university of groningen

University Medical Center Groningen

University of Groningen

Track recommendation bias

Timmermans, A. C.; de Boer, H.; Amsing, H. T. A.; van der Werf, M. P. C.

Published in:
British Educational Research Journal

DOI:
10.1002/berj.3470

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
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. https://doi.org/10.1002/berj.3470

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 20-10-2021

BERJ British Educational Research Journal

BERA BRITISH EDUCATIONAL RESEARCH ASSOCIATION

British Educational Research Journal
Vol. 44, No. 5, October 2018, pp. 847–874

DOI: 10.1002/berj.3470

Track recommendation bias: Gender, migration background and SES bias over a 20-year period in the Dutch context

A. C. Timmermans*, H. de Boer, H. T. A. Amsing and M. P. C. van der Werf

University of Groningen, The Netherlands

Bias in track recommendations is an important mechanism, which causes education inequity in a tracked educational system (streaming). If teacher biases in track recommendations change over time, inequity in society and in the education system may also change. We investigated changes in track recommendation bias over time for gender, immigration status and socioeconomic status (SES), based on a longitudinal empirical study of nine cohorts of Dutch students in their final year (grade 6) of primary education in the period 1995–2014. An overview of educational and societal trends was provided, alongside the empirical analysis, to explain the findings in variation over time in track recommendation bias. Results indicate that the level of track recommendations provided to the students gradually increased over time. For a similar performance, a higher track recommendation was awarded in 2014 compared to 1995. This development coincided with an increase in parental education level, the valuing of education and the introduction of lower-status pre-vocational education tracks. Track recommendation bias favouring students with a migrant background and female students decreased, which coincided with growing cultural intolerance and attention to the 'boy problem'. Bias in track recommendations related to SES appeared stable, with only small deviations from year to year. The results of this study indicate that track recommendation bias and teacher considerations are dependent on time and context.

Keywords: equal opportunities; longitudinal analysis; track recommendations

Introduction

Tracked educational systems, such as many European secondary education systems, are frequently associated with higher levels of educational inequity compared with comprehensive systems (e.g. Croxford, 1994; Hanushek & Wößmann, 2006; Brunello & Checchi, 2007). In these tracked (or streamed) systems, students transfer from comprehensive primary school classes to specific track levels, usually varying in level and content, in secondary education. In the Netherlands, a recommendation has to be made at the end of primary education on which secondary school track (or stream) is the most appropriate, given the student's aptitude. A track

This research did not receive any specific grants from funding agencies in the public, commercial or not-for-profit sectors.

*Corresponding author. University of Groningen, GION Education/Research, Grote Rozenstraat 3, 9712 TG, Groningen, The Netherlands. E-mail: a.c.timmermans@rug.nl.

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

recommendation is therefore the expression of the primary school teacher's professional and informed expectation for the student's future performance during secondary education (Inspectorate of Education, 2007; De Boer *et al.*, 2010). Track assignment decisions made by secondary schools are frequently based on several sources of information, including the primary school teacher's track recommendations and test scores. Basing track assignment decisions wholly or partly on teacher recommendations is a frequently debated practice, because several studies have indicated that teachers appear to be biased in their track recommendations, and consider not only student performance but also gender, socioeconomic status (SES) and migrant background (e.g. De Boer *et al.*, 2010; Glock *et al.*, 2015; Timmermans *et al.*, 2015; Pit-ten Cate *et al.*, 2016).

Bias in track recommendations may be an important mechanism that causes inequity in education in the long term (Lüdemann & Schwerdt, 2013). Track recommendations in the Netherlands are highly predictive of initial track assignment in secondary education, also after controlling for student achievement (De Boer *et al.*, 2010), which in turn significantly determines future educational and career options (e.g. Checchi & Flabbi, 2006; Bol & van de Werfhorst, 2013; Hopwood *et al.*, 2016). Biased track recommendations and subsequent biased placement decisions may therefore lead to unequal educational opportunities for particular categories of students, which not only remain throughout the entire educational and working careers of individuals, but might also be a mechanism by which the status quo regarding inequity in society, such as in the Netherlands, is maintained.

It is arguable that if bias in track recommendations changes over time, inequity in the educational and societal system could also change. Therefore, the stability of the mechanisms which contribute to inequity should be the focus of research, and not only the development of inequity in the later stages of a student's career (e.g. Croxford, 1994; Büchner & van der Velden, 2013; Buis, 2013). One way to study this is to look at whether the bias in track recommendations in favour or against particular groups of students is stable over a long period or fluctuates with educational and societal developments. We therefore conducted a longitudinal empirical study into bias in track recommendations for nine cohorts of Dutch students who were in their final year of primary education in the period from 1995 to 2014. We investigated whether bias in track recommendations in favour or against specific groups [according to gender, migrant background and parental education (SES)] changed over time. The findings of the empirical analyses of variation over time in track recommendation bias will be discussed against the background of the educational and societal changes that occurred in the Netherlands during this 20-year period.

Track recommendation bias stability

Research into the stability of track recommendation bias (or more generally, teacher expectation bias) over time is very scarce as it requires data collection from several generations of students, or a careful review of studies conducted within the same context over a long period. Thus far, only one study, the meta-analysis by Tenenbaum and Ruck (2007), has considered the possible instability of teacher expectation bias in the USA in the period from 1968 to 2003. Although teacher expectations were

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

higher for majority students compared with minority students in all decades, the difference in teacher expectations between these groups was dependent on the period. When comparing the difference between European-American students and minority groups in general, studies conducted in the 1980s evidenced smaller effect sizes than those conducted in the 1990s. However, with respect to the largest group of minority students, African-American students, a contrasting effect was found. For this particular comparison, studies published in the 1980s evidenced larger effect sizes than studies published during other decades (1960s, 1970s, 1990s and 2000s). Although the Black–White test score gap narrowed in the 1970s and 1980s (Grissmer *et al.*, 1998), teacher expectations differentiated the most between ethnic minority and European-American children in the 1980s. The authors could not clearly link these findings to wider social, political or educational developments.

Track recommendations may show a similar level of bias instability. Studies of track recommendation bias in the Dutch context conducted in the late 1980s indicated that students with a migrant background received on average higher track recommendations compared with Dutch students with similar performance levels (de Jong, 1987; Driessen, 1991). This was seen as positive discrimination (Driessen, 1991; Koeslag & Dronkers, 1994). Relatively high track recommendations for students with a migrant background may also be a result of the relatively high aspirations of their parents (De Boer & van der Werf, 2015). If teachers become aware of the immigrant parents' higher aspirations, they adjust their recommendations accordingly. Studies based on data collected in the late 1990s showed a different pattern: the differences between the track recommendations for Dutch students and students with a migrant background appeared far less substantial or sometimes non-existent once student performance was taken into account (Dagevos *et al.*, 2003; De Boer *et al.*, 2010). The earlier pattern of positive discrimination reappeared in the early 2000s, although it was discussed differently. The difference in track recommendations between Dutch low-SES students and low-SES students with a migrant background was interpreted as teachers recognising the developmental potential in students with a migrant background but failing to recognise the potential of Dutch students from low-SES backgrounds (Claassen & Mulders, 2003). The previous studies indicate some level of instability for track recommendation bias with respect to immigrant status.

Some indications of instability in track recommendation bias were found in the Dutch context with respect to SES, which in the educational context in the Netherlands is almost exclusively measured by the level of parental education. After students' performance and migration backgrounds were taken into account, students from high-SES families received higher track recommendations than students from low-SES families (e.g. De Boer *et al.*, 2010; Timmermans *et al.*, 2015). One of the most frequently mentioned explanations is that teachers take into account the parents' ability and resources to support their children (Ditton *et al.*, 2005). Teachers deem parents from lower-SES backgrounds to be less well equipped to assist their children with school work. Furthermore, parents from higher social classes exert more pressure on teachers to get academic track recommendations (e.g. Dronkers *et al.*, 1998), while poorly educated parents rarely object to low track

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

recommendations (Hillmert & Jacob, 2010). The differences in track recommendations between students from high and low-SES families systematically increased in the period from 2009 to 2014 (Inspectorate of Education, 2016), indicating instability in bias based on SES. This may be explained by increasing pressure from parents from higher social classes due to their better knowledge of the system and awareness of the growing importance of education.

The Dutch education system and changes between 1995 and 2014

Dutch primary education is intended for all children from age 4 (pre-kindergarten) up to and including age 12 (grade 6). In 2013, approximately 1,500,000 pupils were enrolled in 6,500 primary schools (Ministry of Education, 2014). Primary education is the same for all students and takes 8 years (Stevens *et al.*, 2011). Students are clustered into age groups, but grade retention is possible. To reduce educational inequity, the Dutch government provided additional financial resources to schools to provide additional support for students with a disadvantaged background. Between 1985 and 2007, this additional funding was based on the parents' education level combined with the parents' immigrant status. In 2007, the policy changed and the additional financial resources were solely based on the level of parental education (Claassen & Mulders, 2011; Driessen, 2012). Furthermore, since 2000, the Dutch government has endeavoured to help children get a good start in their education career by offering preschool programmes for young children with a disadvantaged background (Bronneman-Helmers, 2011).

