



Achievement development is less important for school-placement recommendations when students are stereotyped

Florian Klapproth¹ · Birthe Doreen Fischer²

Received: 16 April 2020 / Accepted: 16 September 2020 / Published online: 29 September 2020
© The Author(s) 2020

Abstract

With this study we examined with a sample of $N=102$ primary-school teachers whether their use of information about students' achievement development for placement recommendations depended on student ethnicity. We applied student vignettes to mimic real students, and orthogonally varied student ethnicity, their GPA development, suggested by their last two school reports in primary school, and their grand mean of grades. We found that students were more likely to be recommended for the highest track when their grand mean of grades indicated higher achievements and when their GPA improved rather than declined. Moreover, we found strong evidence that teachers applied ethnic stereotypes when making school-placement recommendations. If the students fitted an ethnic stereotype, teachers tend to ignore information about their achievement development, whereas if the students did not fit an ethnic stereotype, teachers' judgments were rather based on all information that was provided about the students. Hence, achievement development was less important for school-placement recommendations when students were stereotyped than when they were not stereotyped. Implications of the results were discussed.

Keywords Achievement development · School-placement recommendations · Teachers · Social stereotypes · Ethnicity · Experiment

1 Introduction

School-placement is a form of ability grouping that is used in Germany and some other European countries (e.g. Austria, Switzerland, Luxembourg) to assign primary-school students to different school tracks in secondary school by the students'

✉ Florian Klapproth
florian.klapproth@medicalschooll-berlin.de

¹ Medical School Berlin, Calandrellistrasse 1-9, 12247 Berlin, Germany

² University of Trier, Trier, Germany

ability level. The ability level is gauged by students' school grades, which are the most important and decisive sources of information for teachers when they have to opt for a school track that a student should attend in secondary school. Recent evidence (Caro et al. 2009; Klapproth and Fischer 2019) suggests that teachers do also regard the students' academic development for placement decisions. Students who improve in achievement usually are placed in a higher track than students who deteriorate. However, teachers might neglect students' academic development when salient attributes of the students would allow for stereotyping them. Stereotyping students might be based on physical, social, or behavioral attributes (Fiske 1998). The present study examined whether students' ethnicity contributes to stereotyping and results in biased school-placement decisions.

2 Theoretical and empirical background

2.1 Secondary school in Germany

In Germany, school-placement is a permanent school administrative arrangement that leads to restrictions on students' graduation and career paths. Traditionally, three separate tracks constitute the German secondary school. The lowest track ("Hauptschule") is a lower secondary education program and offers students with major learning difficulties and below-average achievement profiles qualifications for vocational training. The intermediate track ("Realschule") provides students general education and vocational-training courses and allows for attending the highest track when completed successfully. However, only the highest track ("Gymnasium") offers students with rather flawless achievement profiles the qualification for university entrance, providing they successfully accomplish this track. The German secondary school has undergone several reforms, leading to the rise of comprehensive schools and permeable tracks (Becker et al. 2016), which makes the German secondary school more similar to the secondary school of most European countries. Nevertheless, the highest track is still the most favorable track for students aiming at post-secondary university education (Nikolai 2019).

At the end of primary school, which is—depending on the federal state—either the 4th school year or the 6th school year, teachers will recommend the school track a student should attend in secondary school. In some federal states, the teachers' recommendation is mandatory (e.g. in Bavaria), whereas in other states, the students' parents make the final decision (e.g. in Berlin). After a student has been placed in a certain track, changing the track is rather unlikely (Bellenberg and Forell 2012).

German transition regulations intend that teachers value students' achievements as the major factor when making school-placement recommendations (KMK 2010). The students' achievements considered for placement recommendations are mainly represented by their grades given on the second to last school report of primary school, although in some federal states (e.g. Berlin) the grades of two successive school reports are used. Moreover, teachers use working habits and social behavior that are mentioned in the school report for track recommendations. School grades in Germany vary between 1 and 6, with lower scores meaning higher achievements.

School grades are based on predetermined educational standards representing the knowledge and skills that should be mastered at each stage of the school system. However, school grades are subject to bias (Tobisch and Dresel 2017), and they are therefore not as objective as they might appear.

2.2 Effects of students' academic development on placement decisions

When teachers in Germany recommend students for one of the different tracks in secondary school, they mainly resort to school grades given in the school reports as indicators of the students' academic status (Arnold et al. 2007). However, recent studies have shown that teachers may also use data of academic growth (positive or negative) when it comes to placement recommendations. Caro et al. (2009), for example, demonstrated that students growing more rapidly in their mathematics skills, measured by standardized achievement tests, were more likely to get a high-track recommendation than students with a slower degree of growth. Likewise, Klapproth and Fischer (2019) found with preservice teachers that students who improved their school marks in the last year of primary school were more than twice as likely to be recommended for the highest track in secondary school than those whose school marks deteriorated, even when their grand mean of grades was the same.

Effects of students' achievement development on teachers' school-placement decisions could be explained by assuming that teachers (or preservice teachers) are influenced by expectations they have based on information regarding students' prior development. Cooper and coworkers (Cooper 1985; Good and Brophy 2003) suggested the term "sustaining expectation effects" to describe the phenomenon that teachers expect students to continue to perform with respect to previously established achievement patterns. According to the hypothesis of sustaining expectation effects, students who performed well in the past are expected to perform well in the future, whereas students who performed poorly in the past are expected to perform poorly in the future (Cooper et al. 1976; Rolison and Medway 1985).

2.3 Effect of students' ethnicity on placement decisions

As has been shown in a variety of studies, placement decisions are not only determined by students' achievements, but are also affected by variables of the students' social characteristics (Baumert and Schümer 2002; Glock et al. 2013a; Stubbe and Bos 2008; Tiedemann and Billmann-Mahecha 2007). In particular, students' ethnicity is one of the factors that does affect teachers' decisions regarding track orientation in Germany (e.g. Ditton et al. 2005) and in other European countries (e.g. Darmody et al. 2014; Pásztor 2010; Sneyders et al. 2018). Several studies in Germany have shown that the probability of receiving a high-track recommendation is significantly lower for immigrant students than for native students, with native students being twice (Kristen 2006; Lintorf et al. 2008) or three times (Stubbe 2009) as likely to be placed at the higher level than immigrant students.

