärchner, & Glock, 2018). The description of the student's characteristics is often presented in written format but may also include photos. In some experimental studies, teachers are asked to evaluate students on the basis of video- or audiotapes (Tenenbaum & Ruck, 2007). Recent studies also use simulated

computer games in which teachers can interact with hypothetical students in a class (Kaiser, Südkamp, & Möller, 2017).

Experiments tend to rely on small samples and can come at the cost of low *external validity*, which is the extent to which the results can be generalized to the larger population and/or other settings (e.g., everyday life situations). A teacher's behaviour in an experiment is not a direct equivalent of teachers' actual behaviour in classrooms. Observational studies, on the other hand, are conducted in real-world settings (i.e. actual classrooms) and are therefore more likely to represent existing dynamics in classroom settings. In experimental settings, teachers possess relatively little information about the students and may realize that their decisions are fictitious. Teachers could be more biased in such a setting, because they lack the motivation or the possibility to invest effort into correcting potential biases. Nevertheless, a recent study on teacher track recommendations in Germany and Luxembourg has demonstrated the ecological validity of case vignettes (Krolak-Schwerdt, Hörstermann, Glock, & Böhmer, 2017). This study shows that the distribution of the tracks that teachers recommend, as well as the student factors that predict these recommendations, are similar for a sample of actual students and a set of hypothetical students.

3. Student traits and teacher expectations

3.1 How 'accurate' are teachers?

Before discussing why teacher evaluations may be distinct for students from different demographic groups, it is important to shed light on how '(in)accurate' teacher evaluations actually are (Jussim & Harber, 2005). Accuracy is defined as the extent to which teacher evaluations *reflect* social reality, without influencing it (Jussim, Eccles, & Madon, 1996). For example, a teacher is accurate when his or her expectation of a student's test score matches the objective ability of a student to score high or low on this test. This implies that the student's test score is not affected by the expectations of the teacher.

Teacher accuracy is often measured in observational studies by a simple correlation between a teacher's judgement of a student's academic achievement and his or her actual

performance on a standardized test (Südkamp, Kaiser, & Möller, 2012). This correlation tends to be quite high. For example, a meta-study found an overall mean correlation of 0.63 across 75 studies (Südkamp et al., 2012). However, a correlation merely indicates the extent to which teachers are able to put their students in a correct rank order, and not the extent to which teachers over- or under-estimate their students (i.e., the correlation will also be high when teachers systematically over- or underestimate *all* of their students) (Südkamp et al., 2012). Moreover, a correlation of 0.63 still implies that about 60 percent of the variation in teacher expectations remains unexplained.

Most importantly, a high average correlation between teacher evaluations and a student's actual performance can also be due to the fact that a student's performance is influenced by a teacher's earlier evaluations (Jussim & Harber, 2005). Because of this, some studies explicitly examine whether teacher expectations predict, but do not cause student outcomes. These studies either make use of longitudinal research designs (i.e., they account for the prior performance of a student), or they compare the simple correlation between teacher expectations and student performance to the average effect that teacher expectations tend to have on student performance in experimental studies. These studies conclude that teacher expectations are generally quite accurate (i.e., about 75% is accurate). However, this does not preclude that teacher expectations can also be a powerful causal predictor of students' performance (i.e., self-fulfilling prophecy effects), particularly for disadvantaged students whose performance tends to be influenced most by teacher expectations (Jussim, 2018).

A few studies specifically focus on the accuracy of teacher track recommendations in the Netherlands and indicate that teacher track recommendation tend to correspond well with a student's actual performance in school. For example, a cross-sectional study on 7550 students in the final grade of primary school in the school year of 2004/2005, indicates that 80% of the variance in teacher track recommendations is explained by a student's scores on the school leaver's test (e.g., CITO Eindtoets) and tests from school monitoring systems in math, language, and reading comprehension (Timmermans, Kuyper, & Werf, 2015). However, as the study is cross-sectional, the possible influence of teacher expectations on students' test scores is not accounted for. Moreover, the data for this study was collected in the school year of 2004/2005.

In that year, teachers still formulated their track recommendations *after* they received a student's score on the school leaver's tests, a practice that has changed in the Netherlands in 2015. Hence, the correspondence between teacher track recommendations and students' test scores may currently be smaller. A recent study examining track recommendations in the Netherlands over a 20-year period confirms that predictors of teachers' track recommendation are time and context dependent (Timmermans, de Boer, Amsing, van der Werf, 2018).

The Dutch Inspectorate of Education reports that, in 2017, only 37.8 percent of the Dutch students received a track recommendation that matched their score on the school leaver's test (Inspectie van het Onderwijs, 2018b). However, the Dutch Inspectorate of Education does not examine whether deviations between a student's track recommendation and his or her score on the school leaver's test is explained by a student's performance on other standardized tests. Moreover, a mismatch may be due to random variation, rather a systematic over- or underestimation of (particular groups of) students (Timmermans et al., 2015).

3.2 Why are teachers inaccurate? Theoretical considerations

The abovementioned evidence suggests that teachers primarily base their evaluations of students on achievement related-information, but their judgements and expectations do not always fully reflect student performance. Multiple theoretical accounts, from different disciplines, provide reasons for why this might be the case and suggest that teacher judgements and expectations vary as a function of students' demographic groups.

3.2.1. Discrimination in sociology: Statistical vs. taste-based discrimination

In the field of economics and sociology, the literature focuses on the concepts of statistical and taste-based discrimination. Statistical discrimination refers to evaluations of individuals based on the *average* characteristics of the social group to which they belong (Becker, 2010). For example, teachers may hold higher expectations for students from higher socioeconomic backgrounds, because on average these students *are* truly more successful in school. Statistical discrimination is often seen as rational. For instance, when teachers are

unsure and lack information about the actual educational potential of a child, they may rely on average group traits that can serve as a proxy for the child's potential. While statistical discrimination could lead to more 'accurate' decisions at a group-level (i.e., on average certain students *are* also less or more successful in school), they can lead to inaccurate decisions for specific individuals that belong to the group.

Taste-based discrimination implies that people evaluate and treat individuals belonging to different groups in distinct ways, because they rely on prejudices and/or they prefer to interact with the members of some groups over others. Hence, taste-based discrimination reflects a true bias. In contrast to statistical-based discrimination, taste-based discrimination is not rational and can be costly. For example, female teachers may prefer female students, leading to inaccurately high track recommendations for girls.

3.2.2. Evaluations based on individual students

Inequality in the judgements, expectations, and track recommendations of teachers do not necessarily stem from the fact that teachers evaluate students on the basis of the group to which they belong (i.e., discrimination). Possibly they are also caused by the fact that teachers evaluate students on basis of individual traits that are conducive to school success, yet that are also correlated to demographic student traits. For example, teachers may give higher track recommendations to students who are punctual, plan better, and show responsibility for their own learning. These students may in turn happen to come from higher socioeconomic backgrounds (Boone & Van Houtte, 2013). Similarly, teachers may evaluate a student on the basis of characteristics of the home environment that are conducive to school success, and are more pronounced among specific social groups, such as parental aspirations, help and support with school.

A complicating factor is that the aspects that are conducive of school success are not necessarily 'objective' criteria of evaluation, as they could be unrelated to a child's cognitive competencies. This is quite apparent for aspects of the home environment but may also apply to characteristics of the students themselves.

Sociological work on cultural capital and the cultural (or social) reproduction hypothesis is informative in this respect. According to this approach, the cultural resources that students and their parents possess (i.e., cultural capital) positively influence students' success in school (Bourdieu & Passeron, 1977). While the notion of cultural capital is somewhat imprecise (Jæger, 2011; Lamont & Lareau, 1988), it generally refers to people's acquaintance with the codes of the culture of the dominant class. In other words, it pertains to "institutionalized high-status cultural signals", such as behaviours and mannerisms (e.g., manners of speech), preferences (e.g., music tastes), and attitudes (Lamont & Lareau, 1988). These signals constitute a way for people from the dominant class to distinguish oneself from those who do not belong to it.

Cultural capital affects children's school experiences, but is acquired at home: parents either purposefully transmit it to their children, or they subconsciously expose their children to it (Jæger, 2011). The cultural capital of higher social class children is more aligned to schools' cultural implicit expectations than the cultural capital of lower social class children (Stephens, Markus, & Phillips, 2014). Consequently, higher social class children feel more at ease in school, encouraged by their teachers, and "at their place" (Croizet, Goudeau, Marot, & Millet, 2017). Lower social class students, on the other hand, face additional challenges. They need to identify, without being explicitly taught, the expectations of the school and the teachers. This means that some students have to learn which behaviours are (not) valued by their teachers (e.g. research shows for instance that teachers prefer students who ask for help, ask questions, or take responsibility for their behaviour; Calarco, 2011; Dompnier, Dubois & Pansu, 2006), while other students have been taught to behave intuitively in ways that are seen positively in school. This cultural congruence between a child's cultural capital and the school's implicit rule of behaviours may lead to better performance. Indeed, several experiments testing the effect of cultural congruence in education find that low-SES students perform worse on the same test in conditions of cultural mismatch than in conditions of cultural match (Stephens, Fryberg, Markus, Johnson, & Covarrubias, 2012)). As performance in school is perceived as an objective metric of children's abilities, it is then used to form "accurate" recommendations (i.e. based on current performance). However, there are reasons to doubt that children whose home environment allowed them to produce better performance are necessarily more able than

those who didn't. In which case, teacher recommendations could be technically accurate but nonetheless contribute to creating inequalities by offering better learning opportunities to children who possess more cultural and material resources.

In sum, cultural capital tends to be directly rewarded in school through the criteria that schools apply when they evaluate students. Is this as a form of discrimination? On the one hand, it reflects a true bias (i.e., taste-based discrimination), since students from specific social groups (i.e., the dominant social class) are evaluated more favourably in school, for reasons that are not related to their actual cognitive competencies (Jæger, 2011). On the other hand, a student's cultural capital will *be* conducive of a student's school success and competence, because it *is* in effect rewarded in the school system. Hence, it can be argued that teachers do not discriminate, but rely on *relevant* individual student traits, when they base their evaluations on a student's cultural resources. This shows that there is an additional blurry boundary between what 'objective' and 'biased' evaluations are. Moreover, even if relying on those traits might not be considered to be discrimination as such, it does nonetheless mean that schools indirectly contribute to the perpetuation of social inequalities, by offering more rewards and opportunities to students who are initially better equipped for school.

3.2.3 Conclusion

To sum up, there are three main processes that provide explanations for inequality in teacher track recommendations, expectations, and judgements. First, teachers discriminate by evaluating students on the basis of the social group to which they belong. Second, teachers base their evaluations on personal characteristics of the student that are conducive of school success, but that are also correlated to demographic student traits. Third, teachers rely on characteristics of a student's home environment when forming their evaluations of students, which also happen to be correlated to a student's demographic traits. However, it is sometimes difficult to distinguish between discrimination and the other two processes. Traits of the student or his/her home environment that are conducive of school success, may in fact not be related to a child's cognitive abilities. Hence favouring students on the basis of these traits is also a form of bias.

3.3 Empirical Findings

3.3.1 Student demographic traits

Inaccurate teacher evaluations are biased when they pertain to the systematic errors in teacher evaluations, either for all students, or for specific (groups of) students (Timmermans et al., 2015). Research shows that teachers tend to make systematic errors in their evaluations of (specific groups of) students. This means that demographic traits influence teacher evaluations, above and beyond student performance (and motivation). In this section we discuss how teacher track recommendations, expectations, and judgements with respect to a student's academic performance or ability level are influenced by student gender, socioeconomic background, and race and/or ethnicity. We leave out research on teacher expectations or judgements about a student's behaviour or attitudes. Table 1 presents an overview of the findings. In the table we also include studies that include demographic traits as control variables, yet do not specifically focus on them.

Findings are most consistent with respect to a student's socioeconomic status. With a few exceptions, research shows that teachers evaluate students from high socioeconomic backgrounds more positively than students from low socioeconomic backgrounds, also in the Netherlands (Smeets, van Kuijk, & Driessen, 2014). A Dutch observational study underlines the consistency in teachers' unfavourable evaluations of students from low socioeconomic backgrounds by showing that *all* of the 500 teachers that participated in the study gave lower track recommendations to students from lower socioeconomic backgrounds, irrespective of students' academic performance (Timmermans et al., 2015). Moreover, an observational study using information on nine Dutch cohorts suggests that socio-economic inequality in teacher track recommendation has been relatively stable between 1995 and 2014 (Timmermans, et al., 2018), yet other findings suggest that this inequality has even increased over time (Inspectie van het Onderwijs, 2016). Interestingly, the few studies that find no relationship between a student's socioeconomic status and teacher evaluations are all conducted outside of Europe, in countries such as the United States (Irizarry, 2015) or Chili (Mizala, Martínez, & Martínez, 2015).

While findings are consistent, differences in teachers' evaluations of students from different socioeconomic backgrounds are generally small. Jungbluth (2003) is even able to fully account for the socioeconomic inequality in teacher track recommendations in the Netherlands in the school year of 2000-2001. More specifically, he shows that this inequality disappears when accounting for students' factual performance, as well as teachers' perceptions of students' capacity, work attitudes, and social behaviour. However, it is important to note that there may be social biases in teacher perceptions of (the capacity of) students. Hence, differences in teacher track recommendations for students from different socioeconomic backgrounds may have persisted if the multivariate regression models would have only accounted for factual performance¹ and not for subjective perceptions of student' capacity. In line with this, Jungbluth (2003, 2014) observes that socioeconomic inequality in teacher track recommendations is larger than socioeconomic inequality in students' score on the school leaver's test, and that both are larger than socioeconomic inequality in students' intelligence scores.

Experimental studies also reveal consistent socio-economic inequality in teacher evaluations. When students' information is manipulated to represent identical performance, higher SES students still tend to be favoured as compared to lower-SES students. For instance, two experiments in Switzerland were designed to simulate actual tracking dilemmas that can occur in the Swiss system. Teachers and the principal can offer a second chance to pupils who are borderline cases for the higher track (i.e., slightly below official standards). The results reveal that this opportunity is more readily offered by teachers (and students playing the role of teachers) to the high-SES student than to the low-SES student despite a same prior performance (Batruch, Autin, Bataillard, & Butera, 2018). Similarly, in a study in the United States, 103 middle school counsellors are asked to rate the academic potential and future expectations of a problematic student whose SES is manipulated. Results indicate that the high-SES student is perceived as more academically able and academically promising compared to the low-SES student (Auwarte & Aruguete, 2008).