The track recommendation each student receives in the final year of primary education represents the teacher's view of the school track (or two adjacent tracks) in which the student has the best potential to develop. A teacher's track recommendation should be based on adequate information about the student, usually student performance information, although teachers are allowed to consider other information, such as working habits and engagement. For information on student performance, about 85% of schools administer the CITO (Dutch National Institute for Educational Measurement) primary school leavers' test, a standardised test on basic subjects, designed to help teachers formulate a track recommendation for secondary education; other tests were also allowed.

Schools providing secondary education, for which students register after primary education, place each student in a track according to their scholastic aptitude (track recommendation and score on the school leavers' test). There is a total of seven track levels. The duration of a track varies between 4 and 6 years, and each track offers different access to further education. The pre-university track (the highest track) takes 6 years and is the only one that directly prepares students for university education (ISCED level 3A). Higher general secondary education is the second highest track (lasting 5 years) and prepares students for further education in higher vocational education or universities for applied sciences (ISCED level 3A). Four pre-vocational education tracks (each lasting 4 years) prepare students for further education in senior secondary vocational education, although these pre-vocational education tracks differ in level and the further educational opportunities they provide (ISCED level 2B). Finally, there is a seventh track for students who struggle with education,

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

which largely focuses on practical skills and offers limited access to further education (ISCED level 2C).

Between 1995 and 2014, only one major policy change was implemented which may have influenced track recommendations: a change in 1999 to the structure of the lower tracks in Dutch secondary education by means of the introduction of four pre-vocational secondary education tracks (vmbo). This new type of school merged two previously distinct school types: lower general secondary education and lower vocational education. The introduction of this new structure was meant to upgrade the status of the latter and to provide better preparation for senior secondary vocational education. In practice, this new organisation has led to a decline in the status of the former lower general secondary education, which was now transferred into one of the vmbo tracks, resulting in a tendency to avoid this new unpopular type (Kloosterman & de Graaf, 2010; Education Council, 2015).

Grade repetition within tracks and intermediate upward and downward mobility between tracks is possible during secondary education, as students can change tracks depending on their grades. However, the extent of intermediate downward and upward mobility is limited as, after 3 years in secondary education, 85% of students remain in the track recommended by their teachers (Inspectorate of Education, 2014). Furthermore, the accumulation of tracks is possible, since students who have graduated from one track can enrol in a higher one, although such transitions are not without thresholds, and not all the tracks permit this upward mobility (Education Council, 2010). Finally, secondary schools can offer heterogeneous classes by grouping students from different tracks, which is common, especially in the first 1 or 2 years, thus postponing definitive track assignment. In the period under study, however, there was a decline in the number of schools offering combined tracks (Education Council, 2010; Inspectorate of Education, 2016).

Finally, teaching methods were changed from standardised content-oriented education to approaches more adapted to the needs of individual learners, in which each student's activities and development of study and collaboration skills, and the application of a variety of teaching methods, are central (Hulshof *et al.*, 2015). Constructivist ideas were translated into didactic approaches, which aim to stimulate active involvement and self-regulated learning.

Developments in the positions of female students, students with migrant backgrounds and students from low-SES families

Gender. In general, the average educational track level of the Dutch population rose in the period under study, especially for girls and students with a migrant background (Fettelaar *et al.*, 2014). We observe that by the mid-1990s, girls had already reversed their initially poorer educational performance and started to outperform boys. Girls are over-represented in the higher levels in secondary and further education (CBS, 2012, 2016; Portegijs & Van den Brakel, 2016). Several explanations related to contextual and thus variable factors have been proposed to explain this over-representation, including the influence of the school, with feminisation of education being one of them, referring to an increasing number of female teachers and, as an expected result, a change in didactics. This occurred especially after 1980, which is late from

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

an international perspective, and is associated with an emphasis on collaboration in the classroom and self-regulated learning, didactics that would favour girls (Driessen & Doesborgh, 2004; Driessen & van Langen, 2013). In response to the catch-up made by girls and the perceived feminisation of education, the 'boy problem' was placed on the educational agenda, arguing for education to be adapted to better suit the needs of boys: more male role models (more male teachers) and more structure (Volman, 1999; Timmerman & Van Essen, 2004; Heemskerk *et al.*, 2012).

Migration background. The period under study was very unstable regarding immigration numbers and tolerance towards them, especially concerning Muslims. In 1994, a peak showed up in the Dutch statistics on immigration, due to the civil war in Yugoslavia and instability in Afghanistan, Iran, Iraq and Somalia (Stevens *et al.*, 2011). Numbers kept rising until 2001, but declined in the period 2001–2005 due to stricter immigration regulations. After 2005, numbers increased again due to migrant workers from Eastern Europe. In 2007, 15% of the students in Dutch primary and secondary education had a non-Western background (Gijsberts & Herweijer, 2007). The European refugee crisis started in 2013, with people arriving in particularly large numbers from Syria (CBS Statline, 2016). Students with a migrant background are not equally distributed across schools in the Netherlands, as these students are more likely to enrol in schools with a high percentage of students with a migrant background, due to family school choices and residential segregation (Sykes & Kuyper, 2013).

The multicultural society was rarely a subject of social debate until the late 1980s, even though the participation of non-Western immigrants in the job market was relatively low from the 1970s onwards, when a lot of first-generation migrant workers lost their jobs because of restructuring in the Dutch economy. The only anti-immigrant party established in the 1980s, Hans Janmaat's Centrum Party, was politically marginalised. In the 1990s, the tone of the debate about the multicultural society hardened. Although the participation of non-Western immigrants in the job market increased due to the better schooling of second-generation immigrants and the increase in participation among non-Western migrant women, questions were raised about the ability of Muslims to adapt to life in Western democracies with Western norms and values. From the beginning of the twenty-first century, intolerant views were dominated by a fear of Muslim terrorism. The European refugee crisis began in this atmosphere of declining tolerance (National Archive, 2016).

If we consider the developments in the level of education of people with migrant backgrounds, we see that they lag behind in the Netherlands. Children of non-Western migrants are over-represented in vocational tracks and under-represented in pre-university tracks (Sykes & Kuyper, 2013). However, their level of education rose faster compared with other people, especially for girls (Stevens *et al.*, 2011; Fettelaar *et al.*, 2014; National Archive, 2016). This decreasing backward position of migrant students might be explained by policies to reduce inequity and an increasing utilisation of talent, but also by positive discrimination (de Jong, 1987; Kerkhoff, 1988) or a fear among teachers of being accused of racism (Jungbluth, 1985; Stevens, 2008).

Socioeconomic status. In the Netherlands, socioeconomic background has become less decisive with regard to a person's position in society since the Second World

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

War. Society and the education system became more meritocratic than before the war, and education has become a catalyst for upward social mobility. The educational level of the population rose substantially, and education acquired a more decisive role in people's lives. As a side effect, it also became a line of social demarcation (Wolbers, 2014). Educational success has proved to be a strong predictor of people's position in the labour market, the educational opportunities enjoyed by their offspring, their living conditions, trust in politicians and happiness (Bijl *et al.*, 2011; Bol, 2015). This increase in the decisive role of education could partly explain the recent increase in the differences in opportunities between students from different socioeconomic backgrounds, since, at least for the Dutch context, parents with a higher level of education recognise the value of education and have higher aspirations for their children (De Boer & van der Werf, 2015). Education's more decisive role makes those who lag behind all the more vulnerable.

Since 1985, special financial resources have been allocated to schools for students with migrant backgrounds, and from even earlier for the substantial group of native Dutch children of low SES. However, cohort research conducted from 1988 onwards shows that the latter group was unable to benefit from these policy measures (Mulder *et al.*, 2005). The gap in scholastic achievement in primary school between this (Educational Priority) policy target group and other students has increased over the years, and currently the performance level of Dutch low-SES students even lags behind the performance level of students with migrant backgrounds (Driessen, 2012). As a consequence, relatively few students from this group advance to higher education (Vogels & Bronneman-Helmers, 2003; Guldemond & Bosker, 2006).

In several other countries, policies regarding SES are based on parental income, such as the UK Free School Meal regulations introduced in 2014 (Department for Work & Pensions, 2013) or combined measures based on income, occupation and education, such as the New Zealand decile system introduced in 1995 (The University of Auckland, 2011; NZPPTA, 2013). The Dutch policy implemented in 2007 is based solely on the level of parental education (Claassen & Mulders, 2011; Driessen, 2012). Two groups of disadvantaged students are recognised: (1) students whose parents completed only primary education and (2) students whose parents completed only one of the lowest levels of secondary education but did not continue in further education. Although parental education is strongly related to income, it is a more stable and stronger predictor of a child's school success in the Dutch context (CBS, 2017).