Some authors of studies conducted in order to reveal determinants of school-placement decisions suggest that a plain reason for differences in the likelihood of receiving a high-track recommendation between immigrant and native students is their current achievement in school (e.g. Dollmann 2010; Tiedemann and Billmann-Mahecha 2007). Students with lower school marks or lower scores obtained in standardized achievement tests are less likely to be recommended for the highest track. Since immigrant students show on average lower school marks and lower test scores than their native peers (Dollmann 2010; Gresch 2012), they consequently are less often placed in the higher tracks. In addition, even when achievement variables were controlled for, some studies still show an effect of migration status on placement decisions (Dumont et al. 2014; Klapproth et al. 2013; Lintorf et al. 2008). Moreover, when randomized controlled studies instead of correlational field studies were applied, effects of migration status became even more evident. For instance, in two experimental studies Klapproth et al. (2018) presented preservice teachers with vignettes imitating students' school reports. When these vignettes were accompanied by the name of a German student, the participants were more likely to recommend the student for the highest track than when they were accompanied by a Turkish name, even if the students' reported school grades were identical. Similar results were obtained in experimental studies with inservice teachers (Glock et al. 2013b; Riley and Ungerleider 2008) and with different judgment criteria (Glock 2016; Kleen and Glock 2018).

2.4 Teachers' stereotypes as an explanation for the effect of ethnicity on placement decisions

In educational research, social stereotypes are discussed as factors influencing teacher judgments (e.g. Jussim and Harber 2005). A stereotype is defined as a belief that members of a particular group (e.g. men, women, minorities) have certain attributes or traits (Wilson et al. 2000). Ethnic stereotypes, particularly if the targeted people are students, are often related to achievement (Erensi and Adanli 2004; Glock and Krolak-Schwerdt 2013; Peterson et al. 2016). Negative correlations between students' ethnicity and their achievement in school have reinforced achievement-related stereotypes (Dee 2005; Marx and Stanat 2012). Since immigrant students (particularly those with a Turkish background) have been found to underperform relative to native students in the German PISA study (Stanat et al. 2010), it is reasonable to assume that teachers in Germany expect immigrant students to show lower achievements than their native German peers. Correspondingly, Tenenbaum and Ruck (2007) reported in a meta-analysis that teachers had more positive expectations for the ethnic-majority than for ethnic-minority students and that these effects were even greater in primary school.

According to Fiske's and Neuberg's continuum model of impression formation (Fiske and Neuberg 1990), people almost instantly categorize individuals on the basis of their salient attributes. Categories enable people to judge quickly and efficiently without engaging in much effortful thought (Macrae et al. 1995). Attributes like gender or ethnicity are likely to elicit social categories, or social stereotypes,

respectively (Fiske 1998). Once people have categorized an individual, they automatically tend to think and behave toward that individual in a stereotypic manner (Fiske et al. 2018). Whether people stay at this level of categorization or move their attention to further attributes of the individual to be judged, depends on how much these attributes confirm the category (Fiske and Neuberg 1990). In case of disconfirming attributes, people may try to recategorize the individual, which means that they try to find a new, better fitting category for this individual. Alternatively, they may even integrate each attribute into an overall assessment. In this stage of the impression formation process, the initial category does not vanish, but becomes itself another attribute that contributes to the overall impression (Fiske et al. 2018).

If ethnic stereotypes are activated, teachers should expect students' further achievements in line with their stereotypes. That is, teachers who are to judge a student who fits the stereotype of an immigrant student would be more likely to expect him or her to be low-achieving and lazy (Baur and Ossenbreg 2016) and showing lower increments of achievement in the future, compared to a student fitting a stereotype of a native student who presumably is expected to be rather high-achieving and hard-working (Keller 1991).

The probability that a stereotype is activated depends on context information that corroborates the stereotype (Casper et al. 2010). For instance, when teachers are to judge an immigrant student who has shown rather low achievements, activation of the ethnic stereotype of a low-achieving student might be facilitated. However, if teachers are presented with information regarding prior achievement of students that contradicts the prevailing ethnic stereotype, the probability that the ethnic stereotype would guide teacher's decisions is lowered (Casper et al. 2010). For example, when the to-be-judged immigrant student has shown rather high achievements, activation of the same ethnic stereotype might be inhibited to a certain degree. Likewise, if teachers are presented with a native student showing low achievement, the prevailing stereotype of a high-achiever might also be restrained.

Inhibition of the activation of stereotypes should consequently result in more attribute-based judgments (Fiske and Neuberg 1990). If teachers are presented with an immigrant student showing rather high achievements, hence contradicting the ethnic stereotype, they should be more likely to consider the student's achievement development as a further attribute than teachers who are to judge an immigrant student fitting the ethnic stereotype. Correspondingly, if a teacher is judging a native student showing low achievements, his or her achievement development should also be taken into account for placement decisions, whereas teachers judging a native student with rather high achievement would presumably ignore information about achievement development.

2.5 Research question and hypotheses

To the authors' knowledge, this is the first study that systematically investigated the conditions under which teachers use information about students' achievement development for making placement decisions. The extent to which teachers make use of development information might depend on the degree to which

they stereotype students based on their social attributes. Therefore, we examined whether the effect of students' achievement development on teachers' school-placement decisions differed depending on their assumed ethnicity. Thus, the present study aimed to extend knowledge obtained from both studies investigating the effect of achievement development (e.g. Klapproth and Fischer 2019), and studies examining the effect of ethnicity on teachers' placement decisions (e.g. Glock et al. Glock 2013a; Sneyders et al. 2018). According to the continuum model of impression formation (Fiske et al. 2018; Fiske and Neuberg 1990), teachers should use ethnic stereotypes to judge students if the students' ethnicity is salient and if their previous achievement confirms the ethnic stereotype. Our study was guided by the following hypotheses.

Since grades are the most important information for teachers' predictions of students' future achievement (Arnold et al. 2007), we assumed that teachers would use the mean of the grades for making their placement decisions. Because in Germany grades range on a scale from 1 to 6, where 1 means "very good" and 6 means "insufficient", we expected that students with a lower mean of all grades would be more likely to be recommended for the highest track than students with a higher mean of all grades. Moreover, we assumed that the probability of a highest-track recommendation would depend on the achievement development of the students. Students who improved their GPA between two school reports should have a higher chance of receiving a recommendation for the highest track than students who declined in their GPA. Furthermore, we expected an effect of the ethnicity of the students on teachers' placement decisions. In particular, we hypothesized that teachers would be prone to recommend German students more frequently for the highest track than students with a Turkish background. Thus, we expected three additive effects to occur: a main effect of the grand mean of grades, a main effect of achievement development, and a main effect of students' ethnicity.

In addition to these main effects, we supposed that students' achievement development would play a different role for teachers' judgments depending on whether the ethnic stereotype about a student is confirmed or refuted by the students' GPA. Confirmation of the ethnic stereotype would be the case if a native student shows rather high achievement, or an immigrant student shows rather low achievement. If the stereotype is confirmed, students' achievement development should affect placement decisions only to a relatively small degree. Conversely, disconfirmation of the ethnic stereotype would be the case if a native student exhibits rather low, or an immigrant student shows rather high achievement. If the stereotype is not confirmed, students' achievement development should affect placement decisions to a larger degree, because teachers should consider all available information about the students. Thus, we expected smaller differences in the probability of highest-track recommendations between improving and declining students when students fit the ethnic stereotype than when students don't fit the stereotype. Moreover, we hypothesized that this effect should be moderated by student ethnicity. Whereas with German students, the difference in placement probabilities between improving and declining students should be lowest for rather high achieving students, the reverse should be the case when the students had a Turkish background.