¹ For this reason, Jungbluth's (2003) findings are not presented in table 1

The findings of different studies are relatively inconsistent when considering teacher biases with respect to a student's ethnicity or race. While some studies find that teachers evaluate ethnic minority students more negatively, other studies find no statistically significant effect, or even a positive one. These inconsistent findings may be explained by the country in which the study is set. For example, a meta-study on experimental research on ethnic teacher biases in the United States indicates that teachers hold lower expectations for African American or Latino/a students than for European Americans, and that they are less likely to refer them to gifted programs, and more likely to make referrals for special education or disciplinary action (Tenenbaum & Ruck, 2007). Nevertheless, the average effect sizes of these ethnic biases are small.

Differences in research findings may also be due to the research design. It seems that experimental studies either tend to show that teachers evaluate typically disadvantaged minority groups more negatively or report mixed findings. Notably, several experiments conducted on German or Luxembourgian pre- and in-service teachers find that judgements, expectations, and recommendations are lower for students with ethnic minority background than they are for students from an ethnic majority background (Glock, Krolack-Schwerdt, Klapproth, & Böhmer, 2013; Glock & Krolack-Schwerdt, 2016; Holder & Kessel, 2017; Klapproth, et al., 2018 ; Tobisch & Dresel, 2017). An observational study in New Zealand also finds that teachers' overall academic judgements - which are used for deciding ability grouping - are lower for ethnic minority groups, after accounting for student performance (Meissel, Meyer, Yao, & Rubie-Davies, 2017). However, in other observational studies in which student performance is accounted for, teachers seem to have more positive perceptions of students from disadvantaged minority groups than students from the native majority. This even pertains to observational research in the United States (Gershenson, Holt, & Papageorge, 2016).

Findings may also be inconsistent, because teachers tend to sometimes over- and other times underestimate students from ethnic minorities, leading to positive, negative, as well as null findings. In line with this explanation, a study in Germany and Luxembourg indicates that track recommendations for students from ethnic minority groups are more likely to be 'inaccurate' than those for students from the ethnic majority (Glock et al., 2015). This implies

that students from ethnic minority groups are more likely to receive higher as well as lower track recommendations than would be expected on the basis of their academic profile. Possibly, teachers overestimate some and underestimate other ethnic minority students. For example, a study by Irizarry (2015) in the United States indicates that teachers judge low performing minority students more favourably than low performing majority students, while they judge high performing majority students more favourably than high performing minority students. However, other studies find that teacher expectations are actually more accurate for students from ethnic minority groups (Kaiser et al., 2017; Tobisch & Dresel, 2017), possibly because teachers tend to overestimate students from the ethnic majority (Tobisch & Dresel, 2017).

Findings with respect to teachers' ethnic biases could also be inconsistent because there are large variations *across* teachers in how they evaluate students from ethnic minority groups as compared to students from the native majority. In line with this, a Dutch study indicates that some teachers tend to give *higher* track recommendations to Turkish, Moroccan, and other foreign students from low socioeconomic backgrounds, while other teachers tend to give *lower* track recommendations to these minority groups (Timmermans et al., 2015). These effects cancel each other out, leading to a non-significant overall effect of student ethnicity on teacher track recommendation.

Finally, differences in findings may be due to differences in the year the study was conducted. In the Netherlands, ethnic biases in teacher track recommendations have changed between 1995 and 2014 (Timmermans et al., 2018). While in 1995 teachers tended to give higher track recommendations to students from ethnic minority groups than their equally performing peers from ethnic majority groups, this difference reduced over time, and eventually disappeared. The authors note that this might be due to the fact that equity-related policies stopped focusing on the position of minority groups or possibly because Dutch society became less tolerant towards members of minority groups over this period.

Of the 18 studies considering gender biases in teacher evaluations, 12 suggest that teachers evaluate girls more positively than boys. Especially studies in the Netherlands seem to find consistent gender biases in teacher track recommendations. However, Timmermans et al.



(2015) show that there is variation in the gender bias across teachers. While overall girls tend to receive higher track recommendations in the Netherlands, some recommendations seem to slightly favour boys in the Netherlands. Moreover, the female advantage in teacher track recommendations in the Netherlands has declined between 1995-2014 (Timmermans et al., 2018). In 2014, teachers did not seem to give higher track recommendations to female students as compared to their equally performing male counterparts.

Table 1: overview of study outcomes on the influence of student demographic traits on teacher evaluations, accounting for student performance

Outcome	Demographic traits		
	High(er) socioeconomic status	Ethnically or racially disadvantaged groups	Girl
Track recommendations and referrals to special education or gifted programs	<p>Positive Batruch et al., 2018 (CH) ^e Barg, 2012 (FR) ^o Becker, 2013 (DE) ^o Boone, et al. 2018 (BE) ^o Boone & Van Houtte, 2013 (BE) ^o Dauber, et al., 1996 (US) ^o De Boer et al., 2010 (NL) ^o Driessen, et al., 2008 (NL) ^o Kelly, 2004 (US) ^o Podell & Soodak, 1993 (US) ^e Timmermans et al., 2015 (NL) ^o Timmermans, et al., 2016 (NL) ^o Timmermans, et al., 2018 (NL) ^o</p>	<p>Positive Barg, 2012 (FR) ^o Driessen et al., 2008 (NL) ^o</p> <p>Negative Boone et al., 2018 (BE) ^o Dauber et al., 1996 (US) ^o Glock et al., 2013 (LX) ^e Klapproth, et al., 2018 (DE) ^e Tenenbaum & Ruck, 2007 (US) ^{e*} Riley & Ungerleider, 2008 (CA) ^e</p> <p>Mixed Elhoweris et al., 2005 (US) ^e Glock et al., 2015 (GE; LUX) ^e Timmermans, et al., 2018 (NL) ^o</p> <p>N.S. Boone & Van Houtte, 2013 (BE) ^o De Boer et al., 2010 (NL) ^o Timmermans et al., 2015 (NL) ^o Boone & Van Houtte, 2013 (BE) ^o</p>	<p>Positive De Boer et al., 2010 (NL) ^o Driessen et al., 2008 (NL) ^o Kelly, 2004 (US) ^o Timmermans et al., 2015 (NL) ^o Timmermans et al., 2016 (NL) ^o</p> <p>Mixed Dauber et al., 1996 (US) ^o Timmermans, et al., 2018 (NL) ^o</p> <p>N.S. Boone & Van Houtte, 2013 (BE) ^o Boone et al., 2018 (BE) ^o</p>
Expectations of student performance	<p>Positive Auwarter & Arguete, 2008 (US) ^e Gershenson et al. 2016 (US) ^o Tobisch & Dresel, 2017 (DE) ^e</p> <p>N.S. Mizala et al., 2015 (CL) ^e</p>	<p>Negative McKown & Weinstein, 2008 (US) ^o Sebastian Cherng, 2017 (US) ^o Tenenbaum & Ruck, 2007 (US) ^{e*} Tobisch & Dresel, 2017 (DE) ^e Turner et al., 2015 (NZ) ^o</p> <p>Positive Gershenson et al. 2016 (US) ^o</p>	<p>Positive Gershenson et al. 2016 (US) ^o</p> <p>Negative Mizala et al., 2015 (CL) ^e</p>

Judgements of student performance or ability

Positive

Alvidrez & Weinstein, 1999 (US) °
Ready & Wright, 2011 (US) °
Tobisch & Dresel, 2017 (DE) °

N.S.

Irizarry, 2015 (US) °
Hinnant, et al., 2009 (US) °

Negative

Glock & Krolack-Schwerdt, 2016 (DE) °
Holder & Kessel, 2017 (DE) °
Meissel, et al., 2017 (NZ) °
Ready & Wright, 2011 (US) °
Rubie-Davies et al., 2015 (NZ) °
Sebastian Cherg 2017 (US) °
Tobisch & Dresel, 2017 (DE) °

Mixed

Glock, 2016 (DE) °
Irizarry, 2015 (US) °

N.S.

Kaiser et al., 2017 (DE) °
Hinnant, et al., 2009 (US) °

Positive

Hinnant, et al., 2009 (US) °
Irizarry, 2015 (US) °
Martinez et al., 2009 (US) °
Meissel, et al., 2017 (NZ) °
Ready & Wright, 2011 (US) °
Tournaki & Podell, 2005 (US) °

Negative

Holder & Kessel, 2017 (DE) °

Notes: The country of the study is indicated in brackets. N.S. refers to not statistically significant
° observational data; ° experimental data; °* meta-study that only relied on experimental data

3.3.2 Student behaviour and attitudes

Teacher evaluations of students are influenced by a student's school behaviour and attitudes, or a teacher's perception thereof. Research has specifically focused on the role of a student's motivation and effort, as these are sometimes perceived to be 'legitimate' criteria for evaluation. This means that some scholars define teacher accuracy as the extent to which teacher expectations, judgements, or track recommendations are explained by student performance, ability, as well as motivation (e.g., Becker, 2013; Timmermans et al., 2015).

Research in Luxembourg and Germany indicates that teacher perceptions of a student's work behaviour are positively related to teacher track recommendation (Krolak-Schwerdt et al., 2017). This relationship exists irrespective of a student's school performance (i.e., grades, and in Luxembourg also test scores) and nationality. Dutch research also shows that a student's school behaviour and attitudes are related to teacher track recommendations, although a student's cognitive competencies are a much stronger predictor (Smeets et al., 2014). More specifically, another study using cross-sectional survey data among 5,664 Dutch students in their final year of primary school, shows that teacher track recommendations are slightly higher for students who are perceived to be more self-confident and to have better study attitudes by their teachers (Driessen, Slegers, & Smit, 2008). Teachers also give higher track recommendations to students who report that they put more effort into school. Surprisingly, teacher track recommendations are lower for students who are perceived to exhibit more socially accepted behaviour. These relationships exist while controlling for students' cognitive competences (e.g., intelligence and test scores) and background characteristics², yet only explain an additional 1.5 percent of the variance in teacher track recommendation on top of student performance and demographic traits. It should be noted that this study builds on data

² Although not an explicit focus of the study, the findings suggest that (teacher perceptions of) student behaviour and attitudes do not explain teachers' slightly higher track recommendations for students from higher social backgrounds. After non-cognitive student traits are accounted for, the higher track recommendations for students from ethnic minority groups seem to be slightly higher. However, teacher perceptions of a student's Dutch language skills are also included as a non-cognitive student trait in this model.

that was collected more than 15 years ago, in the school year of 2002/2003, and track recommendation procedures might have changed in the meanwhile.

Differences in students' behaviour and attitudes may (partly) explain differences in teacher expectations along demographic lines. In other words, teachers may evaluate students on the basis of behaviours or attitudes that are related to a student's social class, thereby causing inequality in teacher track recommendations (Boone & Van Houtte, 2013). Using data on more than 390 Flemish teachers, Boone & Van Houtte (2013) show that most teachers (i.e., 85 percent) do not explicitly involve a pupil's social background in their track recommendation. However, 69 percent of the teachers indicate that they take into account student behaviour. To shed more light on the social inequality in teacher track recommendations in Flanders, Boone and Van Houtte (2013) conduct focus groups with a total of seven teachers who participated in the survey. In these focus groups, all teachers mention that they pay attention to pupil qualities other than achievement, such as independence, planning capacity, responsibility, and punctuality. Boone & Van Houtte (2013) suggest that these qualities are more prevalent among students from higher socioeconomic backgrounds, leading to social disparities in track recommendations.

However, findings of a study among 5316 Dutch students who are in their final year in primary school in the school year 2004/2005 are not supportive of these conclusions (Timmermans et al., 2016). While accounting for a student's standardized test scores (i.e., scores on the school leaver test, language test, mathematics test, and reading comprehension test), this study shows teachers tend to give somewhat higher track recommendations to students whom they perceive as self-confident (e.g., a student does not panic quickly) and having positive work habits (e.g., a student works hard); and slightly lower track recommendations to students whom they perceive as showing good social behaviour in class (e.g., a student sticks to the class rules). A teacher's perceptions of a student's popularity and his or her relationship with a student are generally not predictive of his or her track recommendations. Most importantly, teacher perceptions of student attitudes and behaviours do not explain why students from higher social backgrounds receive higher track recommendations, yet partly explain why teachers give higher track recommendations to girls

in comparison to boys. Possibly, there are true gender differences in students' school attitudes and behaviour, but teachers' perceptions of student behaviours and attitudes may also be biased in favour of girls. Nevertheless, teacher perceptions of student attitudes and behaviour only explain 3% of the variance in teacher track recommendations that is not explained for by student performance. The findings also suggest that teachers weigh their perceptions of student attitudes and behaviours differently in their track recommendations, meaning that the track recommendations of some teachers are more positively or negatively affected by student attitudes and behaviours than those of others.

The relationship between a student's performance and/or demographic traits and the evaluations of a teacher may also be *dependent* on a student's behaviour or attitudes. Timmermans et al. (2016) show that the impact of a student's performance on a teacher's track recommendation is relatively independent of a teacher's perception of a student's self-confidence, popularity, and social behaviour. However, a student's performance seems to have a stronger impact on a teacher's track recommendation when the student is perceived to have more positive work habits, and a weaker impact when a teacher has a more positive perception of his/her relationship with the student.

Klapproth, Kärchner, & Glock (2018) study the extent to which the effect of a student's ethnic background on a teacher's track recommendation is dependent on student absenteeism. High student absence rates among ethnic minority students may confirm the stereotype of ethnic minority students as poor academic performers, leading to an activation of the ethnic stereotype, and higher ethnic discrepancies in teacher track recommendations. To test this idea, the authors conduct a vignette experiment among 95 preservice teachers in Germany. In the vignette experiments respondents are asked whether or not they are in favour of placing a hypothetical male student in the highest secondary school track. The GPA, ethnicity (i.e., Turkish or German), and absence rate of the hypothetical student are experimentally manipulated. The authors do not find clear support for their hypothesis. Overall, students with a higher GPA, with a German background, and with low absence rates are more likely to be assigned to the highest track. In line with the hypothesis, higher absence rates are related to a greater reduction in the likelihood of a high track-recommendation for Turkish students with a

high GPA than for German student with a high GPA. However, among students with a low or medium GPA, high absence rates only decrease the probability of a high track recommendation for German, but not for Turkish students.