The current study

The overview of the literature on the stability of bias in teacher expectations in general, and track recommendations in particular, suggests that this bias is not a steady phenomenon. Several educational and societal changes also occurred in the Netherlands in the 20-year period under study, which may have influenced the track recommendations teachers gave their students. This provides the opportunity to investigate whether biases in recommendations have changed over time in the Dutch context, alongside these educational and societal changes. In the following empirical study, we explored bias stability in track recommendations in the period from 1995 to 2014. The following questions guided our research:

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

1. To what extent and in which direction did track recommendations and track recommendation bias towards specific groups of students change during the period from 1995 to 2014?
2. Can general changes in the recommended track level be associated with changes in Dutch society and education?
3. Can changes in track recommendation bias towards specific groups of students be associated with trends in Dutch society and education related to these groups?

Method

Samples for the empirical study

This study is based on nine data sets collected in the final year of primary education in the period from 1995 to 2014. These data sets were collected as part of two large-scale Dutch cohort studies: the PRIMA cohort studies with data collections every 2 years in the period from 1995 to 2005 (Van Langen & Vierke, 1996; Driessen *et al.*, 1998, 2000, 2002, 2006; Van der Veen *et al.*, 2004) and the COOL$^{5–18}$ cohort studies with data collections every 3 years in the period from 2008 to 2014 (Driessen *et al.*, 2009, 2012, 2015). Both cohort studies involved data collection from multiple primary education grades. For the specific purposes of the current study, we only selected students from the final year (grade 6) of primary education. Both cohort studies used two-stage hierarchical sampling in which schools were sampled first and then all the students in grade 6 were sampled.

Schools were encouraged to participate in more than one of the data collection waves and they could also participate in both the PRIMA and COOL$^{5–18}$ cohort studies. This implies that the data analysed in the current study are partially longitudinal at school level. The school-level identifiers in the PRIMA and COOL$^{5–18}$ studies, however, are not directly comparable. The institutions responsible for data collection for the two studies only provided us with partial links to the PRIMA and COOL$^{5–18}$ school identifiers. We therefore cannot be completely certain that all school overlap between these studies can be accounted for with these links. There are 1,490 unique schools in the combined data sets, and the overlap varied between participation in one data collection (N = 717, 48.1%) to participation in all nine (N = 8, 0.5%), with an average of 2.27 participations per school. Table 1 provides an overview of the nine data collection samples for grade 6. The sample sizes presented in the table reflect the number of students and schools with complete records for the variables of interest.

Instruments and variables

Track recommendation. The track recommendation each student receives at the end of primary education is the teacher's informed expectation of which track optimally fits a given student's potential. Track recommendations for the students were obtained from teachers by means of a questionnaire. Teachers were allowed to indicate a single or two adjacent tracks for each individual student. The

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Table 1. Overview of samples

	Data collection date	*N* (%) schools	*N* (%) students
PRIMA-1	1994/1995	367 (10.9)	5,739 (9.1)
PRIMA-2	1996/1997	373 (11.0)	5,962 (9.4)
PRIMA-3	1998/1999	390 (11.5)	7,214 (11.4)
PRIMA-4	2000/2001	428 (12.7)	8,451 (13.3)
PRIMA-5	2002/2003	430 (12.7)	8,785 (13.9)
PRIMA-6	2004/2005	423 (12.5)	8,552 (13.5)
$COOL^{5-18}$-1	2007/2008	395 (11.7)	7,561 (11.9)
$COOL^{5-18}$-2	2010/2011	316 (9.4)	5,961 (9.4)
$COOL^{5-18}$-3	2013/2014	257 (7.6)	5,178 (8.2)

recommendations were recoded on a scale from 1 to 11: this is assumed to be an interval scale. Each point corresponds to half a school track: for example, 11 is a single recommendation for the pre-university track, while 10 is a combined recommendation for the higher general secondary and pre-university tracks.

Student performance. The students' scores on the CITO school leavers' test were available for all nine data sets. The school leavers' test is a highly reliable (MAcc 0.96) standardised high-stakes test with 200–240 items administered in the final year of Dutch primary education. Every year, a new school leavers' test is developed, which is similar in structure but has different questions to previous tests. An extensive pilot test is conducted every year to ensure that similar performance is awarded similar test scores. The test comprises Dutch language, mathematics and information processing. Several school leavers' tests have been investigated in terms of differential item functioning (DIF) for students with different backgrounds (Van Schilt-Mol, 2007; Van Boxtel *et al.*, 2011), indicating that although the tests contained some items which revealed DIF, the tests overall did not advantage or disadvantage particular student subgroups. The students' scores are converted by CITO to a scale from 501 to 550 (Van Boxtel *et al.*, 2011).

Gender. Student background information was retrieved from the school administrations. A dummy variable was created for gender, with boys forming the reference group.

Socioeconomic status. Socioeconomic status was indicated by the highest completed education level of both a student's parents, or of the primary carer in single-parent families, retrieved through school administrations. This was rendered an ordinal variable with four categories: (1) primary education, (2) lower vocational education, (3) senior secondary education and (4) higher education or university. This variable was considered a continuous predictor variable in the analyses.

Migrant background. The students' migrant background was coded into three categories based on the birth countries of the students' parents: (1) native Dutch (students with two native Dutch parents), (2) mixed (students with one native

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Dutch parent and one immigrant parent) and (3) migrant (both parents from an immigrant family). Information on migrant status was retrieved from school administrations.

Measurement points. Dummy variables were created to indicate which information was retrieved from which data collection. The data collected in 1995 functioned as the reference group.

Analysis strategy

The data were analysed using three-level hierarchical models (Snijders & Bosker, 2012; Leckie, 2013) using MLwiN 2.35 software (Rasbash *et al.*, 2009), with students (level 1) nested within school cohorts (level 2) nested within schools (level 3). Separating the school and the school-cohort levels permitted us to account for dependency in the data due to the schools that participated in the data collections more than once, and to decompose the variance in track recommendations into a component which reflects differences between schools that endure over time (school level) and a component which provides an indication of measurement-to-measurement differences in track recommendations within schools (school cohorts).

An unconditional model (Model 0) with teacher recommendations as the dependent variable was estimated to investigate the size of the differences in the teacher recommendations among schools and school cohorts. Whether a significant part of the variance is associated with the school and the school-cohort level was tested by means of deviance tests. The results of these tests indicated that the variance in track recommendations is significant at both the school and the school-cohort levels [school $\chi^2(1) = 998.004$, $p < 0.001$; school cohort $\chi^2(1) = 240.195$, $p < 0.001$], which demonstrates that the three-level model is preferable for testing the stability of track recommendation bias.

In Model 1, student performance and student background variables were entered into the model to investigate the average associations between student background variables and track recommendations across all the measurement points. Grand mean centring was applied for the continuous predictor variables. Model 2 expanded the random part of the model with random slopes of the student-level predictor variables in order to investigate whether general differences in track recommendations among subgroups from Model 1 could be generalised across all schools and all school cohorts. The significance of the random slopes was determined by means of deviance tests, for the separate variables and all the variables simultaneously. These random slopes are presented in the form of coverage intervals (Leckie, 2013) for interpretation purposes. The model is extended by including measurement points and cross-level interactions between the student-level predictor variables and measurement points (Model 3). Measurement points were added as a categorical variable, with the measurement in 1995 as the reference category. The cross-level interactions help investigate whether bias is generally smaller at one measurement point compared with other measurement points.

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Results

Descriptive statistics of the nine cohorts

Table 2 describes the student background variables for the nine cohorts and Table 3 describes the student performance and track recommendations. Consistent with the general increase in the Dutch population's education level, a clear pattern emerged in the students' parents' education level, which developed positively over the 20-year study period: $\chi^2(24) = 3{,}023.25$, $p < 0.001$. The proportion of students whose parents had only completed primary education decreased from 14.8% in 1998/1999 to 6.5% in 2013/2014, and the proportion of students whose parents had completed higher education or university gradually increased from 16.9% in 1994/1995 to 35.4% in 2013/2014. The nine data sets also differ systematically in terms of the participation of students with a migrant background: $\chi^2(16) = 816.61$, $p < 0.001$, consistent with the dynamics of immigration numbers in the period under discussion. The relative participation of students with a migrant background was the greatest in the data collections from the early 2000s. Investigating the students' specific backgrounds yields no clear patterns. The largest minority student groups remained students with a Turkish (5.1–9.3%) or Moroccan (4.0–8.3%) background. As expected, no significant differences were found between the cohorts with respect to gender: $\chi^2(8) = 11.27$, $p = 0.187$. The differences in average performance and track recommendations between the data collections also appeared significant: performance $F(8, 63{,}394) = 22.60$, $p < 0.001$, $\eta_p^2 = 0.003$; track recommendations $F(8, 63{,}394) = 46.39$, $p < 0.001$, $\eta_p^2 = 0.006$.