3 Method

3.1 Participants

We expected a medium effect (around $d=0.50$, which translates to an odds ratio of 2.72 in logistic regression; Borenstein et al. 2009) of students' characteristics on teacher judgments to occur, since main effects of student attributes on preservice teachers' judgments have been shown to be on average $d=0.67$ (Klapproth and Fischer 2019). We conducted a power analysis using G*Power 3.1 (Faul et al. 2009) and we assumed that the event rate under H_0 is $p=.5$. When prespecifying $\alpha=.05$, $\beta=.70$, and the estimate of the squared multiple correlation with the covariates to be $R^2=0$ (since all covariates were noncorrelated), power analysis yielded a total sample size of $N=85$, which we deemed to be the minimum sample size.

Actually, a total of $N=102$ primary-school teachers participated in this study. Of these participants, 87.2% were female and 12.8% were male. The distribution of gender in our sample was in accordance with the distribution of teacher gender in German primary schools (Neugebauer and Gerth 2013). The participants' mean age was 35.40 years ($SD=9.26$), with an average of 9.36 years ($SD=5.73$) teaching experience, and 95.1% of them were German. The participants were recruited across the country, with the majority (64.8%) living in Berlin, 15.7% living in North-Rhine Westphalia, and the remaining living in Hamburg (5.9%), Lower Saxony (5.9%), Saxony (3.9), Thuringia (1.9), and Brandenburg (1.9%). All participants volunteered without any reward.

3.2 Materials

Each participant received 16 male student vignettes. We used only male student vignettes because we deemed presenting a total of more than 16 vignettes a possible burden for the participants, which might increase the likelihood of dropping out from the study. Similar vignettes have been used in previous studies (Klapproth et al. 2018; Klapproth and Fischer 2019). Each vignette contained six grades varying between 1 ("very good") and 4 ("sufficient"), with each grade being related to one school subject. However, the school subjects were not specified (e.g. subject A: "2," subject B: "2," subject C: "4," subject D: "1," subject E: "2," subject F: "3"), so that the participants had to rely entirely on the value of the grades and not be able to apply subjective weighing of the school subjects. To reach a single GPA (e.g. 2.17), the combination of grades (e.g. 1; 2; 2; 2; 2; 4) was always the same. The realized grand means of the grades of the vignettes were $M=\{2.33; 2.50; 2.67; 2.83\}$, with higher means representing lower achievements. These means were used because in some German federal states (e.g. Berlin) the interval between 2.30 and 2.70 necessitates a well-founded judgment by the teacher as to which school track a student has to be recommended for. Grand means of grades lower than 2.30 will result automatically to a highest-track recommendation, whereas grand means of grades higher than 2.70 will lead to lower-track recommendations (e.g. Senatsverwaltung für Bildung, Jugend und Familie 2019).

The grand means emerged from two school reports from the last 2 years of a 6-year German primary school. These school reports showed either improvement or decline of grades. In case of improvement, the GPAs of the first school report (representing grades obtained in school year 5, second semester) were 2.50, 2.67, 2.83, or 3.00 and the corresponding GPAs of the second school report (representing grades obtained in school year 6, first semester) were 2.17, 2.33, 2.50, or 2.67, respectively, so that the magnitude of improvement was always the same between two school reports. Accordingly, in case of decline, the GPAs of the first school report were smaller than those of the second school report. Note that the change of GPAs was always due to the change of grades in one school subject by an amount of 2.0. For instance, when a student improved in his GPA (e.g. from 2.50 to 2.17, yielding a grand mean of 2.33), he realized this improvement by the change of grades in a single school subject from 4 (“sufficient”) to 2 (“good”). The grand means of the GPAs were unrelated to both the students’ ethnicity and whether there was improvement or decline of grades.

The ethnicity of the students was manipulated by the names that were assigned to the students, which were common names of either Turkish or German male students. The names were intended to elicit a social stereotype of either ethnicity, which in turn should affect the participants’ attitudes toward the students described in the vignettes. Very common names were chosen to prevent the activation of other concepts like a certain socioeconomic background by rarely used names that are especially common in certain social and economic milieus (Gerhards 2010). Each participant was given eight vignettes with German and eight vignettes with Turkish names.

Each vignette was supplemented with information regarding the students’ working habits and social behavior in order to make the vignettes more similar to real-life school reports. This information was delivered by two rather short sentences, which were derived from standardized sentences used for appraisal of the working habits and social behavior in schools (Niedersächsisches Kultusministerium 2010). All sentences used in the vignettes displayed behavior that is regarded in school as “meeting the expectations.” Thus, all vignettes showed student behavior that was evaluated in nearly the same way. An example of the vignettes is presented in the “Appendix”.

The three factors (the students’ grand mean of GPAs, their ethnicity, and their achievement development) were varied orthogonally, resulting in a 2 (ethnicity: Turkish or German) \times 2 (development: improvement or decline) \times 4 (grand mean = 2.33, 2.50, 2.67, or 2.83) within-subjects factorial design. The dependent variable was the decision of the participants, which was either in favor of or against placement in the highest school track. We additionally collected data about the participants’ age, gender, and nationality.

3.3 Procedure

The experiment was conducted online on www.soscisurvey.de. Participants performed the task on a computer or any other device that was connected to the Internet.

The study was open for 14 days. The participants were instructed to imagine that they were a teacher of a class in the last grade of primary school and asked to make a decision on the future secondary school track for each student in the class. They were given the options “in favor of the highest track” or “not in favor of the highest track” at the end of each student description. The participants were instructed to use all the information that was presented to them for each of the 16 students. After the general instruction, an example task followed which helped the participants get acquainted with the procedure. After that, the student vignettes followed in random order. In case a participant did not make a judgment (and instead clicked on the “next” button), a prompt popped up which reminded the participant to make a decision. A new vignette was shown on the screen after the preceding vignette was closed by the decision of the participant. There was no time restriction for the participants to read the vignettes. Once the decisions were made for all 16 students, the participants were requested to give some information about their sociodemographic background. Finally, they were all thanked, debriefed about the purpose of the study, and asked for comments and requests.