3.3.3 Home situation

Teachers seem to hold higher expectations for students who receive more school-related support and help in the family. A study among 56 German primary school teachers indicates that teacher track recommendations are higher when parents provide their children more support in problems that occur in school (Krolak-Schwerdt et al., 2017). Similarly, observational research among Dutch students and their teachers in the school year of 2003/2004, indicates that teachers provide higher track recommendations to students who live in educationally supportive homes in which learning and curiosity are stimulated (Driessen et al., 2008). Nevertheless, teachers themselves do indicate that they generally base their track recommendations more on the cognitive and social-emotional competencies of the child than on the support that students receive at home (Driessen, 2006).

The fact that teachers may take into account parental involvement and support in their tracking decisions could lead to social disparities in track recommendations. Students from higher social backgrounds may receive more school-related help and support from their parents, partly because their parents are better equipped to help. While some empirical work seems to be supportive of this idea (e.g., Bonizzoni, Romito, & Cavallo, 2016; Grolnick et al., 1997; Lareau, 2015;), other findings suggest that this relationship is not self-evident (e.g., Driessen, 2006; Bakker, Denessen, & Brus Laeven, 2007; Sui-Chu & Willms, 1996) . For example, Driessen (2006) shows that the educational background of parents only plays a marginal role in students' perceptions of their parents' interest in school, help with school, and school expectations in the Netherlands. Dutch students whose parents belong to the group with the lowest educational credentials even perceive their parents to have the highest educational expectations.

Some scholars suggest that parents of different socioeconomic backgrounds are involved in school in distinct ways (Bakker et al., 2007). While parents with a lower

socioeconomic background may be more involved in school activities in the home, parents with a higher socioeconomic background could be more involved in activities at school (e.g., show interest and encouragement with respect to school-related matters), such as parent-teacher meetings and volunteering. Research in the United States for example suggests that parents with higher educational credentials are more likely to have contact with school, volunteer in school, and to be active in Parent-Teacher organization activities (Kelly, 2004). Similarly, American teachers rate their relationship with White and Hispanics parents and their children more positively than their relationship with African-American parents and children. This in turn predicts teachers' perception of children's abilities, even when controlling for parental education and children's measured abilities (Hughes, Gleason, & Zhang, 2005). Hence, even when there are no demographic differences in the *amount* of parental involvement in school, teachers may still *perceive* a social or ethnic gradient in parental involvement (Bakker et al., 2007), partly because they base their evaluations of parents' involvement in school by their involvement in activities at school, rather than activities at home.

Research in the Netherlands indicates that teacher perceptions, rather than parents' own perceptions of their involvement, vary according to parents' socioeconomic background (Bakker et al., 2007). More specifically, parents with a higher socioeconomic status perceive their participation in school activities to be higher, yet there are no differences by social status in parents' own perceptions of their contact with teachers, influence on school (policies), educational involvement in the home, or the extent to which they are informed about the school. In contrast to this, teachers perceive parents from higher socioeconomic backgrounds to be more involved on all these dimensions, except for the extent to which parents are informed. This social gradient in teachers' perceptions of parental involvement in school may also (partly) explain the social gradient in teachers' expectations and track recommendations, yet we are not aware of any empirical research on this.

Children from higher socioeconomic backgrounds may not only (be perceived to) receive more help and support in school, but their parents are possibly also better equipped with cultural resources that ease the interaction with teachers and principals (i.e., a form of cultural capital) (Barg, 2012; Lareau, 2015). More specifically, parents with a higher

socioeconomic status are believed to communicate more comfortably and effectively with school, because they possess the linguistic abilities to do so, have greater knowledge about the educational system, and their habits and values tend to match those of the school (Barg, 2012; Lareau, 2015). This also makes them more likely to question the decisions of the school and to (successfully) exert (implicit) 'pressure' on the school's track recommendations. Moreover, schools may even expect or assume that parents with a higher socioeconomic status will object to relatively 'low' track recommendations, and may try to avoid such objections by simply giving higher track recommendations to students from higher social backgrounds (Barg, 2012).

Empirical findings with respect to parental influence is still ambiguous. A large-scale quantitative study among more than 13,000 American students suggests that parents with a higher socioeconomic status are more involved in school and intervene more in track placement decisions (Kelly, 2004). However, this does not seem to explain why students from higher class backgrounds are placed into higher tracks in mathematics than their equally performing counterparts with a lower class background.

A study by Driessen (2006) among Dutch teachers suggests that the socio- or ethnic composition of the school is not related to the level of pressure on track recommendations that schools experience. Nevertheless, teachers do indicate that they take into account the wishes and opinions of parents (and children) when they formulate track recommendations. Moreover, they seem to be more likely to do so when parents are born in the Netherlands and are highly educated. Similarly, in schools with a larger proportion of disadvantaged students with a migration background, track recommendations tend to be more fixed, and teachers are less likely to formulate their recommendations together with parents. However, and surprisingly, in schools with a larger proportion of disadvantaged students from the native majority, the involvement of parents in the track recommendation procedure tends to be larger.

Research among 11,000 students in France indicates that differences in teacher track recommendations by social background are heavily reduced when accounting for the school track requests by families (Barg, 2012). It should be noted that in France, schools explicitly take into account parental wishes in their track recommendations (Barg, 2012). In a first stage,

parents are asked to request a track for their child; subsequently, the school staff recommends a track; and, finally, the family can reject the school's recommendation. While differences in track requests by social background seem to contribute to disparities in teacher track recommendations by social background, they do not fully explain them. The study also shows that teacher track recommendations are higher for students whose parents are involved in parent associations, yet lower for students whose parents initiate meetings. This latter finding may be due to reversed causality, as parents may initiate meetings with school when there are problems.

Parents who belong to ethnic minority groups may also face barriers in their communication with school and their involvement in school life, including limited knowledge of the language and educational system, as well as feelings of inferiority and a lack of self-esteem (Bonizzoni, et al., 2016). A qualitative semi-structured interviews among 26 teachers and headmasters of 12 middle schools in Milan supports this idea by indicating that immigrant families are perceived to be less involved in their children's schooling, and to participate less in school activities and meetings, partly because they are constrained by their knowledge of the Italian language and educational system (Bonizzoni et al., 2016). Because of this, they are also less able to challenge and influence teacher track recommendations and are more likely to accept the teacher's track recommendation as an 'expert' decision. Parents from upper and middle-class parents from the native majority are more likely to press for specific track recommendations and to go against teacher decisions.

4. Traits of teachers

4.1. Stereotypes and attitudes

4.1.1 Theoretical considerations

So far, it seems that teachers' distinct judgements and expectations for students from different demographic groups cannot entirely be explained by students' performance, behaviour, attitudes, and/or their home environment. To uncover potential sources of

inequality in educational decisions, research in psychology has focused on teachers' stereotypes and prejudicial attitudes against students of disadvantaged groups.

In the psychological literature, stereotypes, attitudes, prejudice, and discrimination are different theoretical constructs. Prejudices are the negative evaluation of a social group or an individual based on their group membership (Crandall & Schaller, 2005). Prejudice can have an affective component (i.e. like or dislike) as well as a cognitive component (i.e. positive or negative beliefs about the group) (Dovidio, Hewstone, Glick & Esses, 2010). Attitudes refer to the affective component of prejudice, while stereotypes are related to the cognitive component. Stereotypes are thought to be cognitive schemas used by perceivers to process information about other individuals, which manifest as positive or negative beliefs about personal traits or behaviours associated with certain social groups (Al Ramiah, Hewstone, Dovidio, & Penner, 2010). People are prejudiced when they negatively *evaluate* an individual because they dislike the group to which the individual belongs (negative attitudes) or believe that members of the group possess a certain trait (stereotype). When people *treat* others unfairly or unequally on the basis of their group membership, it is considered discrimination.

Debates over the origins of stereotypes and their accuracy remain ongoing. Some researchers contend that stereotypes are grounded in reality and reflect actual differences between groups. In their perspective, applying stereotypes might be inaccurate in individual cases, but stereotypes should be seen as overall probabilistic generalizations, which remain adaptive to changes from the exterior world (Jussim, McCauley & Lee, 1995). Others consider stereotypes as rigid cultural constructions that are the product of a socialization process. In their view, stereotypes fulfil a psychological need to justify unequal differences of resources between groups and are useful for the perceiver to navigate environments efficiently but are often inaccurate or overly restrictive at the individual-level (Fiske, 1998). Importantly, this perspective emphasizes that despite arising from normal cognitive processes (i.e. categorization processes), the consequences of applying stereotypes can be particularly damaging for individuals on the receiving end. They can lead to discrimination (i.e. the enactment of unequal behaviour towards an individual based on their group membership).

Most studies on stereotypes and attitudes can broadly be subdivided in two research lines investigating, respectively, explicit and implicit bias. Explicit stereotypes and attitudes are deliberately and consciously formed and are therefore easy to measure. For this reason, social psychologists have traditionally relied on explicit self-reported measures in questionnaires to study them. More recently, researchers have found that those associations could operate even outside of individual consciousness, leading to the investigation of implicit biases (Fiske, 1998). Implicit stereotypes and attitudes are measured by quantifying the strength of the automatic association between either (1) a social category and particular semantic content, known as an implicit stereotype, or (2) a social category and an evaluation of like /dislike, known as a positive or negative implicit attitude (Greenwald & Krieger, 2006). For instance, if a teacher automatically associates high socioeconomic status children more with 'eloquence' than low socioeconomic status children, then (s)he is holding an implicit stereotype. If teachers automatically evaluate them less favourably than high socioeconomic status children, then (s)he is harbouring negative implicit attitudes. In practice, these associations are measured by calculating the speed of sorting words into different categories associated to social groups. An advantage of measuring implicit stereotypes and attitudes is the possibility to circumvent issues of social desirability to directly access less controllable associations, as most researchers agree that relying exclusively on individuals' self-reported stereotypes and attitudes with explicit measures is limited in context where social norms prohibit public display of discrimination and prejudice.

There is still contention between researchers over how to interpret the meaning of implicit bias measures. There is general agreement over their effects: not all stereotypes or attitudes have to be conscious, and implicit biases also predict behaviour independently of explicit biases. However, other questions pertaining to the differences between the processes underlying implicit and explicit bias and their commonalities are still debated (Dovidio, Hewstone, Glick & Esses, 2010). Most studies on implicit stereotypes and attitudes among teachers rely on the "dual process theories" framework to explain differences of origin between both biases. Dual process theory posits that explicit and implicit biases arise from two distinct mental processes. It distinguishes between (1) automatic mental processes which are fast,

efficient, and unintentional, and (2) deliberate mental processes which are slower, more controlled, and intentional (Forscher et al., 2018). Both automatic and deliberative processes could influence behaviour, but each process is most influential in different contexts.

Deliberative processes are likely to be influential if the individual has sufficient cognitive resources (e.g. are not pressured by time) and a high level of motivation (e.g. they judge someone important to them). In this case, the individual may take more pieces of information into account to form judgements and explicit measures should in theory be more predictive of their behaviour (Glock & Krolak-Schwerdt, 2014). Automatic processes are more likely to arise if people's motivation and cognitive resources are compromised (Devine, 1989; Fazio & Olson, 2014; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). This might occur for instance in the cases of hectic classrooms when teachers may experience cognitive overload and have less opportunity to control their behaviours (Kumar, Karabenick & Burgoon, 2015).

Implicit and explicit measures tend to correlate, but the extent to which they do varies substantially across contexts, ranging from very small correlations ($r=.07$) to strong correlations ($r=.70$), with an overall mean of $r=.24$ (Cameron, Brown-Iannuzzi, & Payne, 2012; Greenwald et al., 2009; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek & Hansen, 2008). When it comes to their relationship to behaviour, meta-analysis find small but significant positive correlations between implicit bias and intergroup behaviour (around $r=.15$ to $.25$) and also between explicit bias and behaviour ($r=.10$ -. $.20$; Talaska, Fiske, & Chaiken, 2008). As most intergroup behaviours are observed in laboratory settings, it is unknown to what extent these small effect sizes cumulate in reality and whether they have small or large societal effects (Oswald, Mitchell, Blanton, Jaccard, & Tetloch, 2013). Moreover, recently some researchers have expressed scepticism over the interpretation of implicit measures in the literature. They suggest that the validity of those measures in explaining real-world phenomena such as individual discrimination might have been overemphasized (Carlsson & Agerström, 2016). Nonetheless, explicit and implicit stereotypes and attitudes are often studied on teacher samples, and researchers assume that they are related to teachers' potential discriminatory behaviour.

4.1.2 Empirical findings

In this section, we'll start by reviewing the current evidence of teachers' or pre-service teachers' explicit stereotypes and attitudes before presenting the scientific literature on teachers' implicit stereotypes and attitudes. Finally, we'll describe a recent study conducted in the Netherlands indicating that implicit attitudes can affect teachers' expectations and students' test scores.

Several studies examine the effects of teachers' explicit stereotypes and attitudes. For instance, Wenz, Olczyk and Lorenz (2016) study gender, ethnicity, and socioeconomic status achievement-related stereotypes among German second-grade teachers. They administered a questionnaire to 52 teachers designed to measure the extent to which they perceived students from these different social groups as good/mediocre students in mathematics and reading. To avoid a social desirability bias, the teachers were told that the researchers wished to compare their perceptions of students' achievement to actual data. The findings indicate that teachers' estimations of groups' achievement follow national trends but are overly negative when it comes to comparisons between native and immigrant students' achievement (particularly Russian students), as well as boys' achievement scores in reading test.

Another study conducted in Switzerland focuses on explicit gender mathematical stereotypes. Keller (2001) tests whether teachers' stereotypes relate to students' mathematic gender stereotypes. Controlling for student-level variables (SES, gender, achievement, interest, self-confidence) and school-level variables (grade, schooling track), the results reveal a small (4.5% explained variance) but significant positive effect of teachers' stereotyping of mathematics as a male domain on students' stereotyping, providing some initial evidence that teachers' stereotyping could affect students' own beliefs.

Focusing on pre-service teachers' social class stereotypes in Germany, Dunkake and Schuchart (2015) find that comparatively to middle-class students, lower-class students are on average perceived as more aggressive, less ambitious, lazier and bothersome, undisciplined, low-performing, and unmotivated. In a second task, participants were exposed to fictitious scenarios of disruptive behaviour to test if the lower-social class students would be treated more harshly in this experiment. The results show that 36.7% of the participants treat children

differently according to their social class. Among participants with weaker stereotypes, 14.1% opt for harsher disciplinary action for lower-class children. Among participants with stronger stereotypes, 26.1% are harsher with lower-class children. More generally, the authors observe that participants prefer to address middle-class children's disruptive behaviour more discreetly. Inversely, they are more likely to choose to reprimand the lower-social class children publicly. In line with this, another study on pre-service teachers in the United States shows that 25% of the teachers endorse explicit stereotypical beliefs about lower-social class and minority students. However, the authors note that throughout the teaching program, pre-service teachers' stereotypical beliefs about lower social class children decreases (Kumar & Hamer, 2013). This result may suggest that experience could be a factor that can change pre-service and by extension in-service teachers' explicit stereotypes. It may also be that with time, pre-service and in-service teachers are socialized to learn not to express those views explicitly. In sum, the current research on explicit measures suggest that some pre-service teachers and in-service teachers explicitly report holding negative stereotypes against disadvantaged students, especially lower-class students.