Track recommendation bias

The results of Models 0 and 1 are presented in Table 4. Model 0 indicates that 10.7% of the variance in track recommendations is related to the school level (i.e. correlation in track recommendations between two random students from the same school) and 13.3% of the variance is associated with the combined school and school-cohort level (i.e. correlation between two random students from the same school cohort within the same school).

Model 1 added student performance and student background variables to the model. The unexplained variance in track recommendation decreased by 75.1% due to the inclusion of these variables. As expected, the students' academic performance was a significant predictor of track recommendations [$b = 0.238$, $t(63{,}397) = 425.4$, $p < 0.001$], indicating that higher track recommendations were provided across all schools and all school cohorts if the students' performance was better. Furthermore, after taking performance into account and across all measurement points, girls received higher track recommendations than boys [$b = 0.113$, $t(63{,}397) = 11.3$, $p < 0.001$], students with a migrant background received higher track recommendations than students with a Dutch background [bi-ethnic $b = 0.079$, $t(63{,}397) = 3.4$, $p < 0.001$; migrant background $b = 0.194$, $t(63{,}397) = 12.1$, $p < 0.001$] and level of parental education was positively associated with track recommendations [$b = 0.189$, $t(63{,}397) = 27.0$, $p < 0.001$].

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Table 2. Overview of samples by student demographic background

		Gender		Socioeconomic status				Migrant background		
	Total	Male	Female	Primary education	Secondary education	Vocational education	Higher education and university	Dutch	Mixed	Migrant
PRIMA-1	5,739	48.2	51.8	13.1	40.7	29.3	16.9	77.0	6.0	17.0
PRIMA-2	5,962	49.6	50.4	11.2	37.9	32.0	19.0	78.8	5.5	15.7
PRIMA-3	7,214	49.4	50.6	14.8	33.1	31.5	20.6	74.7	2.0	23.2
PRIMA-4	8,451	50.2	49.8	14.4	31.2	32.5	21.9	69.9	3.7	26.4
PRIMA-5	8,785	48.5	50.5	13.9	28.9	34.5	22.7	67.5	5.3	27.2
PRIMA-6	8,552	50.0	50.0	11.8	25.7	34.5	27.9	68.4	5.6	26.0
COOL$^{5-18}$-1	7,561	50.2	49.8	10.5	22.3	40.9	26.3	67.8	6.1	26.1
COOL$^{5-18}$-2	5,961	48.9	51.1	7.6	16.6	42.1	33.7	74.5	6.3	19.3
COOL$^{5-18}$-3	5,178	48.5	51.5	6.5	13.0	45.2	35.4	73.9	6.1	20.0

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Stability of bias in track recommendations

Model 2 is presented in Table 5. Compared with Model 1, adding random slopes to the school and school-cohort levels for each of the predictor variables resulted in improved model fit [$\chi^2(34) = 3{,}244.202$, $p < 0.001$] but barely changed the fixed part of the model. The only exception was the school-level random slope comparing mixed-migrant-background students and Dutch students, which did not improve the model fit. The full variance–covariance matrices at the school and school-cohort levels are presented in the random part of Table 5.

Table 3. Overview of samples by student performance and track recommendations

	Total	Performance		Track recommendations	
	N	*M*	SD	*M*	SD
PRIMA-1	5,739	533.22	10.51	7.04	2.47
PRIMA-2	5,962	533.14	10.17	6.98	2.58
PRIMA-3	7,214	532.79	10.31	6.84	2.74
PRIMA-4	8,451	533.03	10.40	6.91	2.78
PRIMA-5	8,785	533.28	10.11	6.91	2.88
PRIMA-6	8,552	532.50	10.50	6.96	2.93
$COOL^{5-18}$-1	7,561	533.00	10.30	7.20	2.86
$COOL^{5-18}$-2	5,961	534.56	9.51	7.50	2.75
$COOL^{5-18}$-3	5,178	533.82	10.49	7.42	2.85

Table 4. Results from Models 0 and 1

	Model 0		Model 1	
	b	SE *b*	*b*	SE *b*
Fixed part				
Intercept	7.081*	0.029	7.013*	0.017
School leavers' test			0.238*	0.001
Gender (girls)			0.113*	0.010
Migrant background (mixed)			0.079*	0.023
Migrant background (migrant)			0.194*	0.016
Socioeconomic status			0.189*	0.007
Random part				
School-level variance (3)	0.827	0.045	0.166	0.014
School-cohort-level variance (2)	0.206	0.019	0.258	0.011
Student-level variance (1)	6.716	0.039	1.509	0.009
Model fit				
−2*log likelihood	303,971.808		211,464.255	
No. schools	1,490		1,490	
No. school cohorts	3,379		3,379	
No. students	63,403		63,403	

*$p < 0.05$.

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Table 5. Results from Model 2

	Intercept		School leavers' test		Gender		Mixed		Migrant		Socioeconomic status	
	b	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*
Fixed part												
Intercept	6.994*	0.018										
School leavers' test	0.238*	0.001										
Gender (girls)	0.116*	0.011										
Migrant background (mixed)	0.073*	0.023										
Migrant background (migrant)	0.205*	0.019										
Socioeconomic status	0.197*	0.008										
Random part												
School level												
Intercept	0.169	0.017										
School leavers' test	−0.003	0.001	0.000	0.000								
Gender (girls)	0.002	0.007	−0.000	0.000	0.004	0.005						
Migrant background (mixed)												
Migrant background (migrant)	−0.026	0.012	0.000	0.001	0.007	0.006			0.023	0.013		
Socioeconomic status	−0.004	0.005	−0.001	0.000	−0.000	0.003			−0.000	0.004	0.009	0.003
School-cohort level												
Intercept	0.305	0.016										
School leavers' test	−0.010	0.001	0.001	0.000								
Gender (girls)	−0.060	0.010	−0.000	0.001	0.041	0.009						
Migrant background (mixed)	−0.024	0.017	−0.001	0.001	0.015	0.012	0.018	0.026				

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Table 5. (Continued)

	Intercept		School leavers' test		Gender		Mixed		Migrant		Socioeconomic status	
	b	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*
Migrant background (migrant)	−0.061	0.015	0.001	0.001	0.013	0.011	0.028	0.019	0.103	0.021		
Socioeconomic status	−0.036	0.006	−0.001	0.000	0.016	0.004	0.005	0.008	0.021	0.007	0.012	0.004
Student level	1.353	0.008										
Model fit												
−2*log likelihood:	208,220.053											
No. schools	1,490											
No. school cohorts	3,379											
No. students	63,403											

*$p < 0.05$.

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

When the slope variance partitions across the school and school-cohort levels are considered, it appears that much more of the variation in slope is related to student cohorts within schools (measurement-to-measurement differences) than to schools (enduring differences). The significant school-level random slope variance indicates that the general associations found in Model 1 did not adequately reflect the enduring relationships between the predictor variables and track recommendations for all schools. The association between performance and track recommendations across all measurement points is positive for all schools, but the 95% coverage interval ranges from $b = 0.196$ to $b = 0.280$. The enduring differences between schools are larger for the other predictors: 95% coverage intervals[1] for gender [−0.011, 0.243], students with a migrant background [−0.095, 0.505] and parental education [0.011, 0.383]. Similarly, the significant school-cohort-level random slope variance indicated that the general associations found in Model 1 did not adequately reflect that, within schools, the extent to which predictor variables contribute to the prediction of track recommendations differs across student cohorts. Greater variation at the school-cohort level is also apparent in the 95% coverage intervals: student performance [0.173, 0.303], gender [−0.282, 0.514], mixed-background students [−0.189, 0.335], students with a migrant background [−0.424, 0.834] and parental education [−0.017, 0.411]. These intervals indicate that the extent and direction of bias can change considerably between cohorts from the same school.

Trends in track recommendation bias

Table 6 presents the results of Model 3. The results of the main effects of the background characteristics now represent bias based on the 1995 cohort, while the cross-level interactions indicate whether bias in the later cohorts was significantly different from the 1995 cohort. The development of the track recommendations across the nine measurement points with respect to migrant background is plotted in Figure 1 and for gender in Figure 2. The track recommendations (y-axis) in these figures are model based, which implies that they represent the estimated track recommendations based on the Model 3 predictors. The lines represent the development of different student groups after controlling for the other predictor variables in the model.