3.4 Data analyses

We used multilevel logistic regression analysis to test our hypotheses. In this analysis, the judgments of the participants are nested within the participants. Hence, the level-1 unit of the analysis consists of the repeated measures for each participant, and the level-2 unit is the participant. The predictors in the regression model were the grand mean of grades (with the values 2.33, 2.50, 2.67, and 2.83) as a metric covariate, as well as the ethnicity of the students and the achievement development as nominal variables. The ethnicity was coded as 0 (German) and 1 (Turkish), and achievement development was coded as 0 (decline) and 1 (improvement). Prior to analyses, we *z*-standardized the metric predictor variable (grand mean of grades) for two reasons. First, the magnitude of its effects was comparable to the magnitude of the effects of the qualitative predictors; second, since through *z*-standardization the value of zero became meaningful, interpretations of interactions were made easier (actually, the value of zero of the *z*-standardized grand mean of grades was equal to the mean of all unstandardized grand means, which was equivalent to 2.58).

We estimated two regression models. The first model contained only main effects, whereas in the second model interaction terms were included.

4 Results

The participants were quite fast with making their decisions. On average, they needed 5.23 min (SD = 1.67) for judging 16 students. Table 1 shows the mean proportions of high-track recommendations for each condition. Actually, most recommendations (80.1%) of the participants were against the highest track.

We report the results of the regression models according to the guidelines suggested by Peng et al. (2002). Table 2 displays the results of the regression analyses.

Table 1 Mean proportions of high-track recommendations

	Grand mean			
	Ethnicity			
	Turkish		German	
	Achievement development			
	Improving	Declining	Improving	Declining
2.33	0.45	0.19	0.15	0.19
2.50	0.18	0.11	0.42	0.16
2.67	0.44	0.12	0.25	0.12
2.83	0.11	0.09	0.15	0.07

The resulting logistic regression equation of Model 1, containing only main effects, was as follows:

- (1) Predicted logit of high-track recommendation = $-2.02 + 0.15 * \text{Ethnicity} + 0.92 * \text{Development} - 0.29 * z(\text{Grand Mean})$.

Except for ethnicity, all main effects were found to be significant. Students showing improvement in achievement were 2.52 times more likely to receive a high-track recommendation than students showing deterioration, $\chi^2 = 28.19$, $p < .001$. Furthermore, students' grand mean of grades was significantly related to school-placement decisions, with higher grand means corresponding to a lower probability for a high-track recommendation, $\chi^2 = 25.12$, $p < .001$. When the z -standardized Grand mean of grades increased by one unit (which corresponded to the standard deviation of the grand mean, $SD = 0.19$), the chance for receiving a high-track recommendation dropped by 0.68.

We next applied a regression model containing both main effects and interaction effects. The resulting regression equation reads as follows:

- (2) Predicted logit of high-track recommendation = $-1.94 - 0.04 * \text{Ethnicity} + 0.80 * \text{Development} - 0.39 * z(\text{Grand Mean}) + 0.27 * \text{Ethnicity} \times \text{Development} + 0.10 * \text{Ethnicity} \times z(\text{Grand Mean}) + 0.29 * \text{Development} \times z(\text{Grand Mean}) - 0.42 * \text{Ethnicity} \times \text{Development} \times z(\text{Grand Mean})$.

Compared to model 1, the goodness of fit of Model 2, indicated by the quasi-likelihood under the independence model criterion (QIC), was slightly smaller and, hence, indicated better fit to the data. As in model 1, the main effect due to the achievement development, $\chi^2 = 16.15$, $p < .001$, and the main effect due to the grand mean, $\chi^2 = 7.45$, $p = .006$, were significant. Model 2 also revealed a significant Development \times Grand mean interaction effect, $\chi^2 = 5.27$, $p = .022$, and a significant three-way interaction effect, $\chi^2 = 7.62$, $p = .006$.

When interaction terms are included in a logistic regression equation, the coefficients for the main effects no longer represent main effects in the traditional sense. Instead, their exponents represent an odds ratio comparing the odds for a

Table 2 Results of multilevel logistic regression analysis

Predictor	Model 1				Model 2			
	B (SE)	Exp (B)	Wald χ^2 (df=1)	p	B (SE)	Exp (B)	Wald χ^2 (df=1)	p
Intercept	-2.02 (0.24)	0.13	73.36	<.001	-1.94 (0.20)	0.24	63.68	<.001
Ethnicity	0.15 (0.13)	1.16	1.20	.274	-0.04 (0.22)	0.96	0.04	.844
Development	0.92 (0.17)	2.52	28.19	<.001	0.80 (0.20)	2.22	16.15	<.001
z(grand mean)	-0.29 (0.06)	0.75	25.12	<.001	-0.39 (0.14)	0.68	7.45	.006
Ethnicity \times development					0.27 (0.19)	1.31	1.96	.161
Ethnicity \times z(grand mean)					0.10 (0.14)	1.10	0.44	.508
Development \times z(grand mean)					0.29 (0.13)	1.34	5.27	.022
Ethnicity \times development \times z(grand mean)					-0.42 (0.15)	0.66	7.62	.006
QIC			1573.84				1571.35	

The reference group for positive change and negative change was zero change

QIC means quasi-likelihood under the independence criterion. Change = change of grade point average

Coding of Ethnicity was as follows: 1 = Turkish, 0 = German

specific predictor scored 1 (on a nominal variable) with the odds for the reference group (scored 0) when all other predictors in interaction terms with this specific predictor are set to zero. For instance, the main effect of achievement development in Equation 2 means that improving students had a 2.22 times higher chance of being recommended for the highest track than students who deteriorated, when students were German (Ethnicity=0) and the z -standardized grand mean of grades was zero (corresponding to an unstandardized grand mean of 2.58). Figure 1 shows the predicted logits for all conditions realized in the experiment, with the grand mean of grades being regressed to achievement development. In the upper panel, predicted logits are shown for students accompanied by a name typical for German males (in the following “German students”), in the lower panel logits are depicted for students accompanied by a name typical for Turkish males (in the following “Turkish students”).