Research on teachers' and pre-service teachers' implicit biases have mostly been conducted in European settings and focus on implicit attitudes and stereotypes towards three groups that are seen as disadvantaged in educational systems: immigrant (ethnic minority students) vs. native students, girls vs. boys, and students of varying socioeconomic status. A recent study in the Netherlands focuses on both explicit and implicit stereotypes and attitudes towards students with parents of varying educational level (Pit-ten Cate & Glock, 2018). The results suggest that implicit attitudes are more positive towards children of highly educated parents. However, teachers do not hold differential explicit stereotypes or attitudes towards children of high vs. low-educated parents. More specifically, participants neither agree nor disagree with propositions that children's motivation, social behaviour or opportunities in the Netherlands vary as a function of their parents' education level.

Similar procedures have been used to assess explicit and implicit attitudes and stereotypes among German teachers, which reveal more overall negative implicit attitudes towards ethnic minority students, independent of student and teacher gender (Kleen & Glock,

2018). They also find gender differences among the ethnic minority conditions: the implicit attitudes are more negative towards female ethnic minority students than male ethnic minority students. However, when looking at the explicit measures, they find that teachers report more enthusiasm teaching female ethnic minority students as compared to the males. All other explicit scales used in the study reveal that teachers report positive attitudes towards teaching ethnic minority students with no systematic variations between students or teachers' gender. These findings are consistent with the results of, yet another study among teachers in Germany (Glock & Kapproth, 2017). This study indicates that elementary and secondary school teachers have more negative implicit attitudes towards ethnic minority students. Secondary school teachers are implicitly more positive towards boys, whereas elementary school teachers are implicitly more positive towards girls. Concerning the results on explicit attitudes, elementary school teachers are more enthusiastic about teaching ethnic minority boys than girls. Finally, another study on preservice teachers in Germany examines the implicit associations towards racial minority vs. majority students on three different implicit measures to compare the implicit attitudes (Glock & Karbach, 2015). The first two implicit measures show more negative implicit attitudes towards racial minority students than towards racial majority students. The third one reveals that implicit attitudes towards racial majority students are positive, whereas those towards racial minority students are neutral.

In sum, current research on teachers' implicit measures towards students in European context provide an overall consensus that disadvantaged students are the target of more implicit negative attitudes as compared to advantaged students. It appears to be particularly the case for students whose parents are immigrants or low-educated. The pattern for gender biases is less clear. Implicit or explicit stereotypes and attitudes sometimes favour girls, and other times favour boys. The results on the implicit and explicit measures reveal different patterns in most of the studies described above. This suggests that the two underlying processes do not seem to be related in these studies, and/or are unevenly influenced by a social desirability or self-presentation bias. While these results are indirectly suggestive that biases in teachers' decision-making processes or expectations could be the result of negative

implicit associations, they do not present direct evidence for this effect. To the best of our knowledge, only one existing study present a case supporting this hypothesis.

The study in question was conducted in the Netherlands on a population of 41 teachers in 17 elementary schools (Van den Bergh, Denessen, Hornstra, Voeten & Holland, 2010). Similar to other findings, the study shows that the explicit and implicit measures of prejudiced attitudes do not relate to each other. While the explicit measures of the prejudiced attitudes do not predict teachers' expectations for students, implicit prejudiced attitudes do. Teachers' *expectations* are found to be lower for students of Turkish or Moroccan origin than for students of Dutch origin. These ethnic differences in expectations are larger when teachers' negative implicit prejudiced attitudes are higher. Similarly, the authors do not find a relationship between teachers' explicit prejudiced attitudes and students' *test scores*, yet more negative implicit prejudiced teacher attitudes are associated with lower test scores for students of Turkish or Moroccan origin. A maximum difference in the teachers' implicit prejudiced attitudes (comparing least prejudiced to most prejudiced teachers) is related to a difference of 1.09 on the standardized mathematic achievement score between the ethnic minority and majority group students, and a difference of 1.05 for the standardized text comprehension test scores. Furthermore, the authors find that implicit prejudiced attitudes among teachers are related to larger ethnic differences in student test scores via teachers' expectations. The conclusion of this study is that teachers implicit prejudiced attitudes could be important to consider when investigating teachers' expectations of students as well as to understand processes underlying the ethnic achievement gap in school.

Taken together, the studies presented above show that teachers hold more negative implicit stereotypes of ethnic minority students and children of lower-educated parents. Teachers and pre-service teachers seem more reluctant to express explicit stereotypes of ethnic minority students and tend to reduce their explicit stereotypes of lower-class students with time. These results may reveal the methodological limits of measuring explicit stereotypes, as it cannot be ascertained if it means that teachers do not hold explicit stereotypes or if they are unwilling to express them publicly. It is also perhaps unsurprising that evidence of explicit stereotypes is stronger for children of lower-class students than for ethnicity. First, it is

congruent with the research found in the previous section about biases in teachers' behaviour (e.g. track recommendations). And second, it could be more undesirable to publicly express ethnic stereotypes than social class stereotypes in contexts where meritocratic beliefs are strong. The results on gender are more mixed, with little systematic evidence of implicit or explicit stereotypes to the (dis)advantage of girls.

These results could in theory explain real-world discrepancies between students' prior achievement and teachers' scholastic judgements and decisions. Indeed, other areas of studies find correlations between implicit and explicit stereotypes and discriminatory behaviour. However, even in those areas, the relationship between both is often small. Before reaching any definitive conclusion, future research should be conducted in the school context to formally test if teachers' recommendations, holding previous performance constant, are linked to their endorsement of negative stereotypes.

Methodologically, there are substantial difficulties in testing such a hypothesis, which might explain why there is not yet any convincing evidence to this effect. If teachers are asked to report stereotypes before recommending the students, the objective of the study becomes inadvertently clear, which could lead teachers to correct any decisions that would have appeared as biased. The same difficulties arise if the questionnaire is distributed after providing students' recommendation: the teachers who are explicitly negatively biased might not be willing to report it to avoid retribution for their decision. It could be for this reason that the only study to date that reports a link between negative attitudes and achievement gap find this effect on implicit and not on explicit measures (Van den Bergh et al., 2010).

4.2 Congruence between a teacher's and students' social group

4.2.1 Theoretical considerations

Another potential mechanism that is often referred to in the literature as an explanation for teachers' biases is intergroup biases. While teachers may hold negative or positive societal stereotypes of specific groups of students independently of their own group membership, it is also possible that teachers manifest a preference for one's own group leading to differential

student treatment. Studies on the congruence between a teacher's and students' social groups have focused on groups such as race, ethnicity, or gender (Cho, 2012; Dee, 2007).

4.2.2 Empirical findings

Studies that examine how the racial or ethnic (mis)match between a teacher and a student influences teachers' evaluation are primarily set in the United States (Driessen, 2015). A review study on this topic that included 17 American studies on *subjective* teacher evaluations of cognitive (e.g., academic performance) and non-cognitive (e.g., classroom behaviour and diligence) school outcomes shows that the majority of these studies (i.e., 10) report a positive relationship between ethnic or racial match and subjective teacher evaluations. Nevertheless, two studies find a negative relationship, three studies find no relationship, and two studies report mixed findings (Driessen, 2015). These differences in findings do not seem to be related to the dependent variable of the study, or the publication year (i.e., publication years varied between 1995 and 2013). Furthermore, the review study indicates that most studies find no relationship between a racial or ethnic match between teacher and student and *objective* achievement measures (e.g., test scores).

A more recent American study also shows that teachers are likely to have higher expectations of same-race students than of different race-students (Gershenson et al., 2016). This study uses data from 2002 on more than 16,000 student-teacher combinations (i.e., dyads), and compares the evaluations of two different teachers for the *same* tenth grade student. This is an innovative design, as most studies examine biases in teachers' evaluations by comparing teacher evaluations for *different* students. When comparing different students with each other, findings can be confounded by characteristics of the students that influenced the teacher evaluations, but that are not visible to the researcher. By using different teacher evaluations for the same student, student characteristics that influence teacher evaluations that are not observed by the researcher (i.e., unobserved student characteristics) are accounted for. The study shows that, compared to black teachers, non-black teachers are 12 percentage points more likely to expect that black students will complete a high school degree or less. This is a substantial effect, as on average about 30 percent of the teachers has this

expectation for black students. Similarly, non-black teachers are nine percentage points less likely to expect that black students will complete a 4-year college degree than their black counterparts. Evaluations for white students do not vary according to the teacher's race. Similar findings are reported in another study using the same data (Fox, 2015).

Several studies have also examined how a gender match between the student and the teacher is related to teacher evaluations. Overall, studies on this subject are inconsistent. The studies by Gershenson et al. (2016) and Fox (2015) find little support for a relationship between a gender (mis)match and teacher expectations. However, Gershenson et al. (2016) do suggest that the effects of having a same-race teacher is more pronounced when there is also a gender match. Dee (2005), who also uses the evaluations of two teachers for the same student, does find a relationship between gender match and teachers' evaluations. He uses data of a longitudinal study in the United States in 1988 on more than 21,000 8th grade students and their teachers. The analyses on teachers' evaluations of student behaviour show that teachers are more likely to evaluate students of a different sex as disruptive, inattentive, and rarely completing homework. For example, the odds of a student being seen as disruptive are 1.36 larger when the teacher and the student do not have the same gender. Both male and female students are considered to be more disruptive by opposite sex teachers than by same-sex teachers. However, only female students are more likely to be perceived as inattentive by opposite-sex teachers, and only males are more likely to be perceived as rarely completing homework by opposite sex teachers. Another study conducted by Dee (2007) on the same sample reveals that on average, an assignment to an opposite gender teacher lowers student achievement by nearly 0.05 standard deviations.

Puhani (2018) investigates the relationship between teachers' gender and students' tracking recommendation with 5 years of administrative panel data on 1100 elementary schools, ± 12000 teachers, and 200,000 students in the German state of Hesse. He finds that there are no teacher gender effects within schools at the end of elementary school on either the recommended or the chosen middle school type. Another study using TIMSS data across 15 OECD countries, and that accounts for unobserved student characteristics, also contradicts a student-teacher gender matching effect on performance (Cho, 2012). Similarly, Holmlund &

Sund (2005) do not find strong support that a same-sex teacher improves students' educational outcomes in Sweden. This study accounts for unobserved student characteristics, the influence of past teacher characteristics, and teacher sorting into subjects.

Even if there is an effect of a student-teacher match on teacher behaviour, this does not have to be due to a (favourable) teacher bias. It is possible that students themselves behave differently when they are taught by a same-race and/or same-gender teacher, because these teachers act as role models for students. A recent study in Chile examines whether the higher performance of girls under female teachers (no effect of gender matching is found for boys) is due to (1) the fact that teachers represent role models or (2) (un)favourable treatment from teachers (Paredes, 2014). The author tests the hypothesis by examining differences in the effect of a teacher-student gender match across subject. If female teachers discriminate against boys (or favour girls), then the effect should not vary across subjects. However, if the gender-match effect is stronger in mathematics (stereotyped as a male domain, lower rate of female teachers), then this suggests that female teachers serve as role models for girls. The results indicate that the effect is stronger for mathematics and social sciences, which are the subjects with the lowest rate of female teachers. Even though the researcher concludes that these results confirm their "role model" hypothesis, it is possible that the researcher's assumption that discrimination does not vary across subject is invalid. Indeed, female teachers may discriminate more in subjects that are considered to be typically male/female, or they could try to compensate for the underrepresentation of women in certain fields by advantaging girls.

In sum, the evidence presented above is mixed. Overall, evidence for a favouritism bias appear stronger in the case of race in the United States than in the case of gender in European studies. If matching teachers' and students by gender do not favour either group's performance, this does not necessarily contradict the finding that teachers could hold stereotypes or differential expectations. It may well be that teachers have consensual negative attitudes or stereotypes of certain disadvantaged groups independently of their own group membership. However, there is currently little systematic evidence that those vary as a function of teacher's gender and that it extends to them enacting gender-biased behaviours (Li,

1999; Sansone, 2017).

4.3 Teachers' self-efficacy

Research in the area of teachers' self-efficacy (i.e. belief that one has the ability to bring about a specific outcome) may also be informative with respect to inequality in teachers' decisions. Indeed, decades of research show that self-efficacy is predictive of multiple desired classrooms outcomes such as: 1) high quality classroom environment (as measured by lessons advancing students' abilities, efforts to involve students and managing students' misbehaviour), 2) improved well-being, and lower levels of stress and burn-out for teachers, 3) higher students' achievement, motivation, and own self-efficacy (Zee & Koomen, 2016).

Teachers' self-efficacy could also have positive effects on teachers' inclusive practices and affect referral decisions. Teachers with high self-efficacy beliefs are more effective at teaching students with disabilities and at taking responsibility for their school difficulties (Brownell & Parajes, 1999; Brady & Woolfson, 2008). They are also less anxious about teaching inclusive classrooms and have more positive attitudes towards inclusive education and sociocultural diversity (Gao & Mager, 2011; Soodak, Podell, & Lehman, 1998). The literature on self-efficacy and referrals is more mixed. One vignette study provides preliminary evidence that self-efficacy could be linked to bias by showing that teachers with low levels of self-efficacy differentiate more by socioeconomic status in their special education referrals than teachers with a high sense of self-efficacy (Podell & Soodak, 1993). However more generally, the results appear inconsistent. Some studies indeed show that higher teacher self-efficacy is associated with more tolerance towards "problematic" students and less referrals to special education classes, but more recent and rigorous studies have failed to replicate this link (Zee & Koomen, 2016).

In the case of the Dutch context, one study finds that teachers feel less efficacious teaching ethnic minority students comparatively to ethnic majority students, especially in classrooms with low proportions of ethnic minority students (Geerlings, Thijs, & Verkuyten, 2018). If indeed Dutch teachers have lower self-efficacy beliefs with disadvantaged students, it

could in theory affect their expectations of these students and lower students' motivation and achievement. In sum, at this point, one study directly links teachers' self-efficacy to biased referrals, but most studies provide only initial indirect evidence that these concepts are related, and that low teacher self-efficacy could result in more inequality in tracking decisions.