The main measurement point effects indicate that students with similar performance levels on the school leavers' test received higher track recommendations in the later measurement points compared with 1995. This can be seen in Figure 1 as the development of the reference group (Dutch, male students with average achievement and average level of parental education). The development of the track recommendations of students with a mixed background followed roughly the trend of native Dutch students. However, after taking student performance into account, the differences in track recommendations between Dutch students and students with a migrant background were considerably larger in 1995 [$b = 0.603$, $t(63{,}356) = 9.88$, $p < 0.001$] than in later measurements. The difference in track recommendations between Dutch students and students with a migrant background decreased rapidly in the period from 1995 to 2001 and then continued to decrease at a lower rate to almost zero in 2014. The difference in track recommendation bias between 1995 and 2014 [$b = -0.587$, $t(63{,}356) = -6.60$, $p < 0.001$] is almost the size of the original bias in 1995.

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Table 6. Results from Model 3

	Intercept		School leavers' test		Gender		Bi-ethnic		Minority		Socioeconomic status	
	b	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*
Fixed part												
Intercept	6.713*	0.041										
School leavers' test	0.238*	0.001										
Gender (girls)	0.260*	0.034										
Migrant background (mixed)	0.070	0.070										
Migrant background (migrant)	0.603*	0.061										
Socioeconomic status	0.170*	0.022										
1997	0.060	0.053										
1999	0.082	0.053										
2001	0.234*	0.052										
2003	0.211*	0.052										
2005	0.361*	0.053										
2008	0.546*	0.055										
2011	0.448*	0.058										
2014	0.512*	0.063										
Cross-level interactions												
1997*Gender (girls)	−0.063	0.048										
1999*Gender (girls)	−0.129*	0.046										
2001*Gender (girls)	−0.240*	0.044										
2003*Gender (girls)	−0.122*	0.044										
2005*Gender (girls)	−0.104*	0.044										
2008*Gender (girls)	−0.225*	0.045										
2011*Gender (girls)	−0.126*	0.048										
2014*Gender (girls)	−0.264*	0.050										
1997*Migrant background (mixed)	0.102	0.099										
1999*Migrant background (mixed)	−0.119	0.123										
2001*Migrant background (mixed)	0.073	0.100										

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Table 6. (Continued)

	Intercept		School leavers' test		Gender		Bi-ethnic		Minority		Socioeconomic status	
	b	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*
2003*Migrant background (mixed)	0.034	0.091										
2005*Migrant background (mixed)	−0.051	0.091										
2008*Migrant background (mixed)	−0.003	0.092										
2011*Migrant background (mixed)	−0.100	0.097										
2014*Migrant background (mixed)	−0.048	0.102										
1997*Migrant background (migrant)	−0.072	0.083										
1999*Migrant background (migrant)	−0.326*	0.077										
2001*Migrant background (migrant)	−0.460*	0.074										
2003*Migrant background (migrant)	−0.480*	0.074										
2005*Migrant background (migrant)	−0.462*	0.075										
2008*Migrant background (migrant)	−0.530*	0.077										
2011*Migrant background (migrant)	−0.611*	0.085										
2014*Migrant background (migrant)	−0.587*	0.089										
1997*Socioeconomic status	0.061*	0.030										
1999*Socioeconomic status	0.041	0.029										
2001*Socioeconomic status	−0.018	0.028										
2003*Socioeconomic status	0.032	0.028										
2005*Socioeconomic status	0.036	0.028										
2008*Socioeconomic status	−0.013	0.029										
2011*Socioeconomic status	0.046	0.031										
2014*Socioeconomic status	0.024	0.033										
Random part												
School level												
Intercept	0.159	0.016										
School leavers' test	−0.004	0.001	0.000	0.000								
Gender (girls)	0.005	0.007	−0.000	0.000	0.005	0.005						
Migrant background (mixed)												

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Table 6. (Continued)

	Intercept		School leavers' test		Gender		Bi-ethnic		Minority		Socioeconomic status	
	b	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*	*b*	SE *b*
Migrant background (migrant)	−0.020	0.012	0.001	0.001	0.006	0.006			0.019	0.012		
Socioeconomic status	−0.003	0.005	−0.001	0.000	0.000	0.003			−0.001	0.004	0.009	0.003
School-cohort level												
Intercept	0.293	0.016										
School leavers' test	−0.011	0.001	0.001	0.000								
Gender (girls)	−0.053	0.010	0.000	0.001	0.033	0.009						
Migrant background (mixed)	−0.023	0.017	−0.001	0.001	0.015	0.011	0.012	0.025				
Migrant background (migrant)	−0.061	0.015	0.002	0.001	0.005	0.010	0.028	0.018	0.089	0.021		
Socioeconomic status	−0.030	0.006	−0.001	0.000	0.012	0.004	0.006	0.007	0.014	0.006	0.011	0.004
Student level	1.354	0.008										
Model fit												
−2*log likelihood:	207,894.413											
No. schools	1,490											
No. school cohorts	3,379											
No. students	63,403											

*$p < 0.05$.

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

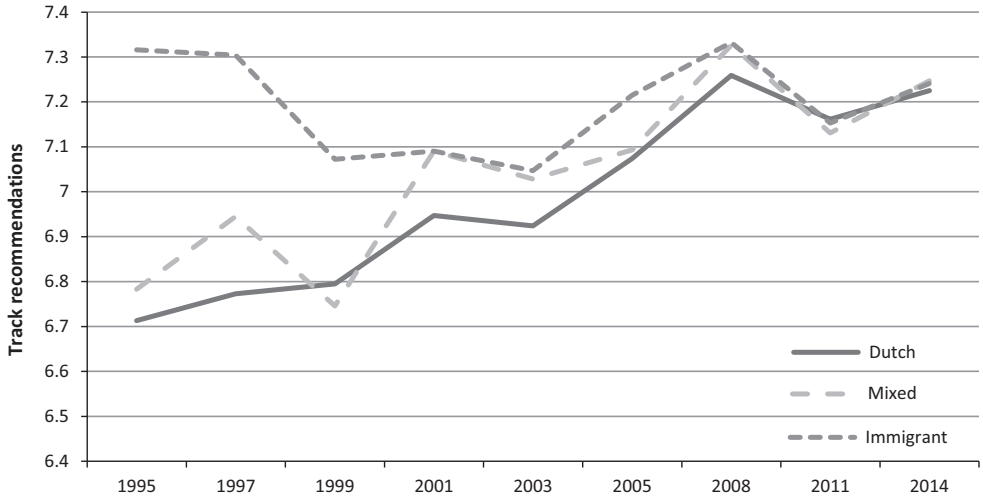

Figure 1. Development in track recommendations between 1995 and 2014 with respect to immigrant background after controlling for other student characteristics (male students, with average achievement and average parental education)

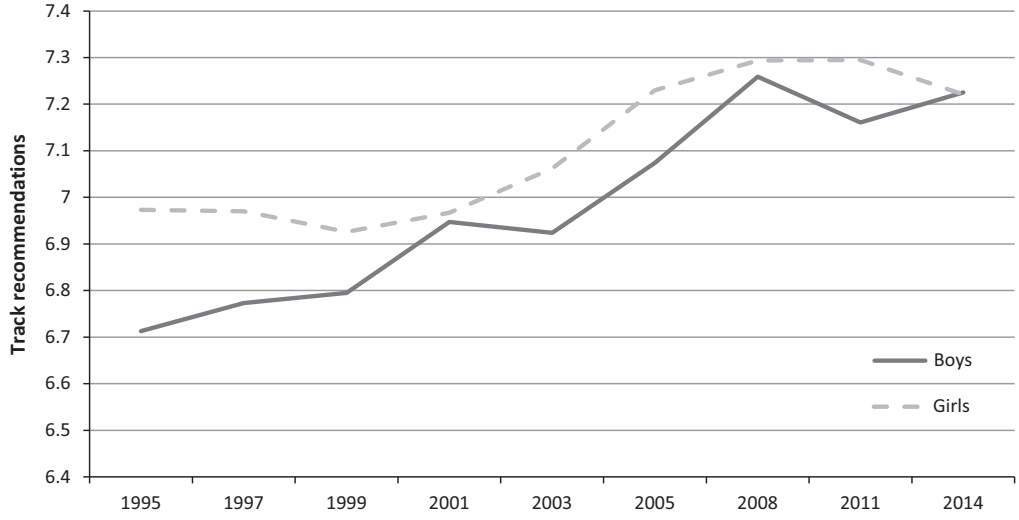

Figure 2. Development in track recommendations between 1995 and 2014 with respect to gender after controlling for other student characteristics (Dutch students, with average achievement and average parental education)

A similar development is apparent for gender. The difference in track recommendations between boys and girls was considerably greater in 1995 [$b = 0.260$, $t(63{,}356) = 4.91$, $p < 0.001$] than in later years. The decreasing rate of differences for gender was more gradual, as was the original difference between groups. While in 1995 girls received track recommendations which were on average 0.260 points higher than the track recommendations of boys, this difference was very close to zero in 2014.