Figure 1 might help understand the interaction effects obtained. The slopes of the regression lines depicted in Fig. 1 represent the effect of students’

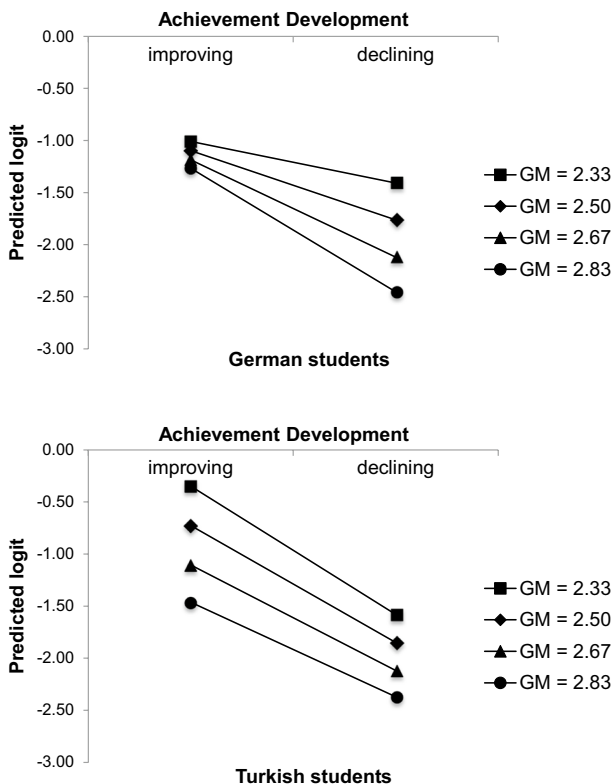


Fig. 1 Predicted logits of the probability of high-track recommendations, obtained from the different conditions in the experiment, depending on the grand mean of all grades and whether the achievement development was improving or declining. Upper panel: predicted logits for German students. Lower panel: predicted logits for Turkish students *Note*: GM means grand mean of grades

achievement development on teachers' placement recommendations. The steeper the slopes are, the more did teachers consider achievement development for their decisions. The Development \times Grand mean interaction means that the slopes of the regression lines, and hence the effect of achievement development on placement recommendations, differed between different grand means of grades. For example, for German students the effect of achievement development was apparently smaller for students showing relatively good achievement than for rather low-achieving students.

The three-way interaction effect means that the Development \times Grand mean interaction significantly differed between German and Turkish students. Whereas with German students the effect of achievement development on placement recommendations decreased with increasing achievement, the reverse was the case with Turkish students. Teachers considered the achievement development of Turkish low-performers to a lesser degree than they did with rather high-performers.

To reveal more meaning from the main and interaction effects, we conducted simple slope tests by estimating whether the slopes of all simple regression equations were significantly different from zero (cf. Cohen et al. 2003). A simple regression equation is the equation for each of the simple regression lines depicted in Fig. 1. Table 3 shows the results of the simple slope tests.

Simple slope tests revealed that for all but one regression line the slopes were significantly different from zero, which means that achievement development significantly affected the probability of high-track recommendations in seven of eight conditions.

We additionally conducted pairwise slope difference tests to examine whether the slopes of the regression lines were significantly different from one another (cf. Robinson et al. 2013). From a total of 28 comparisons of slopes, only two differences were significant. These were the differences between German-GM=2.33 and German-GM=2.83 students ($t = -2.62$, $p = .009$), and between Turkish-GM=2.33 and German-GM=2.33 students ($t = 2.78$, $p = .006$).

Since only 5 (of 102) participants had an immigration background (three were Russian, one was Turkish, and one was Austrian), we abstained from integrating the participants' ethnicity as a predictor variable into the regression models. However, for the sake of a coarse impression, we correlated the participants' ethnicity with each participant's mean frequency of recommendations for the highest track, separated for students with a Turkish or German name. There was no significant

Table 3 Results of the simple slope tests

Grand mean	Ethnicity					
	Turkish			German		
	<i>B</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>t</i>	<i>p</i>
2.33	1.23	5.01	<.001	0.40	1.84	.067
2.50	1.12	5.36	<.001	0.67	3.55	<.001
2.67	1.01	4.44	<.001	0.93	4.15	<.001
2.83	0.91	3.16	.002	1.19	3.96	<.001

correlation between the participants' ethnicity and the frequency of high-track recommendation for students with Turkish ($r=.06$, $p=.524$) or German names ($r=-.01$, $p=.903$).

5 Discussion

5.1 Discussion of the results obtained

With the present study, we examined whether primary-school teachers regarded both the ethnicity of primary-school students and their development of GPAs, indicated by two successive school reports, when making recommendations for a students' track in secondary school. This experimental study yielded several important results.

The participants mostly did not recommend students for the highest track. Less than 1/5 of the participants' recommendations (19.9%) were in favor of the highest track. Obviously, the participants judged most of the students' achievements not suitable for being taught on the highest track. Usually, students are recommended for the highest track if they show a grand mean of grades less than 2.3, whereas if their mean of grades is between 2.3 and 2.7, teachers may opt for the highest track if they find additional supportive information about the students, like for example, productive working behavior or high achievement motivation (Riek and van Ophuysen 2014). In our study only the development of achievement might have served as supportive information. However, since in most conditions students did not increase their achievements, the participants' judgments seem to be rational on the basis of official recommendation guidelines.

As hypothesized, students were more likely to be recommended for the highest track when their grand mean of grades was rather small (indicating higher achievements). This result clearly shows that the teachers participating in our study acknowledged the overall achievement indicated by the grades of two school reports as a basis for their decision.

Moreover, we assumed that teachers would consider the achievement development of the students when making school-track recommendations. In line with this hypothesis, students who improved were about 2.5 times more likely to receive a high-track recommendation than students whose grades deteriorated. These results confirm previous results (e.g. Klapproth and Fischer 2019) and are also in line with the sustaining-expectations hypothesis (Cooper 1985; Good and Brophy 2003), whereby teachers assume that improvement of students' achievements would be followed by further improvement and impairment would be followed by further impairment. Note that this effect occurred due to a rather moderate change in GPA. The difference between the successive school reports was an increase or a decrease of one single grade (out of 6) by the amount of 2 units on the German grade scale. That is, the GPAs between both school reports differed by 1/3 unit.

In contrast to our hypothesis, there was no significant main effect of student ethnicity on teachers' track recommendations. That is, on average, the

participants recommended students for the highest track with equal probability for Turkish and for German students. However, as is indicated by the significant three-way interaction, teachers made a difference between both ethnicities when concurrently considering their school achievement, represented by their school grades, and their development of achievement. The data show that for German students, teachers were affected by students' achievement development to a larger degree when students were rather low achievers—that is, when they obtained a rather large grand mean of grades—whereas when they performed rather well—meaning that they obtained lower grand means of grades—their development was of lower importance for the teachers' recommendations. This effect was reversed with Turkish students. Teachers were less affected by Turkish students' achievement development when they were rather low achievers.

The results confirm our hypothesis that students' achievement development would play a different role for teachers' recommendations depending on whether the ethnic stereotype about a student is confirmed or refuted by the students' GPA. We supposed that the stereotype about a Turkish student is likely to be activated when the school reports that are accompanied by a Turkish name represent rather low achievements, whereas in case of rather high achievements, activation of the Turkish student stereotype should be inhibited (cf. Casper et al. 2010). Conversely, with German students activation of the German student stereotype is likely when school reports represent rather high achievements and are accompanied by a name that is often used for German children. The results also confirm results previously obtained by Klapproth et al. (2018) who could show that students not fitting an ethnic stereotype (e.g. Turkish students identifying themselves with Christianity or German students identifying themselves with Islam) were judged more thoroughly on the basis of actual achievements than students fitting the stereotype (Turkish-Muslim students or German-Christian students).