5. Classroom composition

5.1. Theoretical considerations

Variations in teacher track recommendations may also be explained by a teacher's student population (i.e., classroom composition), as teachers may evaluate students in *reference* to other students in their class (Boone, Thys, Avermaet, & Van Houtte, 2018).

First, the class may function as a comparative reference group, and teachers could evaluate individual students in comparison to their class peers (Boone et al., 2018; Kaiser et al., 2017). Students are more likely to appear academically 'weak' when they attend a class with a high performance level, while they are more likely to appear academically 'strong' when they attend a class with a low performance level. Hence, students in high performing classes will receive lower teacher evaluations than equally performing students in low performing classes. In line with this, research shows that, when accounting for a student's own ability levels, students tend to receive higher grades in classes with a lower average ability level (Marsh et al., 2008).

Teachers may also have 'generalized' expectations for the classes they teach, and judge students *by* their class peers. This means that they evaluate students in the same class similarly, and hold relatively high or low expectations for *all* their students (Friedrich, Flunger, Nagengast, Jonkmann, & Trautwein, 2015; Li & Rubie-Davies, 2017). For example, the average socioeconomic status of the class may lead to lower teacher expectations for the entire class, partly because teachers perceive these students as less teachable (Agirdag, Van Avermaet, & Van Houtte, 2013; Li & Rubie-Davies, 2017). These generalized expectations are possibly (also) due to the normative context in class. For example, Boone et al. (2018) argue that parents from higher socioeconomic backgrounds are more likely to pressure teachers for higher track

recommendations, leading to a normative climate in which teachers are more likely to give higher track recommendations. Generalized teacher expectations are mostly introduced with respect to the influence of the socioeconomic or ethnic composition of the class on teacher expectations (Boone et al., 2018). However, they may also play a role in the influence of the average performance of the class. For example, in classes with a higher average performance level, teachers may perceive their students to be more ‘teachable’, leading to higher teacher expectations for the entire class.

5.2 Empirical findings

Research on the relationship between the composition of the class and teacher expectations mostly focuses on class-level ability or performance, socioeconomic status, or the share of ethnic minority students. Empirical findings on these relationships are equivocal. Some studies suggest that there is a positive relationship between the average ability level and/or the socioeconomic composition of a class and teacher expectations or perceptions. For example, a study on the perception of American teachers of students’ literacy skills among 9,493 students in 1,822 classrooms, shows that teachers have higher perceptions of the literacy skills of children in classes with higher average literacy skill levels and/or social-status (i.e., composite measure of parental income, education, and occupational prestige) (Ready & Wright, 2011). Especially students from lower socioeconomic backgrounds tend to be perceived more positively by their teachers in classes with a higher average social status.

Research on teacher track recommendations among more than 7000 Dutch students in 500 classes in the school year of 2004/2005, also indicates that teacher track recommendations are higher in classes with (1) a higher average score on the school leaver test, and/or (2) a smaller proportion of low-educated parents (i.e., pre-vocational education) (Timmermans et al., 2015). However, the influence of the composition of the class on teacher track recommendations is independent of the individual’s student’s demographic traits.

In contrast to these findings, research in two large cities in Flanders (i.e., Ghent and Antwerp) shows that children in higher ability classes receive *lower* track recommendations

than equally able children in lower ability classes (Boone et al., 2018). Especially children with lower ability levels tend to receive a higher track recommendation when they attend a class with a lower average ability level. These findings correspond with comparative reference group theory, as they indicate that teachers use the average ability of the class as a standard for comparison. This study finds no relationship between the average socioeconomic status of the class and teacher track recommendations, once a student's individual traits are accounted for.

Differences between these Flemish findings and the findings of the Dutch study by Timmermans et al (2015) may be explained by differences in the track recommendation procedures in both countries (c.f. Boone et al., 2018). Possibly, the socioeconomic composition of the class only influences teacher track recommendations in the Netherlands and not in Flanders, as teacher track recommendations carry more consequences the Netherlands than in Flanders (cf. Boone et al., 2018). While track recommendations are officially not binding in both the Netherlands and Flanders, schools in Flanders have no formal entrance criteria, whereas Dutch secondary schools can reject students based on their track recommendation. Hence, in comparison to Flemish schools, Dutch higher socioeconomic status school may experience more parental pressure on their decisions. In these Dutch schools, there may be a culture in which teachers are likely to give children the benefit of the doubt.

For related reasons, the average ability of the class may influence Dutch teachers differently than those in Flanders. In the Netherlands, but not in Flanders, there are national standardized tests on which teachers can base their track recommendations (Boone & Van Houtte, 2013). Hence, Dutch teachers may compare their students to other students in the *country*, whereas Belgian teachers are likely to compare students to their *class* peers. This may lead to 'generalized expectations' among Dutch teachers, causing them to give higher track recommendations to students in higher ability classes.

However, other empirical findings contradict this explanation. First, results from an earlier Dutch study (i.e., in the school year of 2002/2003) correspond with the Flemish findings, as it shows that the percentage of students with a low socioeconomic status in class does not explain variation in teacher track recommendations, and that students in classes with lower performance levels receive *higher* track recommendations (Driessen et al., 2008). Second,

research in New Zealand does not corroborate the above line of reasoning (Meissel et al., 2017). Similar to the Netherlands, ability groups in New Zealand are based on the judgement of the teacher who can rely on standardized test scores. This study also shows that teacher judgements for reading and writing are lower for students who attend schools with a higher average achievement level. Moreover, the average socioeconomic status of students in school is unrelated to teacher judgements. Contrary to the Netherlands and Flanders, students in New Zealand are placed into different ability groups for different subjects.

Discrepant findings may also be due to differences in the operationalization of the class composition variables. For example, some studies operationalize the average performance level of the class by means of standardized tests that are administered by the researcher (e.g., Boone et al., 2018; Driessen et al., 2008). These tests are so called low-stakes tests, as they bear little consequences for the students who make them. Other studies measure the average performance level of the class by test scores that teachers actually used to base their judgements on (e.g., Meissel et al., 2017; Timmermans et al., 2017). Similarly, scholars measure the socioeconomic status of students in different ways. For example, they use the share of disadvantaged pupils in class (Driessen et al., 2008), the proportion of students with low educated parents (Timmermans et al., 2015), or the average occupational status of students' parents (Boone et al., 2018).

Findings with respect to the ethnic make-up of the class and teacher expectations are also inconclusive. Boone et al. (2018) and Driessen et al. (2008) find no relationship between the ethnic composition of the class and teacher track recommendations in, respectively, Flanders and the Netherlands. However, a study by Thys & Van Houtte (2016) indicates that teacher expectations in primary schools in Ghent and Antwerp are higher in classes with a lower share of ethnic minority students. It is possible that these lower expectations do not translate into lower track recommendations (Boone et al., 2018).

A recent study in Germany suggests that teachers are more accurate when students are in the numerical minority in class (Kaiser et al., 2017). This study makes use of a computer game in which teachers can ask questions to students in a simulated class. Students are represented by a photo and a name, and the traits of students in the class can be experimentally

manipulated. Findings indicate that teachers are more accurate about a student's percentage of correct answers when a student's ethnic group is in the numerical minority in class. This result is found irrespective of whether Turkish, Asian, or German students are in the numerical minority. Similarly, teachers were more accurate in their judgements about (fe)male students when they are the minority group in class. Kaiser et al. (2017) suggest that teachers process student information better when students stand out in class, leading to more accurate judgements.

6. Institutional settings

6.1. Theoretical considerations

There are large variations across Dutch schools in the extent to which teachers' track recommendation match students' actual performance on the school leaver's test (Inspectie van het Onderwijs, 2018a). Moreover, schools also seem to vary in the extent to which teachers' track recommendations are biased by students' demographic traits (Timmermans et al., 2015, 2016, 2018). Hence, inequality in teacher track recommendations may be dependent on institutional differences across schools.

Many theoretical perspectives from different disciplines have suggested that schools are cultural contexts that shape the way in which teachers think and behave, thereby contributing to the reproduction of social inequalities (Adams, Biernat, Branscombe, Crandall, & Wrightsman, 2008; Bourdieu & Passeron, 1977; Duru-Bellat & Tenret, 2009; Lamont, Beljean & Clair, 2014; Stephens, et al., 2012; Thornton & Ocasio, 2008). These scholars propose that, historically, schools were set up with one specific objective: to educate an elite. Since the development of mass education, the main objectives of schools have evolved to combine this goal with other tasks: such as to ensure social mobility by providing all students with the same opportunities. However, these cultural and historical theories emphasize that schools contain in their institution old traces of power and cultural dynamics that could resurface as modern subtle biases. They contend that most teachers might strive to treat students similarly, but that schools' dominant culture encourages a representation of "good students" which tends to

advantage privileged students. As a result, these students are subtly perceived as more capable and academically worthy, and disadvantaged students' cultural mismatch is often misconstrued as academic inability or lack of motivation (Croizet, 2008).

Prior findings in social psychology suggest that the expression of prejudice is related to social norms (i.e. accepted behaviours from individual in a specific group or cultural environment), suggesting that schools as institutional environments could in theory influence teacher biases (Crandall, Eshlema, & O'brien, 2002). While previous research has not directly investigated the specific topic of social norms conveyed in schools, several studies have found that the expression of prejudice is not only related to individuals' initial beliefs but are also contingent on the perception that an environment is accepting or encouraging of prejudicial views (Guimond, Begin, & Palmer, 1989). Following this reasoning, if some school environments are perceived as promoting egalitarian norms, this could diminish the likelihood of bias arising among teachers. On the other hand, if teachers tend to perceive groups in a subtle hierarchical manner and perceive social norms as explicitly favouring certain types of cultural capital over others, this could in theory manifest in biased teachers' decisions. At this stage, the extent to which specific types of schools endorse egalitarian social norms is still unknown. It is likely that strong variations exist between more or less elite schools. But previous qualitative studies suggest that schools' dominant culture tends to value high social class cultural capital over other forms of cultural capital (for a review, see Stephens et al., 2014). If so, then certain schools could contribute to the production of biased evaluations from teachers.

School policies and procedures may also influence the extent to which teachers feel accountable for their track recommendations. Several authors contend that when people are held accountable for their decisions, their motivation to form an accurate impression is likely to be higher (Bodenhausen, Macrae, & Sherman, 1999). This additional motivation could lead perceivers to either avoid forming stereotypical impressions or relying on stereotypes. Research on employment decisions provide evidence for this proposition (Reskin, 2003). For example, a meta-analysis on gender bias and stereotypes in employment decisions finds that when participants are held accountable for their decisions, their decisions are less biased. (Koch, D'Mello, & Sackett, 2015). According to social cognition theories, accountability reduces

discrimination, because it deters teachers from relying on small sets of cues, and makes them more motivated to base their judgements on conscious and comprehensive information-integration strategies (Bodenhuasen et al., 1999; Koch et al., 2015; Pit-ten Cate, Krolak-Schwerdt, & Glock, 2016). There is ample evidence confirming that bias is less likely when a person is motivated to develop an accurate impression of the target person and is willing to invest time and cognitive effort and most likely to occur when individuals are motivated to be time-efficient (Pit-ten Cate et al., 2016).

However, research is still unclear on how accountability among teachers could be effectively enhanced. Possibly, accountability for tracking decisions may increase if teachers have to justify their decisions to external actors (for instance children and their parents). There might also be some benefits in encouraging discussions about tracking decisions with the principal or other teachers, but potential drawbacks from involving actors from the same school should also be considered. Accountability may be reduced if teachers from the same school are also striving to maintain positive work relationships with colleagues. Another possibility may be to foster primary schools to evaluate ‘anonymized’ student profiles together with teachers in secondary schools³. Moreover, accountability might increase if primary school teachers who formulate track recommendations engage in a dialogue with secondary schools about the quality of their track recommendations. For example, teachers could learn from cases where the track recommendation was deemed ‘inaccurate’ by the secondary school. In the Netherlands, there is some descriptive work that indicates that there is quite some variation across schools in aspects of the track recommendation procedure that may influence teachers’ level of accountability (Smeets et al., 2014). For example, in most schools multiple actors (e.g., teachers, tutors, and the school principal) tend to be involved in the track formulation procedure, but this is not the case for all schools. Similarly, primary schools seem to differ in the extent to which they evaluate specific pupils together with secondary schools before providing definite track recommendations, and few primary schools systematically evaluate their track recommendation procedure on the basis of children’s performance in secondary school. It is

³ In the municipality of Utrecht, primary schools have the possibility to evaluate anonymized student profiles with secondary schools before they formulate the final track recommendation (see Smeets et al., 2014).

unclear whether and how these differences across schools in track recommendation procedures may influence inequality in track recommendations.

Another potential pathway to reduce inequality in teachers' evaluations of students resides in the structure of the larger educational systems. Based on individual-level theories on evaluative practices, discrimination, and stereotyping, hypotheses can be derived on the role of educational institutions' stratification processes in increasing inequality in teacher expectations and judgements. First of all, early differentiation in school may influence the extent to which teachers evaluate students. The earlier teachers are required to place students into different educational tracks on the basis of their ability, the earlier they will start categorizing and classifying students. This early categorization and classification can contribute to the (re)production of social inequality (Lamont, 2012). Children who are classified as 'able' students, will reap subsequent educational benefits. For example, they will be challenged more by their teachers and will be offered more educational resources, leading to even greater ability differences across students. Moreover, the classifications can act as self-fulfilling prophecies, causing children who are initially classified as 'high ability' students to actually perform better in school. The earlier these self-reinforcing mechanisms are set in motion, the greater social disparities in educational outcomes may become, as children from higher socioeconomic backgrounds are more likely to be classified as 'able' students.

Early differentiation in school could also enhance the likelihood that teachers discriminate in their track recommendations. According to statistical discrimination theory, teachers are more likely to evaluate students on the basis of the traits of the social groups to which they belong, when teachers lack information, and are more uncertain about the educational potential of students. Because the educational potential of a child is harder to evaluate when the child is younger, teachers are more likely to engage in statistical discrimination when they have to formulate track recommendations for younger students.