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

No major differences were found between the nine data collection points with respect to SES, as becomes apparent from the small cohort cross-level interaction coefficients for socioeconomic status. Compared with 1995, the differences in track recommendations between the students from high and low-educated parents were slightly more prominent in 1997 [$b = 0.260$, $t(63{,}356) = 2.03$, $p = 0.042$], but not at later measurement points.

Conclusions and discussion

The main aim of the current study was to investigate whether track recommendations in general and bias in track recommendations favouring or discriminating against specific groups of students changed over time, and whether the changes can be explained by educational and societal developments. Indications of changes in track recommendations and track recommendation bias were indeed present in the 20-year period, as expected based on the Tenenbaum and Ruck (2007) study.

First, over the period described in this study, increasingly higher track recommendations were given once student performance was accounted for. A possible explanation for this finding can be the increasing level of parental education and associated parental aspirations for their children's education (De Boer & van der Werf, 2015). The results of the current study showed that the level of parental education (as an SES indicator) increased steadily over the nine cohorts under examination, which may have led to a growing awareness of the importance of education and the resulting aspirations these parents have for their children, which may also have resulted in an increase in pressure on teachers to recommend a high secondary school track for their children. Another plausible explanation for the steady increase in the average height of track recommendations can be found in the diminished status of pre-vocational education tracks implemented in 1999, resulting in a tendency to avoid this new unpopular type of school and send students to higher general secondary education (Kloosterman & de Graaf, 2010). However, with respect to the latter explanation, the increase in height of the track recommendations cannot be matched directly to the date of introduction of the pre-vocational education tracks.

Second, the results indicate that the positive bias towards students with a migrant background compared with students of Dutch origin largely decreased between 1995 and 2014. Contrary to the study by Tenenbaum and Ruck (2007), the current study shows a very clear trend over a longer period. The decrease in track recommendation bias with respect to immigrant status is consistent with previous studies in the Dutch context. It was not so much that the level of recommendations received by students with migrant backgrounds decreased in absolute terms, but that the level of recommendations for students of Dutch origin increased steadily over time. The initially higher recommendations for students with migrant backgrounds could be due to their parents' relatively high ambitions for them (De Boer & van der Werf, 2015) and positive discrimination based on these students' perceived developmental potential. This favourable position has disappeared. The growing intolerance towards Muslims and other minority groups might also have played a role in this effect. Another factor which may have played a role is the fact that the policies to reduce inequity stopped considering minority

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

students as a specific target group (1985–2007) and came to be based solely on parental education level.

A similar development in track recommendation bias is apparent for gender. While girls received track recommendations that were higher than boys in 1995, this difference was very close to zero in 2014. The initially higher track recommendations for girls might be explained by the teachers' perceptions that girls have better work habits and engagement in primary education (Timmermans *et al.*, 2016), better fitting the new didactic approaches implemented in secondary education in the period. The decrease in gender bias could be explained by a growing awareness of the feminisation of education and the greater attention being paid to the talents and needs of boys, referred to as the 'boy problem' (Volman, 1999). An awareness that boys are less successful in the classroom might have added to the improvement in opportunities for this group. It should be kept in mind that the differences in track recommendations between boys and girls are considerably smaller than the differences found for the other variables included in this study. This may explain why, although included in many studies, gender differences received very little attention in the track recommendation debates. However, the instability in track recommendations regarding gender indicates that these differences in recommendations also need to be monitored.

Track recommendation bias with respect to SES appeared stable over the 20-year period of this study. Compared with 1995, the differences in track recommendations between students from better and more poorly educated parents were slightly more prominent in 1997, but not at later measurement points. Given equal performance on the school leavers' tests, students from low-SES families consistently had fewer opportunities to enter the higher secondary education tracks. This finding may be an indication of the differences in positioning, knowledge and power in the educational decision-making processes between high and low-SES families (Gazeley, 2012). Differences in track recommendations may remain for several reasons, including teachers consistently taking into account parents' ability and resources to support their children (Ditton *et al.*, 2005; Böhmer *et al.*, 2017) and the limited power of low-SES parents' interactions with educational professionals (Gazeley, 2012). Regarding track recommendations, the existence of such differences has been found several times. Parents from higher social classes exert greater pressure on teachers to get recommendations for academic tracks (e.g. Dronkers *et al.*, 1998) and poorly educated parents object more rarely to lower track recommendations (Hillmert & Jacob, 2010; Korpershoek *et al.*, 2016). Our findings on the stable influence of SES on track recommendations over time, however, are inconsistent with the findings of the Dutch Inspectorate of Education (2016) that the influence of the level of parental education on track recommendations has increased lately. Differences in data, methodology (i.e. not considering the longitudinal nature of the data) and timeframe (i.e. 2009–2015) under study could explain the differences in outcomes.

Building on previous work, the current study presents evidence that track recommendations can be subject to time influences at the societal level. These indications of instability in track recommendation bias suggest that teacher considerations when formulating track recommendations change over time and are at least partly subject to contextual changes in society and education, which could imply that track recommendation bias is changeable. In addition, differences between and within schools

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

also became apparent. Bias in track recommendations for particular subgroups may not be generalised across all schools, as the bias in some schools seemed substantially greater than in others, which is consistent with previous research (Timmermans *et al.*, 2015, 2016). Furthermore, the level of bias within schools changed from one student cohort to another. These differences within and between schools are a further indication that biased track recommendations are related to time and context.

Study limitations

We used two cohort studies available in the Netherlands to examine the trends in teacher recommendations over a 20-year period. Although there was a large overlap between the two cohorts with respect to the type of information gathered and the data collection methods, there is also an inherent limitation: the link between the school-level identifiers was incomplete. Schools within the cohort studies were of course linked, and the schools which participated in the final PRIMA cohort and the first COOL$^{5-18}$ cohort were also linked. However, schools participating in earlier PRIMA cohorts, but not in the final PRIMA cohort, could not be linked to any of the COOL$^{5-18}$ cohorts. It is therefore possible that some schools may have been identified as two different schools while in fact being one. This affected slightly the hierarchy of the data set and might have influenced slightly the findings with respect to the estimated school and school-cohort-level variance, mainly due to relationships within the schools and the different measurement points remaining unnoticed.

Another study limitation is that there was only one prior achievement test available to indicate student performance levels and to determine whether the track recommendations were biased. It remains open to question whether this source, although reliable, highly predictive of student performance in secondary education and widely available in primary schools, was sufficient to calibrate teacher recommendations and whether other valid sources of information that teachers could use in forming recommendations were omitted (De Boer *et al.*, 2010; Timmermans *et al.*, 2015).

Finally, the study was structured as an empirical analysis of relevant data alongside a description of the period and potentially relevant societal and educational developments in the Netherlands. Any links between the results of the empirical analysis and the developments described should be treated carefully, as they cannot be interpreted as causal relationships.

Implications and future directions

Despite its limitations, this study's results indicate that track recommendation bias is related to time and context. This finding has several implications. First, it might explain why several researchers have interpreted the existing evidence of teacher expectations and track recommendation bias as inconsistent (McKown & Weinstein, 2008; Ready & Wright, 2011). Studies conducted in different contexts and at different times can yield different findings, and thus when reviewed result in inconsistent findings. Second, because of the effects on the initial placement in secondary education (e.g. De Boer *et al.*, 2010) and subsequent educational careers (e.g. Checchi & Flabbi, 2006; Bol & van de Werfhorst, 2013; Hopwood *et al.*, 2016), the instability in

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

track recommendation bias requires continuous monitoring. It also implies that policy changes related to track recommendations should consider the most recent evidence on bias to prevent implementation of interventions based on outdated information. Third, the instability in track recommendation bias may also indicate that even within tracked educational systems, levels of inequity may change and are not per se fully caused by the system of tracking. This, of course, adds considerable complexity to attempts to compare tracked or streamed systems with comprehensive educational systems. The current study underlines the importance of detailed investigation of selection mechanisms and their associated biases. Fourth, track recommendation bias appears very robust with respect to parental education. Without further action, teachers cannot be expected to reduce this bias just by being aware of it. It also questions the practice of taking placement decisions based—even in part—on teachers' recommendations and warrants the use of objective measures as a significant source of information for placement decisions. Finally, if bias in recommendations for specific groups of students were to persist or grow, it is not yet clear what type of interventions would be likely to reduce this bias. The variability in track recommendation bias found in this study indicates that it is time and context dependent, and also that schools may be in a position to change such bias. Further research could therefore target which characteristics of schools, classes or teachers are related to greater or lesser bias, and which interventions would best reduce teacher expectation bias.

NOTE

[1] Coverage intervals are centred around the fixed coefficient of the corresponding variable (Leckie, 2013).