Once the stereotype is activated and confirmed by salient achievement information, teachers should—according to the continuum model of impression formation (Fiske et al. 2018; Fiske and Neuberg 1990)—be likely to ignore further information about the student, like, for example, the students' achievement development. Decision making will then be category-based (Fiske and Neuberg 1990), and even if some attributes of the to-be judged individual are stereotype-inconsistent, category-based judgments are likely to prevail (Fiske 1998).

The three-way interaction effect obtained in this study also points out how the importance of grades differed for teachers when they judged students of different ethnicities. With German students, the grand mean of grades played a minor role for improving and a major role for declining students. That is, students who improved were judged more or less despite their actual grades documented in their school reports, whereas students who deteriorated were judged more carefully by considering their actual grades more thoroughly. The reverse was the case with Turkish students. Teachers considered the grand mean of grades to a lesser degree when Turkish students declined than when they improved. These results imply that stereotypes about German and Turkish students led the participants to rely their judgments more on assumed rather than on actual abilities of these students, if descriptions of these students fitted an ethnic stereotype.

However, data of this study also suggest that stereotypes about German students affected teachers' judgments more than stereotypes about Turkish students. The effect of achievement development on placement recommendations was on average smaller for German than for Turkish students, which means that teachers valued achievement development as an additional information rather for Turkish than for German students. There is evidence from previous research showing that ethnic majority or native students elicit bias in teachers' judgments (e.g. Ready and Chu 2015; Tobisch and Dresel 2017). Moreover, Tobisch and Dresel (2017) found that teachers' bias in judging future achievement and achievement aspiration of both Turkish and German students was higher for German than for Turkish students. The authors concluded that teachers probably were aware of stereotypes and prejudices about Turkish students and therefore were motivated to control for stereotypic judgments. However, the teachers may not have been aware of stereotypes about German students and their potential effect on their judgments.

5.2 Limitations

Four limitations pertinent to this study should be mentioned. First, this study was experimental in nature. While we therefore could expect the realization of a high level of internal validity, field investigations are nevertheless needed in order to replicate the effects obtained in a natural environment. Second, grades were not associated with specific school subjects. This might have been confusing for some participants, as in practice grades are always related to school subjects. However, to make the design of the study feasible and the results comparable to previous studies, we abstained from specifying school subjects. Moreover, teachers reported after the experiment that they were able to form an impression about the students despite being informed about the different school subjects. Third, all vignettes we used described male students. We therefore were unable to evaluate whether the results were affected by students' gender. Fourth, we did not ask the participants whether or not they perceived the students as either German or Turkish students. It may be possible that participants perceived some "Turkish" students as students with an Arabic immigration background because some of the names we used (e.g. Mustafa) are also common names in the Arabic countries. However, all names that we applied were frequently used as male names in either Germany or Turkey.

Follow-up studies could shed more light into the complex decision-making processes when teachers evaluate students regarding their appropriateness for a secondary school track. In order to validate the experimental studies, case studies could be applied in which participants would be presented with more elaborative student descriptions, and participants' responses could be coded qualitatively. Moreover, further studies should also include evaluations of both male and female students. Finally, the provision of instructions that explicitly state how to handle information from both school reports could reduce the bias in placement recommendations.

5.3 Conclusion and practical implications

In addition to previous studies (Caro et al. 2009; Klapproth and Fischer 2019), this study has confirmed that students' academic achievement development at the end of primary school affects teacher judgments concerning the eligibility of these students for the highest track in secondary school. Moreover, we found strong evidence that teachers apply ethnic stereotypes when making school-placement recommendations. If the students fitted an ethnic stereotype, teachers judged them on the basis of assumed abilities according to the respective ethnic stereotype. However, if the students did not fit an ethnic stereotype, teacher judgments were rather based on all information that was provided about the students, that is, information about their achievement development. Hence, achievement development was less important for school-placement recommendations when students were stereotyped than when they were not stereotyped. These results let us arrive at two conclusions. First, teachers should be made aware of their proneness to favor students who show a recent improvement over students showing a recent decrement of achievements because recent changes may not be predictive for students' future achievements in school (Klapproth and Fischer 2019). This might also be important because teachers in some German federal states (e.g. Berlin) are legally not allowed to deliberately consider the students' academic development for placement recommendations, but instead have to stick to status information provided by the school reports (Senatsverwaltung für Bildung, Jugend und Familie 2019). Second, teachers should also know about their tendency to categorize students according to their salient attributes, which—in this study—led to partial disregarding of information on achievement development for students fitting a category. Making teachers aware of judgment biases could be done within teacher training seminars (Pit-ten Cate et al. 2014). In addition, we think that it might be reasonable to start a debate on how information about students' achievement development could be formally integrated into teachers' placement decisions. One problem to be solved is that about the validity of the data from which achievement development is inferred. Usually, two times of measurement as used in the present investigation do not suffice to picture a valid trajectory of achievement, as measurement error might be high and the change of grades may only be coincidental. A possible solution to this problem might lie in using more times of measurement, as is done, for instance, in several types of curriculum-based measurement (e.g. Fuchs 2016).

Acknowledgements This study conformed to the Declaration of Helsinki and has been evaluated and approved by the Ethical Committee of the Medical School Berlin (MSB-2020/43).

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflicts of interest to disclosure.

Informed consent All participants provided informed consent about their participation.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Name of student: Mustafa

School report semester 5/2

Subject A	2	Subject D	2
Subject B	4	Subject E	2
Subject C	1	Subject F	1

School report semester 6/1

Subject A	2	Subject D	2
Subject B	4	Subject E	2
Subject C	3	Subject F	1

Notes on social behavior and working habits:

Mustafa's working habits meet the expectations of his teachers. He frequently takes part at common social activities.

Would you recommend Mustafa for the highest track (Gymnasium) in secondary school?