Formalization of the tracking procedure could limit discrimination (Reskin, 2003). Formal guidelines may reduce a teacher's insecurity about the educational potential of a child. For example, teachers may be less likely to statistically discriminate, when they are formally required to base their recommendations on standardized test scores. When teachers have to

follow specific guidelines in their track recommendation, there is also less room for individual taste. Nevertheless, formal guidelines could lead to a legitimization of ‘arbitrary’ evaluation criteria that can reproduce social inequalities (Lamont, 2012). For example, schools may explicitly take into account parental wishes, thereby implicitly legitimizing higher evaluations for students from higher socioeconomic backgrounds.

Educational systems also vary in the extent to which teacher track recommendations are binding. For example, in some German states and in Flanders, parents can decide themselves whether they follow the teacher’s track recommendation (Dollmann, 2015; Boone & Van Houtte, 2013). This may not only reduce the extent to which teachers *feel* accountable for their recommendations, but also factually reduces the extent to which they are accountable for socioeconomic inequality in track attendance. By leaving the decision to parents, and thereby reducing the role of teachers in track allocations, the influence of teacher biases in track *allocations* may diminish. However, one should bear in mind that social inequality in parental aspirations is even larger than social inequality in teacher evaluations (Dollmann, 2015).

The role of educational institutions (e.g., policies in educational systems or schools) in teachers’ biases has not been extensively empirically tested. Yet, many theoretical perspectives in social sciences suggest that institutions enable and constrain teachers’ decisions. In this section, we will review two empirical research lines developed in psychology which have established a link between the role of institutions and teachers’ biases as well as a third research line developed in sociology that focus on how the educational systems’ (track recommendation) policies are related to social (and ethnic) inequality in educational outcomes.

6.2. Empirical findings

6.2.1 School selection function

The first research line directly follows from the theoretical and historical perspective on educational institutions mentioned above and investigates the role of selective educational practices on teachers’ biases. Experiments have been conducted to test whether the use of selective educational practices could encourage the likelihood of potential biases to manifest.

The hypothesis is based on the idea that schools operate with two, at times competing, functions. The first is to provide all students with equal access, treatment, and learning opportunities: the educational function of school. The second is to assess the students to determine at later stages which are deserving of pursuing higher education: the selection function of school. Those two tasks represent two distinct institutional objectives of the educational institution that have been assigned by society (Darnon, Dompnier, Delmas, Pulfrey, & Butera, 2009; Dornbusch, Glasgow, & Lin, 1996). Teachers have to reconcile these two objectives in their daily practices. At times, they have to focus on comparing students to each other to detect which show academic promise, and at other times, they are devoted to help all students to improve and ignore students' initial level of competence. Educational practices that direct evaluators towards their selective role in the institution (of hierarchization of students), instead of their role of helping all students, could encourage bias in assessments of students of different social class with identical performance.

In two experiments, participants (university students playing the role of teachers) were asked to assess a dictation test which was supposedly produced by a low or a high socioeconomic status student (Autin, Batruch, & Butera, 2018). To assess the test, participants had to either use a selective assessment method (i.e., grading) or an educational assessment method (i.e., providing comments). The results of these experiments show that when evaluators use a 'selective assessment method', they find more mistakes in the test of a student with a low socioeconomic background than in the test of a student with a high socioeconomic background. This difference is not found when assessors use an 'educational assessment method'. The authors conducted a follow-up study to ascertain that the underlying effect is due to the selective component of grading and not to the practice of grading as such. In this study, the function of assessment (selection vs. educational) was manipulated along with the assessment method, and the student's socioeconomic status. Depending on the experimental condition, participants had to either provide grades or comments. Additionally, participants were told that they had to look for mistakes in the student's test to either select the student for another class (i.e. selection condition), or to help the student improve (i.e. educational). The results reveal that the function of assessment (selection vs. educational), and

not the assessment method (grading vs. comments), predicts a biased assessment (i.e. more mistakes found for the same test). This suggests that it is the selective purpose of assessment, rather than the assessment method itself, that lead evaluators to artificially differentiate between students of advantaged vs. disadvantaged background (Autin et al., experiment 3). Batruch et al. (2018) test a similar hypothesis in one of their experiments on tracking decisions. More specifically, they examine whether socio-economic differences in teachers' tracking decisions are dependent on a school's function (i.e. selection vs. educational). In this experiment, university students were asked to play the role of teachers. Their findings are consistent with the studies presented above. The selective function of school appears to be more related to biased tracking decisions than the educational function of school (Batruch et al., 2018).

These studies indicate that it is not the use of those practices per se, or only individual biases that are directly responsible for the artificial gap, but that encouraging evaluators to focus on selecting students, rather than teaching them, could encourage biased assessments. In these experiments, the artificial gap either disappeared -or appeared smaller- when the practices served an educational purpose rather than a selective purpose. These experiments further suggest that teachers do not entirely practice independently of institutional forces; their work partly reflects the expectations of institutions. As such, institutions may want to consider what are the underlying function of educational practices and to consider their effects on students as well as teachers. This may matter more in the context where selective practices such as tracking procedures are institutionalized. Implementing practices that make teachers compare and hierarchize students may enhance the likelihood that they rely on irrelevant information such as students' background to unconsciously find objective justification of students' ranking in ambiguous cases. These experiments additionally provide preliminary evidence that educational institutions can reduce inequality if they allow teachers to entirely devote their efforts to help all students improve. It should be noted however that the experiments were conducted on varied and small samples (students playing the role of students, pre and in-service teachers). Before reaching any definitive conclusions on the role of schools' function of selection, these experiments should be either replicated on bigger samples

of teachers or be tested in real-world settings to determine if these studies have sufficient external validity.

6.2.2 Accountability

A second research line focuses on the role of accountability in increasing accuracy in teachers' decisions (Pit-ten Cate et al., 2016). The authors hypothesize that accountability can play a role in increasing individuals' motivation to invest effort in the decision, and use an experimental longitudinal design to test this specific hypothesis: School teachers from Luxembourg were asked to make tracking decisions at three separate points in time for students of ethnic minority vs. majority background. After making tracking decisions for the first set of case vignettes, participants were asked to answer how accountable they felt for their decision on a 7-point scale. This scale was used to render accountability salient in the mind of the participants. Participants then reviewed another set of vignettes and were asked to come back six months later. The accuracy of the participants' tracking decision was evaluated by comparing the tracking decision made for vignette students to a criterion developed by the researchers which reflects the extent to which a student fits in a particular track based on his or her achievement. Additionally, the researchers assessed the extent to which participants were overconfident when making a wrong decision or underconfident when making a right decision.

The results of this experiment reveal that the average level of accuracy for the tracking decisions is high. Nevertheless, tracking decisions for ethnic majority students have a higher accuracy than the tracking decisions for ethnic minority students. Accountability seems to reduce these ethnic differences: after respondents had completed questions about accountability, the accuracy of the tracking decisions for ethnic minority students increased. The results also indicate that participants are more likely to be overconfident in their decision for ethnic minority students before answering the accountability questions than after. Just after responding to the accountability questions (i.e. at time 2), participants' level of confidence was more in line with the actual accuracy of the decision. There are no significant differences in accuracy for ethnic minority students between Time 2 and Time 3 (i.e. six months later);

however, accuracy for ethnic majority students improved at Time 3, creating again a discrepancy in accuracy between both groups.

The results are consistent with a previous experiment conducted on secondary school German teachers (Krolak-Schwerdt, Böhmer & Gräsel, 2013) which shows that in context of low accountability, but not high accountability, there are ethnic differences in performance ratings in mathematics and reading. Overall, the results of these studies provide some initial support for raising teachers' feeling of accountability to reduce differences in accurate tracking decisions between advantaged and disadvantaged group members.

6.2.3 Track allocation policies and inequality in educational outcomes

A third line of research examines the relationship between educational policies pertaining to tracking practices and inequality in educational outcomes. While this research line does not specifically focus on inequality in *teacher evaluations* of students, teacher practices may be an important underlying mechanism that link educational policies to inequality in educational outcomes. We discuss empirical evidence with respect to three policies: (1) tracking age, (2) standardized examinations, and (3) binding versus non-binding track recommendations.

Research consistently shows that social inequality in teacher track recommendations is larger in educational systems in which students are tracked at a younger age (Van de Werfhorst & Mijs, 2010; Van de Werfhorst, 2018). Teacher practices may play a role in this relationship, as teachers may be more uncertain about children's potential when children are younger, making them more likely to rely on the average traits of the demographic group(s) to which children belong.

Findings are mixed with respect to the relationship between standardized examinations and inequality in educational outcomes, with studies showing a positive, a negative, or no relationship (Bol et al., 2014; Wössmann et al., 2009; Horn, 2009). However, one study that *does* suggest that standardized examinations reduce educational inequality, specifically focuses on an outcome that is related to track placement, namely student placement in gifted programs (Card & Giuliano, 2015). This study, which was based on a natural experiment, standardized

examinations only reduce inequality in some educational systems. Bol et al. (2014) show that social inequality in the mathematics performance of 15-year-olds is larger in highly tracked educational systems (e.g., the Netherlands), and that this relationship is especially pronounced in educational systems *without* central examinations. Bol et al. (2014) argue that central examinations may reduce social inequalities in tracked systems, because they make teachers more likely to allocate students on the basis of their factual performance. However, Bol et al. (2014) find that central examinations are related to *higher* levels of social inequality in educational outcomes in highly comprehensive educational systems.

The extent to which teacher track recommendations are binding seems to influence inequality in educational outcomes. When track recommendations are not binding, the teacher's recommendation is merely a suggestion. Conversely, when track recommendations are binding, parents are required to follow the teacher's recommendation. In a recent study, Dollmann (2015) uses a natural experiment to test the role of binding track recommendations in inequality in the transition from lower to upper secondary school. He relies on data from two successive student cohorts in North Rhine-Westphalia in Germany in which the government reformed the teacher track recommendation procedure. Because of this, the first cohort received a non-binding track recommendation, whereas the second cohort received a binding track recommendation. While social inequality in the transition from lower to upper secondary school is lower among students in the second cohort (i.e., who received binding teacher track recommendation), ethnic inequality is higher. In the second cohort students from Turkish backgrounds are less likely to make ambitious transitions. These findings are probably explained by social and ethnic differences in parental aspirations. Parents from high socioeconomic backgrounds have higher educational aspirations for their children than parents from low socioeconomic backgrounds, even if their children's performance in school is equal. However, immigrants and ethnic minority groups tend to have higher educational aspirations compared to their equally performing native counterparts.

7. Interventions

There are few studies on intervention strategies to improve teachers' evaluations. Existing intervention studies focus on heightening teacher expectations, rather than reducing biases. However, a recent intervention study explicitly addresses the extent to which teacher trainings can enhance accuracy and reduce biases in teacher evaluations (Krolak-Schwerdt, Pit-ten Cate, & Hörstermann, 2018). In this unpublished study, two different trainings are tested with respect to their effectiveness in reducing ethnic bias, by comparing ethnic biases in teacher track recommendations for hypothetical students before and after a training. In the first training, nine teachers in Luxembourg received an overview of theories on information processing strategies, judgement formation, and (accuracy in) decision making (Pit-ten Cate, Krolak-Schwerdt, Hörstermann, & Glock, 2017). Subsequently, teachers constructed strategies for forming track recommendations in an interactive way and received feedback on the application of these strategies on their track recommendations and student inferences. In the second training, 21 teachers were asked to rate the importance of different student traits for the formation of track recommendations. Subsequently, a computer calculated 'track recommendation' rules for each individual teacher that matched the track formation strategy that teachers intended to use. Teachers were then asked to make tracking decisions in which they received feedback on the extent to which their decision was consistent with their intended strategy. Findings indicate that ethnic biases are generally small. However, before the training, teachers made more accurate tracking decisions for students from the ethnic majority than for students from the ethnic minority. After the training this difference disappeared, because teachers became more accurate in their track recommendations for students from the ethnic minority, but not for students from the ethnic majority. The same findings are reported for both trainings.

There is also a wide array of psychological interventions studies that could indirectly be relevant to changing teachers' biases as they have been designed to either reduce bias among the general population or at changing institutional practices to close students' achievement gaps (Dee & Gershenson, 2017; Dittman & Stephens, 2017). Their results may not be yet

directly applicable to tracking decisions, but they provide initial guidance on how future interventions could be designed to implement changes for teachers to reduce inequality in education. The common themes explored in the literature designed to reduce biases could be organized around three specific axes: trainings to reduce implicit biases, increasing awareness of biases, and changing emotional responses to intergroup relations (Blair, 2002; Burgess, Van Ryn, Dovidio, & Saha, 2007; Devine & Monteith, 1993; Dovidio et al., 2004; Okonofua, Paunesku, & Walton, 2016; Schellhaas & Dovidio, 2016). An unpublished meta-analysis on implicit biases reduction provides mixed findings (Forscher et al., 2018). Methods currently in use are able to produce changes in implicit biases, but the changes do not extend to individuals' behaviours. These results suggest that existing implicit bias trainings may not yet be effective at changing teachers' decisions.

Increasing awareness of biases and changing emotional responses to intergroup relations have not been as widely or systematically tested, but research provides a preliminary indication that increasing bias awareness; encouraging a focus on the individual rather than on the group membership; or building positive emotions through training, contact, or by encouraging perceived similarity or partnership could be effective to improve teachers' relationship to minority students.

Finally, an important section of the intervention literature has found that structural and institutional factors such as school' implicit cultural norms (Stephens et al., 2014), schools' use of competitive practices (Goudeau & Croizet, 2017; Jury, Darnon, & Smeding, 2015; Smeding, Darnon, Souchal, Toczec-Capelle, & Butera, 2013; Souchal, Toczec, Darnon, Smeding, Butera, & Martinot, 2014), or the meaning schools have assigned to evaluative settings can have an impact on achievement gaps (Harackiewicz, Canning, Tibbetts, Priniski, & Hyde, 2016; Oyserman & Destin, 2010; Paunesku et al., 2015; Tibbetts et al., 2016; Yeager et al., 2016b). The central objective of these latter interventions has been to change the meaning of evaluative settings to help students adapt the attributions they make for their academic difficulties, increase their expectations and motivation, and foster their feeling of belonging in school. Specifically, the interventions either involve a reinterpretation of the individual's self-image or feelings (image of self or group, emotions, interpretation of stress), of the meaning of the

evaluation (change students' attribution for success or failure, encourage a mind-set focused on effort), or their perception of contextual fit (encourage feelings of social-belonging, change perceptions of schools' cultural norms to fit with one's own). These interventions find that they can increase minority students' performance by making them more resilient in the face of academic challenges. They do so by adjusting their psychological understanding of the different components of performance to avoid them assuming that they are unable to succeed (Aronson, Fried, & Good, 2002; Paunesku et al., 2015; Sherman, 2013; Walton & Cohen, 2007; Walton & Cohen, 2011; Yeager et al., 2016a; Yeager et al., 2016b). These could be considered as promising directions for future interventions studies on teachers, as the changes are likely to have long-lasting effects in educational settings if implemented on teachers rather than on students. While they do not directly address biases in tracking decisions, they can prevent tracking decisions being made on students who have not been given the best opportunity to express their academic potential.