References

Bijl, R., Boelhouwer, J., Cloïn, M. & Pommer, E. (2011) *De sociale staat van Nederland 2011 [The social state of the Netherlands 2011]* (Den Haag, Sociaal Cultureel Planbureau).

Böhmer, I., Gräsel, C., Krolak-Schwerdt, S., Hörstermann, T. & Glock, S. (2017) Teachers' school tracking decisions, in: D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds) *Competence assessment in education: Research, models and instruments* (Heidelberg, Springer).

Bol, T. (2015) Has education become more positional? Educational expansion and labor market outcomes, 1985–2007, *Acta Sociologica*, 58, 105–110. https://doi.org/10.1177/0001699315570918

Bol, T. & van de Werfhorst, H. G. (2013) Educational systems and the trade-off between labor market allocation and equality of educational opportunity, *Comparative Education Review*, 57, 285–308. https://doi.org/10.1086/669122

Bronneman-Helmers, B. (2011) *Overheid en onderwijsbestel. Beleidsvorming rond het Nederlandse onderwijsstelsel (1990 - 2010). [Government and educational systems. Policies regarding the Dutch educational system (1990 - 2010)]* (The Hague, Sociaal en Cultureel Planbureau).

Brunello, G. & Checchi, D. (2007) Does school tracking affect equality of opportunity? New international evidence, *Economic Policy*, 22, 781–861.

Büchner, C. & van der Velden, R. (2013) How social background affects educational attainment over time in the Netherlands, in: J. Johnson & M. Jackson (Eds) *Determined to succeed? Performance versus choice in educational attainment* (Stanford, CA, Stanford University Press), 89–115. https://doi.org/10.11126/stanford/9780804783026.001.0001

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Buis, M. L. (2013) The composition of family background: The influence of the economic and cultural resources of both parents on the offspring's educational attainment in the Netherlands between 1939 and 1991, *European Sociological Review*, 29, 593–602. https://doi.org/10.1093/esr/jcs009

CBS (2012) *De Nederlandse samenleving 2012 [The Dutch society 2012]* (Den Haag, Centraal Bureau voor Statistiek).

CBS (2016) *Nederland in 2015. Een economisch overzicht [The Netherlands in 2015. An economic overview]* (Den Haag, Centraal Bureau voor de Statistiek)

CBS (2017) *Herziening gewichten regeling primair onderwijs [Revision financial resources policy primary education]* (Den Haag, Centraal Bureau voor de Statistiek).

CBS Statline (2016) Asielverzoeken; nationaliteit, vanaf 1975 *[requests for asylum, from 1975 onwards]*. Available online at: http://statline.cbs.nl/Statweb/publication (accessed 1 November 2016).

Checchi, D. & Flabbi, L. (2006) *Intergenerational mobility and schooling decisions in Italy and Germany*. IZA Discussion Paper 2876 (Bonn, Germany).

Claassen, A. & Mulders, L. (2003) *Leerlingen na de overstap [Students after the transition]* (Nijmegen, ITS).

Claassen, A. & Mulders, L. (2011) *Een afgewogen weging? De effecten van de gewijzigde gewichtenregeling in het basisonderwijs [The balance of weights? The effects of changes in the financial support regulations to reduce inequity in primary education]* (Nijmegen, ITS).

Croxford, L. (1994) Equal opportunities in the secondary-school curriculum in Scotland, 1977–91, *British Educational Research Journal*, 20, 371–391. https://doi.org/10.1080/0141192940200402

Dagevos, J., Gijsberts, M. & Van Praag, C. (2003) *Rapportage Minderheden 2003. Onderwijs, arbeid en social–culturele integratie [Report immigrant populations 2003. Education, work, and social–cultural integration]* (Den Haag, Sociaal en Cultureel Planbureau).

de Boer, H. & van der Werf, M. P. C. (2015) Influence of misaligned parents' aspirations on long-term student academic performance, *Educational Research and Evaluation: An International Journal on Theory and Practice*, 21, 232–257. https://doi.org/10.1080/13803611.2015.1039548

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

Department for Work & Pensions (2013) *Free school meal entitlements and child poverty in England* (London, DWP).

Ditton, H., Krüsken, J. & Schauenberg, M. (2005) Bildungsungleichheit – der Beitrag von Familie und Schule, *Zeitschrift für Erziehungswissenschaft*, 8, 285–304. https://doi.org/10.1007/s11618-005-0138-x

Driessen, G. (1991) Discrepanties tussen toetsresultaten en doorstroomniveau. Positieve discriminatie bij de overgang basisonderwijs – voortgezet onderwijs? *Pedagogische Studiën*, 68, 27–35.

Driessen, G. (2012) *De doelgroepen van het onderwijsachterstandenbeleid: ontwikkelingen in prestaties en het advies voortgezet onderwijs. [The targets of achievement gap reduction policies: Developments in achievement and track recommendations]* (Nijmegen, ITS).

Driessen, G. & Doesborgh, J. (2004) *De feminisering van het basisonderwijs* (Nijmegen, ITS).

Driessen, G. & van Langen, A. (2013) Gender differences in primary and secondary education: Are girls really outperforming boys? *International Review of Education*, 59, 67–86. https://doi.org/10.1007/s11159-013-9352-6

Driessen, G., van Langen, A., Portegen, R. & Vierke, H. (1998) *Basisonderwijs: Veldwerkverslag, leerlinggegevens en oudervragenlijsten – Basisrapportage PRIMA-cohortonderzoek – Tweede meting 1996–1997* (Nijmegen, ITS).

Driessen, G., van Langen, A. & Vierke, H. (2000) *Basisonderwijs: Veldwerkverslag, leerlinggegevens en oudervragenlijsten – Basisrapportage PRIMA-cohortonderzoek – Derde meting 1998–1999* (Nijmegen, ITS).

Driessen, G., van Langen, A. & Vierke, H. (2002) *Basisonderwijs: Veldwerkverslag, leerlinggegevens en oudervragenlijsten – Basisrapportage PRIMA-cohortonderzoek – Vierde meting 2000–2001* (Nijmegen, ITS).

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Driessen, G., van Langen, A. & Vierke, H. (2006) *Basisonderwijs: Veldwerkverslag, leerlinggegevens en oudervragenlijsten – Basisrapportage PRIMA-cohortonderzoek – Zesde meting 2004–2005* (Nijmegen, ITS).

Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J. & van der Veen, I. (2009) *Cohortonderzoek COOL$^{5–18}$. Technisch rapport basisonderwijs, eerste meting 2007/08* (Nijmegen/Amsterdam, ITS/Kohnstamm Instituut).

Driessen, G., Mulder, L. & Roeleveld, J. (2012) *Cohortonderzoek COOL$^{5–18}$. Technisch rapport basisonderwijs, tweede meting 2010/11* (Nijmegen/Amsterdam, ITS/Kohnstamm Instituut).

Driessen, G., Elshof, D., Mulder, L. & Roeleveld, J. (2015) *Cohortonderzoek COOL$^{5–18}$. Technisch rapport basisonderwijs, derde meting 2013/14* (Nijmegen/Amsterdam, ITS/Kohnstamm Instituut).

Dronkers, J., van Erp, M., Robijns, M. & Roeleveld, J. (1998) Krijgen leerlingen in de grote steden en met name in Amsterdam te hoge adviezen? De relaties tussen taal- en rekenscores en advies binnen en buiten de Randstad onderzocht, *Tijdschrift voor Onderwijsresearch*, 23, 17–30.

Education Council (2010). *Vroeg of laat. [Early or late]* (The Hague, Author).

Education Council (2015). *Herkenbaar vmbo met sterk vakmanschap. [Discernable pre-vocational education with strong craftmanship]* (The Hague, Author).

Fettelaar, D., Mulder, L. & Driessen, G. (2014) *Ouderlijk opleidingsniveau en onderwijsachterstanden van kinderen. Veranderingen in de periode 1995–2011 [Parental level of education and at risk students. Changes in the period 1995–2011]* (Nijmegen, ITS).

Gazeley, L. (2012) The impact of social class on parent professional interaction in school exclusion processes: Deficit or disadvantage?, *International Journal of Inclusive Education*, 16, 297–311. https://doi.org/10.1080/13603116.2010.489121

Gijsberts, M. & Herweijer, L. (2007) Allochtone leerlingen in het onderwijs [Ethnic minorities in education], in: J. Dagevos & M. Gijsberts (Eds) *Jaarrapport integratie 2007 [Yearly report Integration 2007]* (Den Haag, Sociaal en Cultureel Planbureau), 102–130.

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, 169–188. https://doi.org/10.1007/s10212-014-0237-2

Grissmer, D. W., Flanagan, A. & Williamson, S. (1998) Why did Black test scores rise in the 1970s and 1980s? in: C. S. Jencks & M. Phillips (Eds) *The Black–White test score gap* (Washington, D.C., Brookings Institution), 182–228.