YES	NO
-----	----

References

- Arnold, K.-H., Bos, W., Richert, P., & Stubbe, T. C. (2007). Schullaufbahn-präferenzen am Ende der vierten Klassenstufe. [Preferences for secondary-school tracks at the end of primary school]. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, & R. Valtin (Eds.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 271–297). Münster: Waxmann.
- Baumert, J., & Schümer, G. (2002). Familiäre Lebensverhältnisse, Bildungs-beteiligung und Kompetenzerwerb im nationalen Vergleich. [Family affairs, educational participation, and skill acquisition in national comparison]. In J. Baumert, et al. (Eds.), *PISA 2000—Die Länder der Bundesrepublik Deutschland im Vergleich* (pp. 159–202). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Baur, R. S., & Ossenberg, S. (2016). *Informationen zum Forschungsprojekt "Deutsche und türkische Stereotype im Vergleich"*. [Information about the research project "Comparison of German and Turkish stereotypes"]. Duisburg: Universität Duisburg-Essen.
- Becker, M., Neumann, M., & Dumont, H. (2016). Recent developments in school tracking practices in Germany: An overview and outlook on future trends. *Orbis Scholae*, 10, 9–25. <https://doi.org/10.14712/23363177.2017.8>.
- Bellenberg, G., & Forell, M. (2012). *Schulformwechsel in Deutschland. Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I*. [School track changes in Germany. Permeability and selection in school systems of the 16 federal states in compulsory secondary school]. Gütersloh: Bertelsmann Stiftung.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex: Wiley.
- Caro, D. H., Lenkeit, J., Lehmann, R., & Schwippert, K. (2009). The role of academic achievement growth in school track recommendation. *Studies in Educational Evaluation*, 35, 183–192. <https://doi.org/10.1016/j.stueduc.2009.12.002>.
- Casper, C., Rothermund, K., & Wentura, D. (2010). Automatic stereotype activation is context dependent. *Social Psychology*, 41, 131–136. <https://doi.org/10.1027/1864-9335/a000019>.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.
- Cooper, H. (1985). Models of teacher expectation communication. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 135–158). Hillsdale, NJ: Erlbaum.
- Cooper, H., Lowe, C., & Baron, R. (1976). Pattern of past performance and expected future performance: A reversal of the unexpected primacy effect. *Journal of Applied Social Psychology*, 6, 31–39. <https://doi.org/10.1111/j.1559-1816.1976.tb01309.x>.
- Darmody, M., Delma, B., & McGinnity, F. (2014). Cumulative disadvantage? Educational careers of migrant students in Irish secondary schools. *Race, Ethnicity and Education*, 17, 129–151. <https://doi.org/10.1080/13613324.2012.674021>.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95, 158–165. <https://doi.org/10.1257/000282805774670446>.
- Ditton, H., Krüskens, J., & Schauenberg, M. (2005). Bildungsungleichheit—der Beitrag von Familie und Schule. [Inequality in education—The contribution of family and school]. *Zeitschrift für Erziehungswissenschaft*, 8, 285–304. <https://doi.org/10.1007/s11618-005-0138-x>.
- Dollmann, J. (2010). *Türkischstämmige Kinder am ersten Bildungswegübergang: Primäre und sekundäre Herkunftseffekte*. [Turkish children at the first transition in education: primary and secondary effects of origin]. Berlin: Springer.
- Dumont, H., Maaz, K., Neumann, M., & Becker, M. (2014). Soziale Ungleichheiten beim Übergang von der Grundschule in die Sekundarstufe I: Theorie, Forschungsstand, Interventions- und Fördermöglichkeiten. [Social inequalities at the transition from primary to secondary school: Theory, research, and possibility of intervention]. *Zeitschrift für Erziehungswissenschaft*, 17, 141–165. <https://doi.org/10.1007/s11618-013-0466-1>.
- Erensi, S., & Adanlı, Y. A. (2004). Turkey in the eye of the beholder: Tracking perceptions on Turkey through political cartoons. *Kontur*, 10, 58–72.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 357–411). New York, NY: Oxford University Press.
- Fiske, S. T., Lin, M., & Neuberg, S. L. (2018). The continuum model: Ten years later. In S. T. Fiske (Ed.), *Social cognition* (pp. 41–75). London: Routledge.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, 23, 1–74. [https://doi.org/10.1016/S0065-2601\(08\)60317-2](https://doi.org/10.1016/S0065-2601(08)60317-2).
- Fuchs, L. S. (2016). Curriculum-based measurement as the emerging alternative: Three decades later. *Learning Disabilities Research & Practice*, 32, 5–7. <https://doi.org/10.1111/ldrp.12127>.

- Gerhards, J. (2010). *Die Moderne und ihre Vornamen: Eine Einladung in die Kulturosoziologie. [Modernity and its forenames: An invitation to sociology of culture]*. Wiesbaden: VS Verlag. <https://doi.org/10.1007/978-3-531-92353-6>.
- Glock, S. (2016). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers' judgments of students. *Social Psychology of Education, 16*, 493–509. <https://doi.org/10.1007/s11218-016-9349-7>.
- Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education, 16*, 111–127. <https://doi.org/10.1007/s11218-012-9197-z>.
- Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013a). Prädiktoren der Schullaufbahnenempfehlung für die Schulzweige des Sekundarbereichs I: Ein Überblick. [Predictors of school-placement recommendations: An overview]. *Pädagogische Rundschau, 67*, 349–367.
- Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013b). Beyond judgment bias: How students' ethnicity and academic profile consistency influence teachers' tracking judgments. *Social Psychology of Education, 16*, 555–573. <https://doi.org/10.1007/s11218-013-9227-5>.
- Good, T. L., & Brophy, J. E. (2003). *Looking in classrooms* (9th ed.). Boston, MA: Allyn and Bacon.
- Gresch, C. (2012). Migrantenkinder auf dem Weg zum Abitur: Wie kommen die Übergangsempfehlungen nach der Grundschule zustande? [Immigrant children on their way towards university-entrance diploma: What determines school-placement recommendations?]. *WZBrief Bildung, 21*, 1–7.
- 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*, 131–155. https://doi.org/10.1207/s15327957pspr0902_3.
- Keller, G. (1991). Stereotypes in intercultural communication: Effects of German–British pupil exchanges. In D. Buttjes & M. Byram (Eds.), *Mediating languages and cultures: Towards an intercultural theory of foreign language education* (pp. 120–136). Clevedon: Bank House.
- Klapproth, F., & Fischer, B. D. (2019). Preservice teachers' evaluations of students' achievement development in the context of school-track recommendations. *European Journal of Psychology of Education, 34*, 825–846. <https://doi.org/10.1007/s10212-018-0405-x>.
- Klapproth, F., Glock, S., Krolak-Schwerdt, S., Martin, R., & Böhmer, M. (2013). Prädiktoren der Sekundarschulempfehlung in Luxemburg. Ergebnisse einer large-scale-Untersuchung. [Predictors of school-placement recommendations in Luxembourg: Results of a large-scale analysis]. *Zeitschrift für Erziehungswissenschaft, 16*, 355–379. <https://doi.org/10.1007/s11618-013-0340-1>.
- Klapproth, F., Kärchner, H., & Glock, S. (2018). Do students' religion and school absences moderate the effect of ethnic stereotypes on school-placement recommendations? *The Journal of Experimental Education, 86*, 173–194. <https://doi.org/10.1080/00220973.2017.1293602>.
- Kleen, H., & Glock, S. (2018). A further look into ethnicity: The impact of stereotypical expectations on teachers' judgments of female ethnic minority students. *Social Psychology of Education, 21*, 759–773. <https://doi.org/10.1007/s11218-018-9451-0>.
- KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2010). *Übergang von der Grundschule in Schulen des Sekundarbereichs I. [The transition from primary to secondary school]*. Retrieved December 12, 2019 from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_10_18-Uebergang-Grundschule-S_eI1-Orientierungsstufe.pdf.
- Kristen, C. (2006). Ethnische Diskriminierung in der Grundschule? Die Vergabe von Noten und Bildungsempfehlungen. [Ethnic discrimination in primary school? Teachers' decisions on grades and track recommendations]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 58*, 79–97. <https://doi.org/10.1007/s11575-006-0004-y>.
- Lintorf, K., Guill, K., & Bos, W. (2008). Grundschulübergang in Nordrhein-Westfalen: Einflussfaktoren auf das Lehrerurteil im Prognoseunterricht. [The transition from primary school in North Rhine-Westphalia: predictors of teachers' judgments]. In E.-M. Lankes (Ed.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (pp. 99–110). Münster: Waxmann.
- Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*, 397–407. <https://doi.org/10.1037/0022-3514.69.3.397>.
- Marx, A. E., & Stanat, P. (2012). Reading comprehension of immigrant students in Germany: Research evidence on determinants and target points for intervention. *Reading and Writing, 25*, 1929–1945. <https://doi.org/10.1007/s11145-011-9307-x>.