8. Implications for the Dutch debate on the role of teachers

When the Inspectorate for Education published *The State of Education* on April 13th 2016, then Minister of Education, Culture and Sciences was interviewed at TV show *Nieuwsuur*. Jet Bussemaker was asked to reflect on the finding that the social gaps in teacher recommendations increased after the final primary school test was postponed until after the recommendation was given. She more or less responded that teachers do not do their work well: "the beliefs of teachers themselves have come to play a role, which is very undesirable" ["De eigen opvattingen van leraren spelen een rol, dat kan dus niet, dat is zeer, zeer ongewenst"]. Something similar is expressed by then member of parliament of the Partij van de Arbeid Loes Ypma in *NRC-Handelsblad*, who said that the national test should "correct human mistakes, like under-recommendation when mother wears a headscarf or father works in the factory" (*NRC-Handelsblad* 14-4-2016).

The process of school recommendations is very complex, and it is likely that teachers evaluate other school-relevant factors in determining a rationally informed track

recommendation for each child. Such school-relevant factors make it understandable how recommendations are formed. Two sets of factors may be taken into account this way, child-related attributes (such as academic motivation) and family-related factors (such as being stimulated at home). However, while such factors may make the recommendation of teachers understandable, they do not make the recommendations fair. Socially or ethnically biased recommendations can be unfair even if there is no taste-based discrimination among teachers, in the sense that teachers would be discriminating against disadvantaged children even if their potential would allow recommendations for higher-level programmes. It could be considered unfair if other people's (e.g. teachers', or parents') expectations determine a child's options in the school career, and if group characteristics are imposed on individual students (statistical discrimination). Children may not be able to fully demonstrate their potential if they are sorted into the lower-level tracks. Moreover, research suggests that Dutch students who receive a higher track recommendation than one would expect, also tend to end up higher in the educational distribution (Smeets et al., 2014; De Boer et al., 2010). Contrarily, students who receive a lower track recommendation than one would expect, tend to end up lower. Misplacements could be particularly consequential for children of disadvantaged backgrounds (c.f. Jussim & Harber, 2005). If the Dutch educational reforms have raised the inadequacies in track recommendations, such reforms need to be scrutinized.

In the Netherlands, one important policy issue is that the final standardized test in primary school (Eindoets Basisonderwijs) is taken after the track recommendation is given. The reform that postponed the standardized test was introduced in 2015, and socioeconomic gaps rose after the reform. While schools are now obliged to reconsider the advice if the final test results in higher-level recommendations than the initial advice, it is important that schools are offered guidelines how to organize the reconsideration. If schools receive little guidance, it is plausible that the recommendations are disproportionately adjusted for children of well-educated families, as their parents may be more actively involved in the reconsideration. In that case inequality would not decrease, but rather rise as a consequence of reconsideration.

One debate in the literature is whether standardized forms of examination and screening can reduce social inequalities in track placement (Card and Giuliano 2015; Bol et al.,

2014, versus Wössmann et al., 2009; Horn, 2009). There are some current tides in the education discourse in the Netherlands that would want to downplay the role of centralized forms of testing. The council for secondary education (VO-Raad) is evaluating the central examinations in secondary school, and the final primary school test has already been made less important for the track allocation. Decreasing standardization, which corresponds to increasing the autonomy of schools which is advocated in contemporary debates such as Onderwijs2032 and Curriculum.nu, may lead to higher socioeconomic inequalities in education (Van de Werfhorst, 2015).

Besides the final primary school test, there is a lot more standardized information available about students' performance. In particular, the Student Performance System (Leerlingvolgsysteem, LVS) contains a lot of information on repeated standardized tests throughout primary education. The LVS tests are formative, and not summative, meaning that these are meant to help to define which aspects of learning need further attention rather than evaluate the mastery of the past materials. Formative tests are not meant for track recommendations, unlike the final primary school test. Nevertheless, it is worthwhile to study the potential role of the LVS for the track recommendations, and to make the use of the LVS more systematic. Current recommendation practices already take notice of the LVS results, and one avenue of further thinking would be to develop more formal usage of the LVS data in the recommendation process. The potential of the LVS has not yet been fully examined. To the extent that the LVS system is useable in the recommendation process, it would provide a main advantage over the final school test, namely that it follows student performance for a longer time instead of during one week in the final school year. However, one should also be aware of an important potential drawback of using the LVS scores, as it would imply that the track recommendation is based on very early information of students, bringing the moment of selection (or at least of the decisions leading to it) to the fore. The Netherlands is already exceptional with regard to the early age at which students are sorted, and early selection is associated to larger socioeconomic gaps. Pulling the decision-making process to the fore may also make early selection to factually happen even sooner than now.

Given the importance of early tracking for enlarging socioeconomic gaps in educational achievement (Van de Werfhorst, 2018), it is worth emphasizing that later (and less rigid) tracking may pay more tribute to the dynamic character of the formation of academic competences during adolescence. Developing less rigidity in the tracking structures, offering broader schools offering multiple tracks, and reviving longer ‘bridge years’ that postpone the moment of selection are all sensible strategies to make the primary school track recommendation less decisive for students’ future school careers (likely reducing socioeconomic inequalities in educational careers). The current development in the Dutch school system goes in the opposite direction: broad schools are slowly disappearing from the big cities, bridge years are reduced and become smaller. Secondary schools report that their first year can become unitary because the quality of the recommendations by the primary schools has improved. However, as our review suggests there is still room for improvement when it comes to reducing socioeconomic gaps. The current practice to encourage ‘combined recommendations’ of two school tracks (“meervoudige adviezen”) seems a step in the right direction. It helps to address the issue that disadvantaged students are more “myopic” when it comes to educational decision making, unable to oversee the long-term consequences of their preferences (Lucas, 2001). Thinking about policies to improve the recommendation process (and reducing inequities therein) can, however, also focus on the structure of the secondary education sector itself, not just at what happens in primary schools.

It is the combination of pieces of information available to schools that likely forms the optimal basis for the track recommendation. Currently schools lack the information about the later school career of their former students. Potentially improved information to schools coming from various sources, including the Vensters website and the reports sent to schools by the National Cohortstudy Educational Careers (Nationaal Cohortonderzoek Onderwijs, NCO) will inform schools how well their recommendation worked. An additional useful step would be to then evaluate future school careers of former students at the school level (see Diris 2012 for an example at the aggregate level), so schools can compare their recommendation-career correlation with that of other (comparable) schools. The evaluation of track recommendation procedures may also make teachers feel more accountable for their decisions, increasing

considered decision making, and reducing automatic judgement processes that are more prone to biases.

With this literature review we aimed to enlarge the scope of knowledge on socioeconomic inequalities in education. While most of these inequalities originate from the home, making it hard for schools to tackle inequalities, also the role of educational professionals needs further consideration. Like in other professional fields, experts who have the best intentions to the opportunities of all of their clients, can still unknowingly contribute to the creation and maintenance of inequalities (Schwalbe et al., 2000).

From the perspective of education policy an important question is how institutions can be formed that promote equal opportunities. Part of the solution may lie at the level of schools and teachers. Teachers need to become aware of the existence of social, ethnic, and gender inequalities in the Dutch education system, of their possible subjectivity in the treatment of different groups of students, and of their potential role in reducing inequalities. It is striking that inequalities do not get much attention in teacher training programmes, and one avenue of improvement may be found there.

References

- Adams, G., Biernat, M., Branscombe, N. R., Crandall, C. S., & Wrightsman, L. S. (2008). Beyond prejudice: Towards a sociocultural psychology of racism and oppression. In G. Adams, M. Biernat, N. R. Branscombe, C. S. Crandall, & L. S. Wrightsman (Eds.), *Decade of Behavior. Commemorating Brown: The social psychology of racism and discrimination* (pp. 215-246). Washington, DC, US: American Psychological Association.
- Agirdag, O., Van Avermaet, P., & Van Houtte, M. (2013). School segregation and math achievement: A mixed-method study on the role of self-fulfilling prophecies. *Teachers College Record, 115*(3), 1–50.
- Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*(4), 731-746.
- Al Ramiah, A., Hewstone, M., Dovidio, J. F., & Penner, L. A. (2010). The social psychology of discrimination: theory, measurement and consequences. In L. Bond, F. McGinnity, & H. Russell (Eds.), *Making equality count: Irish and international research measuring equality and discrimination* (pp. 84–112). Dublin: The Liffey Press.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125.
- Autin, F., Batruch, A., & Butera, F. (in press). The function of selection of assessment leads evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology*.
- Auwarter, A. E., & Aruguete, M. S. (2008). Effects of student gender and socioeconomic status on teacher perceptions. *The Journal of Educational Research, 101*(4), 242-246.
- Bakker, J., Denessen, E., & Brus-Laeven, M. (2007). Socioeconomic background, parental involvement and teacher perceptions of these in relation to pupil achievement. *Educational Studies, 33*(2), 177–192.

- Barg, K. (2012). The influence of students' social background and parental involvement on teachers' school track choices: reasons and consequences. *European Sociological Review*, 29(3), 565–579.
- Batruch, A., Autin, F., Bataillard, F., & Butera, F. (in press). School selection and the social class divide: How tracking contributes to the reproduction of inequalities. *Personality and Social Psychology Bulletin*.
- Becker, D. (2013). The impact of teachers' expectations on students' educational opportunities in the life course: An empirical test of a subjective expected utility explanation. *Rationality and Society*, 25(4), 422–469.
- Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), 42–261.
- 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), 1545–72.
- Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 271–292). New York: Guilford.
- 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.
- 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.
- 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.
- Bourdieu, P., & Passeron, J.-C. (1977). *Reproduction in education, culture and society*. London: Sage.

- Brady, K., & Woolfson, L. (2008). What teacher factors influence their attributions for children's difficulties in learning? *British Journal of Educational Psychology*, 78(4), 527-544.
- Brownell, M. T., & Pajares, F. (1999). Teacher efficacy and perceived success in mainstreaming students with learning and behavior problems. *Teacher Education and Special Education*, 22(3), 154-164.
- Burgess, D., Van Ryn, M., Dovidio, J., & Saha, S. (2007). Reducing racial bias among health care providers: Lessons from social-cognitive psychology. *Journal of General Internal Medicine*, 22(6), 882-887.
- Calarco, J. M. (2011). "I need help!" Social class and children's help-seeking in elementary school. *American Sociological Review*, 76(6), 862-882.
- Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, 16(4), 330-350.
- Card, D. and Giuliano, L. (2015). *Can Universal Screening Increase the Representation of Low Income and Minority Students in Gifted Education? Working Paper*. 21519. National Bureau of Economic Research.
- Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology*, 57(4), 278-287.
- Cho, I. (2012). The effect of teacher–student gender matching: Evidence from OECD countries. *Economics of Education Review*, 31(3), 54-67.
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: the struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359-378.
- Crandall, C. S., & Schaller, M. (Eds.). (2005). *The social psychology of prejudice: Historical and contemporary perspectives*. Seattle, WA: Lewinian Press.
- Croizet, J.-C. (2008). The pernicious relationship between merit assessment and discrimination in education. In G. Adams, M. Biernat, N. R. Branscombe, C. S. Crandall, & L. S. Wrightsman (Eds.), *Decade of Behavior. Commemorating Brown: The social psychology*

- of racism and discrimination* (pp. 153-172). Washington, DC, US: American Psychological Association.
- Croizet, J. C., Goudeau, S., Marot, M., & Millet, M. (2017). How do educational contexts contribute to the social class achievement gap: documenting symbolic violence from a social psychological point of view. *Current Opinion in Psychology, 18*, 105-110.
- Darnon, C., Dompnier, B., Delmas, F., Pulfrey, C., & Butera, F. (2009). Achievement goal promotion at university: Social desirability and social utility of mastery and performance goals. *Journal of Personality and Social Psychology, 96*(1), 119-134.
- Dauber, S. L., Alexander, K. L., & Entwisle, D. R. (1996). Tracking and transitions through the middle grades: Channeling educational trajectories. *Sociology of Education, 290*–307.
- De Boer, H., Bosker, R. J., & van der Werf, M. P. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology, 102*(1), 168-179.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95*(2), 158–165.
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources, 42*(3), 528-554.
- Dee, T., & Gershenson, S. (2017). *Unconscious Bias in the Classroom: Evidence and Opportunities*. Mountain View, CA: Google Inc. Retrieved from <https://goo.gl/O6Btqi>.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5-18.
- Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in intergroup perception* (pp. 317–344). Orlando, FL: Academic Press.
- Diris, R.E.M. (2012). *The economics of the school curriculum* Maastricht: Datawyse /Universitaire Pers Maastricht.

- Dittmann, A. G., & Stephens, N. M. (2017). Interventions aimed at closing the social class achievement gap: changing individuals, structures, and construals. *Current Opinion in Psychology, 18*, 111-116.
- Dollmann, J. (2015). Less choice, less inequality? A natural experiment on social and ethnic differences in educational decision-making. *European Sociological Review, 32*(2), 203-215.
- Dompnier, B., Pansu, P., & Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, halo effect and internal attributions. *European Journal of Psychology of Education, 21*(2), 119-133.
- Dornbusch, S. M., Glasgow, K. L., & Lin, I. C. (1996). The social structure of schooling. *Annual Review of Psychology, 47*(1), 401-429.
- Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping, and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *SAGE handbook of prejudice, stereotyping, and discrimination* (pp. 3–28). London: Sage.
- Dovidio, J. F., Ten Vergert, M., Stewart, T. L., Gaertner, S. L., Johnson, J. D., Esses, V. M., ... & Pearson, A. R. (2004). Perspective and prejudice: Antecedents and mediating mechanisms. *Personality and Social Psychology Bulletin, 30*(12), 1537-1549.
- Driessen, G. (2006). De totstandkoming van de adviezen voortgezet onderwijs: Invloeden van thuis en school. *Pedagogiek, 25*(4), 279–298.
- Driessen, G. (2015). Teacher ethnicity, student ethnicity, and student outcomes. *Intercultural Education, 26*(3), 179–191.
- Driessen, G., Slegers, P., & Smit, F. (2008). The transition from primary to secondary education: Meritocracy and ethnicity. *European Sociological Review, 24*(4), 527–542.
- Dunkake, I., & Schuchart, C. (2015). Stereotypes and teacher characteristics as an explanation for the class-specific disciplinary practices of pre-service teachers. *Teaching and Teacher Education, 50*, 56-69.
- Duru-Bellat, M., & Tenret, E. (2009). L'emprise de la méritocratie scolaire: quelle légitimité?. *Revue Française de Sociologie, 50*(2), 229-258.