Guldemond, H. & Bosker, R. J. (2006) *Onderwijsachterstanden en ontwikkelingen in leerprestaties [Educational deprivation and developments in learning achievements]* (Groningen, GION).

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

Heemskerk, I. M. C. C., van Eck, E., Kuiper, E. & Volman, M. (2012) *Succesvolle onderwijsaanpakken voor jongens in het voortgezet onderwijs* (Amsterdam, Kohnstamm Instituut) (rapport 878). http://www.kohnstamminstituut.uva.nl

Hillmert, S. & Jacob, M. (2010) Selections and social selectivity on the academic track: A life-course analysis of educational attainment in Germany, *Research in Social Stratification and Mobility*, 28, 59–76. https://doi.org/10.1016/j.rssm.2009.12.006

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

Hulshof, H., Kwakernaak, E. & Wilhem, F. (2015) *Geschiedenis van het talenonderwijs in Nederland. Onderwijs in de moderne talen van 1500 tot heden* (Groningen, Passage).

Inspectorate of Education (2007). *Teacher recommendations in the picture. [Onderadvisering in beeld]* (Utrecht, Author).

Inspectorate of Education (2014) *The quality of track recommendations [De kwaliteit van het basisschooladvies]* (Utrecht, Dutch IoE).

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Inspectorate of Education (2016) *The state of education. Education report 2014–2015 [De staat van het onderwijs. Onderwijsverslag 2014–2015]* (Utrecht, Dutch IoE).

de Jong, M. J. (1987) *Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs* (Lisse, Swets & Zeitlinger B.V.).

Jungbluth, P. (1985) *Verborgen discriminatie: Leerlingbeeld en onderwijsaanbod op de basisschool [Hidden discrimination: Views on students and educational provisions in primary education]* (Nijmegen, ITS).

Kerkhoff, A. (1988) *Taalvaardigheid en schoolsucces: De relatie tussen taalvaardigheid nederlands en schoolsucces van allochtone en autochtone leerlingen aan het einde van de basisschool [Language skills and educational success: The relationship between Dutch language skills and success at the end of primary school for ethnic minority and majority pupils]* (Lisse, Swets & Zeitlinger B.V.).

Kloosterman, R. & de Graaf, P. M. (2010) Non-promotion or enrolment in a lower track? The influence of social background on choices in secondary education for three cohorts of Dutch children, *Oxford Review of Education*, 36(3), 363–384. https://doi.org/10.1080/03054981003775244

Koeslag, M. & Dronkers, J. (1994) Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs, *Tijdschrift voor Onderwijsresearch*, 19, 240–258.

Korpershoek, H., Beijer, C., Spithoff, M., Naaijer, H. M., Timmermans, A. C., vanRooijen, M., Vugteveen, J. & Opdenakker, M. -C. (2016). *Overgangen en aansluitingen in het onderwijs: Deelrapportage 1: reviewstudie naar de po-vo en de vmbo-mbo overgang. [Transitions and alignment in education: Report part 1: Review study of the po-vo and vmbo-mbo transitions]* (Groningen, GION Onderwijs/Onderzoek).

Leckie, G. (2013). *Three-level multilevel models: Concepts. LEMMA VLE Module 11* (Bristol, Centre for Multilevel Modelling, University of Bristol).

Lüdemann, E. & Schwerdt, G. (2013) Migration background and educational tracking. Is there a double disadvantage for second-generation immigrants? *Journal of Population Economics*, 26, 455–481. https://doi.org/10.1007/s00148-012-0414-z

McKown, C. & Weinstein, R. S. (2008) Teacher expectations, classroom context, and the achievement gap, *Journal of School Psychology*, 46, 235–261. https://doi.org/10.1016/j.jsp.2007.05.001

Ministry of Education (2014) *Key figures 2009–2013: Education, culture, science* (The Hague, Dutch MoE).

Mulder, L., Roeleveld, J., Van der Veen, I. & Vierke, H. (2005) *Onderwijsachterstanden tussen 1988 en 2002. Ontwikkelingen in basis- en voortgezet onderwijs [Educational deprivation between 1988 and 2002. Developments in primary and secondary education]* (Nijmegen, ITS).

National Archive (2016) Een samenleving in beweging. Nederland tussen 1976 en 2005 in 200 trends en 100 hotspots *[A society in progress. The Netherlands between 1976 and 2005 in 200 trends and 100 hotspots]*. http://www.nationaalarchief.nl/sites/default/files/docs/een_samenleving_in_beweging_trendanalyse.pdf

NZPPTA (2013) *NZ schools: The decile system* (Wellington, NZ, PPTA).

Onderwijsraad (2010) *Vroeg of laat. Advies over de vroege selectie in het Nederlandse onderwijs [Early or late tracking]* (The Hague, Onderwijsraad).

Onderwijsraad (2015) *Herkenbaar vmbo met sterk vakmanschap [Distinctive preparatory secondary vocational education (vmbo) with strong end-to-end learning pathways]* (The Hague, Onderwijsraad). https://www.onderwijsraad.nl/english/publications/2015/item7291

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, 225–243. https://doi.org/10.1007/s10212-015-0259-4

Portegijs, W. & Van den Brakel, M. (2016) *Emancipatiemonitor 2016* (DenHaag, SCP/CBS).

Rasbash, J., Charlton, C., Browne, W. J., Healy, M. & Cameron, B. (2009) *MLwiN Version 2.1* (Bristol, Centre for Multilevel Modelling, University of Bristol).

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

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, 335–360. https://doi.org/10.3102/0002831210374874

Snijders, T. A. B. & Bosker, R. J. (2012) *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, 2nd edn (London, Sage).

Stevens, P. A. J. (2008) Exploring pupils' perceptions of teacher racism in their context: A case study of Turkish and Belgian vocational education pupils in a Belgian school, *British Journal of Sociology of Education*, 29(2), 175–187. https://doi.org/10.1080/01425690701837539

Stevens, P. A. J., Clycq, N., Timmerman, C. & van Houtte, M. (2011) Researching race/ethnicity inequality in the Netherlands: A critical review of the research literature between 1980 and 2008, *British Educational Research Journal*, 37, 5–43. https://doi.org/10.1080/01411920903342053

Sykes, B. & Kuyper, H. (2013) School segregation and the secondary-school achievements of youth in the Netherlands, *Journal of Ethnic and Migration Studies*, 39(10), 1699–1716. https://doi.org/10.1080/1369183x.2013.833707

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, 689–705. https://doi.org/10.1037/0022-0663.99.2.253

The University of Auckland (2011) The great decile debate, *Te Kuaka*, 3, 4–8.

Timmerman, G. & Van Essen, M. (2004) De mythe van het 'vrouwengevaar' [The myth of the 'women peril'], *Pedahogiek*, 24(1), 57–71.

Timmermans, A. C., Kuyper, H. & van der Werf, M. P. C. (2015) Accurate, inaccurate or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, 85, 459–478. https://doi.org/10.1111/bjep.12087

Timmermans, A. C., de Boer, H. & van der Werf, M. P. C. (2016) An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes, *Social Psychology Education*, 19, 217. https://doi.org/10.1007/s11218-015-9326-6

van Boxtel, H. W., Engelen, R. & de Wijs, A. (2011).*Wetenschappelijke verantwoording van de Eindtoets Basisonderwijs 2010 [Scientific report on the School Leavers test 2010]* (Arnhem, CITO).

Van der Veen, I., van der Meijde, A. & Ledoux, G. (2004) *School-en klaskenmerken Basisonderwijs: Basisrapportage Prima-cohortonderzoek, vijfde meting 2002–2003* (Amsterdam, SCO-Kohnstamm Instituut van de Faculteit der Maatschappijen Gedragswetenschappen, Universiteit van Amsterdam).

van Langen, A. & Vierke, H. (1996). *PRIMA-cohortonderzoek. Deel 1 veldwerkverslag basisonderwijs* (Nijmegen, Instituut voor Toegepaste Sociale Wetenschappen).

van Schilt-Mol, T. M. M. L. (2007) *Differential item functioning en itembias in de Cito-Eindtoets Basisonderwijs* (Amsterdam, Aksant).

Vogels, R. & Bronneman-Helmers, R. (2003) *Autochtone achterstandsleerlingen: Een vergeten groep [Native Dutch children at risk: A forgotten group]* (Den Haag, Sociaal en Cultureel Planbureau).

Volman, M. (1999) Girl power en boys problems. Het onderwijssucces van meisjes in de Jaren negentig [Girl power and boys problems. Educational success of girls in the 1990s] *Tijdschrift voor genderstudies*, 2(2), 49–59.

Wolbers, M. H. J. (2014) *De twee gezichten van het onderwijs [The two faces of education]* (Nijmegen, ITS).

© 2018 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.