- Neugebauer, M., & Gerth, M. (2013). Weiblicher Schulkontext und Schulerfolg von Jungen. [Female school context and male school success]. In R. Becker & A. Schulze (Eds.), *Bildungskontexte* (pp. 431–455). Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-18985-7_15.
- Niedersächsisches Kultusministerium. (2010). *Bewertung des Arbeits- und Sozialverhaltens. Runderlass des Kultusministeriums. [Judgment of working behavior and social competence: Edict of the Ministry of Education]*. Hannover: Niedersächsisches Kultusministerium.
- Nikolai, R. (2019). After German reunification: The implementation of a two-tier school model in Berlin and Saxony. *History of Education*, 48, 374–394. <https://doi.org/10.1080/0046760X.2018.1545932>.
- Pásztor, A. (2010). “Go, go on and go higher an’ higher”. Second-generation Turks’ understanding of the role of education and their struggle through the Dutch school system. *British Journal of Sociology of Education*, 31, 59–70. <https://doi.org/10.1080/01425690903385451>.
- Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, 96, 3–14. <https://doi.org/10.1080/00220670209598786>.
- Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers’ explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, 42, 123–140. <https://doi.org/10.1016/j.learninstruc.2016.01.010>.
- Pit-ten Cate, I., Krolak-Schwerdt, S., Glock, S., & Markova, M. (2014). Improving teachers’ judgments: Obtaining change through cognitive processes. In S. Krolak-Schwerdt, S. Glock, & M. Böhmer (Eds.), *Teachers’ professional development. Assessment, training, and learning* (pp. 45–61). Rotterdam: Sense Publishers. <https://doi.org/10.1007/978-94-6209-536-6>.
- Ready, D. D., & Chu, E. M. (2015). Sociodemographic inequality in early literacy development: The role of teacher perceptual accuracy. *Early Education and Development*, 26, 970–987. <https://doi.org/10.1080/10409289.2015.1004516>.
- Riek, K., & van Ophuysen, S. (2014). Kriterien der Übergangsempfehlung—Eine qualitative Interviewstudie mit Grundschullehrkräften. [Criteria of school-placement recommendations—A qualitative interview study with primary-school teachers]. In B. Kopp, S. Martschinke, S. Munser-Kiefer, M. Haider, E. Kirschhock, G. Ranger, & G. Renner (Eds.), *Individuelle Förderung und Lernen in der Gemeinschaft. Jahrbuch Grundschulforschung, Bd* (Vol. 17, pp. 270–273). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-04479-4_51.
- Riley, T., & Ungerleider, C. (2008). Preservice teachers’ discriminatory judgments. *Alberta Journal of Educational Research*, 54, 378–387.
- Robinson, C. D., Tomek, S., & Schumacker, R. E. (2013). Tests of moderations effects: Difference in simple slopes versus the interaction term. *Multiple Linear Regression Viewpoints*, 39, 16–24.
- Rolison, M. A., & Medway, F. J. (1985). Teachers’ expectations and attributions for student achievement: Effects of label, performance pattern, and special education intervention. *American Educational Research Journal*, 22, 561–573. <https://doi.org/10.3102/00028312022004561>.
- Senatsverwaltung für Bildung, Jugend und Familie. (2019). *Wohin nach der Grundschule? [Where to go after primary school?] Berlin*. Berlin: SenBJF.
- Sneyders, E., Vanhoof, J., & Mahieu, P. (2018). Primary teachers’ perceptions that impact upon track recommendations regarding pupils’ enrolment in secondary education: A path analysis. *Social Psychology of Education*, 21, 1153–1173. <https://doi.org/10.1007/s11218-018-9458-6>.
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. [Students with migration background]. In E. Klieme, et al. (Eds.), *Pisa 2000. Bilanz nach einem Jahrzehnt* (pp. 200–230). Münster: Waxmann.
- Stubbe, T. C. (2009). *Bildungsentscheidungen und sekundäre Herkunftseffekte: Soziale Disparitäten bei Hamburger Schülerinnen und Schülern der Sekundarstufe I. [Decisions on education and secondary effects of origin: Social disparities in students of Hamburg secondary school]*. Münster: Waxmann.
- Stubbe, T. C., & Bos, W. (2008). Schullaufbahneempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. [Teachers’ and parents’ school-placement decisions at the end of primary school]. *Empirische Pädagogik*, 22, 49–63.
- 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, 253–273. <https://doi.org/10.1037/0022-0663.99.2.253>.
- Tiedemann, J., & Billmann-Mahecha, E. (2007). Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I. [On the effect of migration and

- school-class affiliation on school-placement recommendations]. *Zeitschrift für Erziehungswissenschaft*, 10, 108–120. <https://doi.org/10.1007/s11618-007-0009-8>.
- 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, 731–752. <https://doi.org/10.1007/s11218-017-9392-z>.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101–126. <https://doi.org/10.1037//0033-295X.107.1.101>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Florian Klapproth Ph.D., is Professor of Educational Psychology at Medical School Berlin, Berlin, Germany. His research areas of interests include social stereotypes and stereotyping, teacher education, and perception of time and change.

Birthe Doreen Fischer M.Sc. in Psychology, is currently working as a clinical psychologist at Verbundkrankenhaus Bernkastel/Wittlich, Germany. Her main research interest is on Educational Psychology and learning.