- Elhoweris, H., Mutua, K., Alsheikh, N., & Holloway, P. (2005). Effect of children's ethnicity on teachers' referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education, 26*(1), 25–31.
- Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski & Y. Trope (Eds.), *Dual process theories of the social mind*. New York, NY: Guilford.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357–411). Boston: McGraw-Hill.
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (Unpublished). A meta-analysis of change in implicit bias. Retrieved from <https://osf.io/b5m97/?show=revision>
- Fox, L. (2015). Seeing potential: The effects of student–teacher demographic congruence on teacher expectations and recommendations. *AERA Open, 2*(1).
- Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology, 41*, 1–12.
- Gao, W., & Mager, G. (2011). Enhancing preservice teachers' sense of efficacy and attitudes towards school diversity through preparation: A case of one US inclusive teacher education program. *International Journal of Special Education, 26*(2), 92-107.
- Geerlings, J., Thijs, J., & Verkuyten, M. (2018). Teaching in ethnically diverse classrooms: Examining individual differences in teacher self-efficacy. *Journal of School Psychology, 67*, 134-147.
- Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review, 52*, 209–224.
- Glock, S., & Karbach, J. (2015). Preservice teachers' implicit attitudes towards racial minority students: Evidence from three implicit measures. *Studies in Educational Evaluation, 45*, 55-61.

- Glock, S., & Klapproth, F. (2017). Bad boys, good girls? Implicit and explicit attitudes towards ethnic minority students among elementary and secondary school teachers. *Studies in Educational Evaluation, 53*, 77-86.
- Glock, S., & Krolak-Schwerdt, S. (2014). Stereotype activation versus application: How teachers process and judge information about students from ethnic minorities and with low socio-economic background. *Social Psychology of Education, 17*(4), 589–607.
- 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.
- 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.
- Goudeau, S., & Croizet, J.-C. (2017). Hidden advantages and disadvantages of social class: How classroom settings reproduce social inequality by staging unfair comparison. *Psychological Science, 28*(2), 162–170.
- Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review, 94*(4), 945-967.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17-41.
- Grolnick, W. S., Benjet, C., Kuriwiski, C. O. and Apostoleris, N. H. (1997). Predictors of parent involvement in children's schooling. *Journal of Educational Psychology, 89*: 538–548.
- Guimond, S., Begin, G., & Palmer, D. L. (1989). Education and causal attributions: The development of " person-blame" and " system-blame" ideology. *Social Psychology Quarterly, 52*(2), 126-140.
- Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology, 111*(5), 745-769.

- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology, 101*(3), 662-670.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369-1385.
- Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher?. *Labour Economics, 15*(1), 37-53.
- Horn, D. (2009). Age of selection counts: a cross-country analysis of educational institutions. *Educational Research and Evaluation, 15*(4), 343-366.
- Hughes, J. N., Gleason, K. A., & Zhang, D. (2005). Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of School Psychology, 43*(4), 303-320.
- Inspectie van het Onderwijs. (2016). *De staat van het onderwijs. Onderwijsverslag 2014-2015*. Inspectie van het Onderwijs.
- Inspectie van het Onderwijs. (2018a). *De staat van het onderwijs. Onderwijsverslag 2016-2017*. Inspectie van het Onderwijs.
- Inspectie van het Onderwijs. (2018b). *Technisch Rapport Sectorhoofdstuk Primair Onderwijs: Staat van het Onderwijs 2016/2017*. Inspectie van het Onderwijs.
- Irizarry, Y. (2015). Selling students short: Racial differences in teachers' evaluations of high, average, and low performing students. *Social Science Research, 52*, 522-538.
- Jæger, M. M. (2011). Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data. *Sociology of Education, 84*(4), 281-298.
- Jungbluth, P. (2003). *De ongelijke basisschool*. Nijmegen: ITS.
- Jungbluth, P. (2014). *De prijs van zwak onderwijs*. Essay. UM, Maastricht.
- Jury, M., Smeding, A., & Darnon, C. (2015). First-generation students' underperformance at university: the impact of the function of selection. *Frontiers in Psychology, 6*.
- Jussim, L. (In press). Teacher expectations. To appear in *The Psychology of Classroom Learning*, (E. Anderman, & L. Anderman, eds.). Farmington Hills, MI: Thomson Gale Publishers.

- Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology*, 28, 281–388.
- 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.
- Jussim, L. J., McCauley, C. R., & Lee, Y.-T. (1995). Why study stereotype accuracy and inaccuracy? In Y.-T. Lee, L. J. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Towards appreciating group differences* (pp. 3-27). Washington, DC, US: American Psychological Association.
- Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, 109(6), 871-888.
- Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology*, 141(2), 165-173.
- Kelly, S. (2004). Do Increased Levels of Parental Involvement Account for Social Class Differences in Track Placement? *Social Science Research*, 33(4), 626–659.
- Kleen, H., & Glock, S. (2018). The roles of teacher and student gender in German teachers' attitudes towards ethnic minority students. *Studies in Educational Evaluation*, 59, 102-111.
- Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100(1), 128-161.
- 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.
- Klapproth, F., Kärchner, H., & Glock, S. (2018). Do Students' Religion and School Absences Moderate the Effect of Ethnic Stereotypes on School-Placement Recommendations? *The Journal of Experimental Education*, 86(2), 173–194.

- Krolak-Schwerdt, S., Hörstermann, T., Glock, S., & Böhmer, I. (2017). Teachers' Assessments of Students' Achievements: The Ecological Validity of Studies Using Case Vignettes. *The Journal of Experimental Education*, 1–15.
- Krolak-Schwerdt S., Pit-ten Cate I.M., Hörstermann T. (2018). Teachers' Judgments and Decision-Making: Studies Concerning the Transition from Primary to Secondary Education and Their Implications for Teacher Education. In: Zlatkin-Troitschanskaia O., Toepper M., Pant H., Lautenbach C., Kuhn C. (eds) *Assessment of Learning Outcomes in Higher Education. Methodology of Educational Measurement and Assessment*. Springer, Cham.
- Kumar, R., & Hamer, L. (2013). Preservice teachers' attitudes and beliefs towards student diversity and proposed instructional practices: A sequential design study. *Journal of Teacher Education*, 64(2), 162-177.
- Kumar, R., Karabenick, S. A., & Burgoon, J. N. (2015). Teachers' implicit attitudes, explicit beliefs, and the mediating role of respect and cultural responsibility on mastery and performance-focused instructional practices. *Journal of Educational Psychology*, 107(2), 533-545.
- Lamont, M. (2012). Towards a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38, 201-221.
- Lamont, M., Beljean, S., & Clair, M. (2014). What is missing? Cultural processes and causal pathways to inequality. *Socio-economic Review*, 12(3), 573-608.
- Lamont, M., & Lareau, A. (1988). Cultural capital: Allusions, gaps and glissandos in recent theoretical developments. *Sociological Theory*, 153–168.
- Lareau, A. (2015). Cultural knowledge and social inequality. *American Sociological Review*, 80(1), 1–27.
- Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research*, 41(1), 63-76.
- Li, Z., & Rubie-Davies, C. M. (2017). Teachers matter: expectancy effects in Chinese university English-as-a-foreign-language classrooms. *Studies in Higher Education*, 42(11), 2042–2060.

- Lucas, S. R. (2001). Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects. *American Journal of Sociology*, *106*(6), 1642–90.
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.-T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*(3), 319–350.
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, *46*(3), 235-261.
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, *65*, 48–60.
- Mizala, A., Martínez, F., & Martínez, S. (2015). Pre-service elementary school teachers' expectations about student performance: How their beliefs are affected by their mathematics anxiety and student's gender. *Teaching and Teacher Education*, *50*, 70–78.
- Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition & Emotion*, *22*(4), 553-594.
- Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences*, *113*(19), 5221-5226.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171-192.
- Oyserman, D., & Destin, M. (2010). Identity-based motivation: Implications for intervention. *The Counseling Psychologist*, *38*(7), 1001-1043.
- Paredes, V. (2014). A teacher like me or a student like me? Role model versus teacher bias effect. *Economics of Education Review*, *39*, 38-49.
- Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, *26*(6), 784-793.

- Pit-ten Cate, I. M., & Glock, S. (2018). Teachers' attitudes towards students with high-and low-educated parents. *Social Psychology of Education, 21*(3), 725-742.
- 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.
- Pit-ten Cate, I. M., Krolak-Schwerdt, S., Hörstermann, T., & Glock, S. (2017). Intervention strategies to improve the quality of teachers' judgments: changes in the accuracy of teachers' transition decisions. Presented at the EARLI 2017, Tampere, Finland.
- Podell, D. M., & Soodak, L. C. (1993). Teacher efficacy and bias in special education referrals. *The Journal of Educational Research, 86*(4), 247-253.
- Puhani, P. A. (2018). Do boys benefit from male teachers in elementary school? Evidence from administrative panel data. *Labour Economics, 51*, 340-354.
- 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.
- Reskin, B. F. (2003). Including mechanisms in our models of ascriptive inequality. *American Sociological Review, 68*(1), 1-21.
- Riley, T., & Ungerleider, C. (2008). Preservice teachers' discriminatory judgments. *Alberta Journal of Educational Research, 54*, 378-387.
- Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G., & Rosenthal, R. (2015). A teacher expectation intervention: Modelling the practices of high expectation teachers. *Contemporary Educational Psychology, 40*, 72-85.
- Sansone, D. (2017). Why does teacher gender matter? *Economics of Education Review, 61*, 9-18.
- Schellhaas, F. M., & Dovidio, J. F. (2016). Improving intergroup relations. *Current Opinion in Psychology, 11*, 10-14.
- Sherman, D. K. (2013). Self-affirmation: Understanding the effects. *Social and Personality Psychology Compass, 7*(11), 834-845.

- Schwalbe, M., Godwin, S., Holden, D., Shrock, D., Thompson, S., & Wolkomir, M. (2000). Generic Processes in the Reproduction of Inequality: An Interactionist Analysis. *Social Forces*, 79, 419-452.
- Sebastian Cherng, H.-Y. (2017). If they think I can: Teacher bias and youth of color expectations and achievement. *Social Science Research*, 66, 170-186.
- Smeding, A., Darnon, C., Souchal, C., Toczec-Capelle, M. C., & Butera, F. (2013). Reducing the socio-economic status achievement gap at university by promoting mastery-oriented assessment. *PLoS One*, 8(8), e71678.
- Soodak, L. C., Podell, D. M., & Lehman, L. R. (1998). Teacher, student, and school attributes as predictors of teachers' responses to inclusion. *The Journal of Special Education*, 31(4), 480-497.
- Souchal, C., Toczec, M. C., Darnon, C., Smeding, A., Butera, F., & Martinot, D. (2014). Assessing does not mean threatening: The purpose of assessment as a key determinant of girls' and boys' performance in a science class. *British Journal of Educational Psychology*, 84(1), 125-136.
- Smeets, E., van Kuijk, J., Driessen, G. (2014). *Handreiking bji het opstellen van het basisschooladvies*. Nijmegen: ITS.
- Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, 102(6), 1178-1197.
- Stephens, N. M., Markus, H. R., & Phillips, L. T. (2014). Social class culture cycles: How three gateway contexts shape selves and fuel inequality. *Annual Review of Psychology*, 65, 611-634.
- 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.
- Sui-Chu, E. H., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of education*, 126-141.

- Talaska, C. A., Fiske, S. T., & Chaiken, S. (2008). Legitimizing racial discrimination: Emotions, not beliefs, best predict discrimination in a meta-analysis. *Social Justice Research, 21*(3), 263-296.
- Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253-273.
- Thornton, P. H., & Ocasio, W. 2008. Institutional logics. In R. Greenwood, C. Oliver, R. Suddaby & K. Sahlin, Eds., *The Sage handbook of organizational institutionalism*: 99-129. London: Sage.
- Thys, S., & Van Houtte, M. (2016). Ethnic composition of the primary school and educational choice: Does the culture of teacher expectations matter? *Teaching and Teacher Education, 59*, 383–391.
- Tibbetts, Y., Harackiewicz, J. M., Canning, E. A., Boston, J. S., Priniski, S. J., & Hyde, J. S. (2016). Affirming independence: Exploring mechanisms underlying a values affirmation intervention for first-generation students. *Journal of Personality and Social Psychology, 110*(5), 635-659.
- Timmermans, A. C., de Boer, H., Amsing, H. T. A., & van der Werf, M. P. C. (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.
- Timmermans, A. C., Boer, H., & Werf, M. P. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education, 19*(2), 217–240.
- Timmermans, A. C., Kuyper, H., & Werf, G. (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.
- Tobisch, A., & Dresel, M. (2017). Negatively or positively biased? Dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds. *Social Psychology of Education, 20*(4), 731–752.

- Tournaki, N., & Podell, D. M. (2005). The impact of student characteristics and teacher efficacy on teachers' predictions of student success. *Teaching and Teacher Education, 21*(3), 299-314.
- 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.
- Van de Werfhorst, H. G. (2015). De achterkant van autonomie: maatwerk, sturing en ongelijkheid. *Nederlands Tijdschrift voor Onderwijsrecht en Onderwijsbeleid*, December: 286-293.
- 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*. In press.
- Van de Werfhorst, H. G. & Mijs, J. B. (2010). Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. *Annual Review of Sociology, 36*(1), 407-428.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1), 82- 96.
- Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*(6023), 1447-1451.
- Wenz, S. E., Olczyk, M., & Lorenz, G. (2016). Measuring teachers' stereotypes in the NEPS (NEPS Survey Paper No. 3). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.
- Wössmann, L., Luedemann, E., Schütz, G., & West, M. R. (2009). *School Accountability, Autonomy, and Choice around the World*. Cheltenham, UK: Edward Elgar.
- Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... & Trott, J. (2016a). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*(3), 374-391.



- Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., ... & Gomez, E. M. (2016b). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences, 113*, E3341-E3348.
- Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*(4), 981-1015.