



When ethnic minority students are judged as more suitable for the highest school track: a shifting standards experiment

Anna K. Nishen¹ · Katja Corcoran² · Katharina Holder¹ · Ursula Kessels¹

Received: 27 July 2021 / Revised: 29 October 2021 / Accepted: 16 December 2021 /
Published online: 29 January 2022
© The Author(s) 2022

Abstract

When students are grouped into school tracks, this has lasting consequences for their learning and later careers. In Germany to date, some groups of students (boys, ethnic minority students) are underrepresented in the highest track. Stereotypes about these groups exist that entail negative expectations about their suitability for the highest track. Based on the shifting standards model, the present research examines if and how stereotypes influence tracking recommendations. According to this theory, members of negatively stereotyped groups will be judged more leniently or more strictly depending on the framing of the judgment situation (by inducing minimum or confirmatory standards). $N=280$ teacher students participated in a vignette study in which they had to choose the amount of positive evidence for suitability they wanted to see before deciding to recommend a fictitious student to the highest track. A 2 (judgment standard: minimum vs. confirmatory) $\times 2$ (target student's gender: male vs. female) $\times 2$ (target student's ethnicity: no migration background vs. Turkish migration background) between-subjects design was used. No effects of target gender occurred, but the expected interaction of target's ethnicity and judgment standard emerged. In the minimum standard condition, less evidence was required for the ethnic minority student to be recommended for the highest track compared to the majority student. In the confirmatory standards condition, however, participants tended to require less evidence for the ethnic majority student. Our experiment underlines the importance of the framing of the recommendation situation, resulting in a more lenient or stricter assessment of negatively stereotyped groups.

Keywords Shifting standards · Stereotypes · Ethnic bias · Gender bias · Tracking recommendations · Teacher judgments

✉ Anna K. Nishen
anna.nishen@fu-berlin.de

¹ Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany

² Institute of Psychology, Karl-Franzens-Universität Graz, Universitätsplatz 2, 8010 Graz, Austria

Introduction

In educational settings, tracking is often used in the hopes that classrooms that are more homogenous in their performance will receive lessons that are better tailored to their needs (Hattie, 2002). It is argued that teachers are able to scaffold students better in more performance-homogenous groups, neither posing too difficult tasks nor challenging students insufficiently. However, tracking is often criticized because it reproduces existing social class differences and because the lower track provides a lower-quality learning environment, contradicting the aim of catering better to the needs of all students (Baumert et al., 2006; Hattie, 2002; Maaz et al., 2008; Retelsdorf et al., 2010; Van Houtte, 2004).

Tracking can be used both within schools, so that students attend classes at different performance levels in their school, and between schools, meaning that students of different performance levels are taught in different schools (Becker et al., 2016). In Germany, students are tracked between schools after 4th or 6th grade (ages nine or eleven) depending on the federal state, and the school track they attend has long-term consequences. Students in the highest track (academic track, *Gymnasium*) receive a diploma permitting them to study at the university, whereas students with a diploma from the vocational track, which some federal states further separate into two tracks, qualify for apprenticeships (see Becker et al. (2016) for an overview). Importantly, there is evidence that the different tracks constitute unequal learning environments. Teachers in lower tracks show less adaptive goal orientations and instructional methods, and they are more likely to experience burnout (Retelsdorf et al., 2010; Van Houtte, 2004). Moreover, students in higher tracks make greater achievement gains even when prior achievement is controlled (*scissor effect*; Becker et al., 2006, 2012; Hattie, 2002; Maaz et al., 2008; Retelsdorf et al., 2012). In theory, the tracking system is meant to be permeable, though in practice this is a mostly downward movement of students to a lower track (Bellenberg, 2012).

With effects on both learning in school and later-life outcomes, tracking decisions are clearly impactful for students' lives and should not be made lightly. The German government established that tracking recommendations should consider "not only performance related to goals established in the curriculum but also the general skills important for success in school" (Standing Conference of the Ministers of Education and Cultural Affairs [KMK], 2015; p. 6). The recommendation of a student for a specific track should be based on their performance, but also their "suitability, affinity, and willingness [...] to work intellectually" (KMK, 2015; p. 5). While in most German federal states, teachers' recommendations are *not binding* anymore, they carry weight in the admission process to different schools of the same track. Some states do have *binding* recommendations, meaning that students may need to further prove their abilities if they want to attend a higher-than-recommended track (KMK, 2015).

In addition to more objective aspects such as grades, teachers do consider motivational and social factors when making a recommendation as suggested by the guideline (Anders et al., 2010; Kaiser et al., 2013; Sneyers et al., 2018). Because teachers' perceptions play a role in their recommendations, they may be vulnerable to bias along gender and ethnicity lines. Indeed, studies show that school is perceived as a more feminine space by students and teachers alike (Heyder & Kessels, 2013); Kessels, 2015, and boys are underrepresented in the academic track (Spinath et al., 2014, Federal Statistical Office, 2019). Similarly, students with a Turkish migration background are underrepresented in the academic track (Rauch et al., 2016). German teacher students have lower expectations for students with a Turkish migration background (Lorenz et al., 2016), and rate them as less qualified for the

academic track (Tobisch & Dresel, 2017) based on the same performance. Our study aims to complement this research by examining whether the framing of a recommendation situation can influence the degree and directionality of bias against male and ethnic minority students.

We draw from the shifting standards model (SSM; Biernat, 2012), which posits that people shift their standards when assessing a person's suitability to a specific position or task according to the social group this person belongs to and the stereotypes about this group's abilities and characteristics. Depending on the framing of the judgment situation—whether the assessment is meant to be just a first, tentative impression or it is meant to be a judgment made with certainty—more or less evidence is required to judge members of negatively stereotyped groups as possessing the characteristic in question (Biernat et al., 2008, 2010). Translating the predictions of the SSM to tracking recommendations, the present study tests how many positive learning-related behavioral examples teacher students will want to see to make an academic track recommendation for negatively stereotyped students (compared to positively stereotyped students) in two different framings of the judgment situation.

Shifting standards in stereotyped judgments

Focusing on gender and racial group membership, the basic premise of the SSM is that stereotypes about various social groups are used as expectations or standards. A stereotype-based judgment standard entails expectations about “the likely mean and range of group members on the attribute” (Biernat, 1995, p. 89). Thus, a specific target is being judged relative to the average expectations for their group and perceivers tend to shift their standard of judgment to correspond to this expectation (Biernat, 1995). This is possible if the language used is subjective, that is, when it allows for different interpretations of the same word (Biernat, 2012). For instance, when describing a woman as “tall,” she is perceived as tall compared with an average woman (Biernat & Manis, 1994), while a man of the same height would not be rated as “tall” because the average man is expected to be taller than the average woman. Research has shown that members of negatively stereotyped groups are rated more negatively on objective scales, but equally or even more positively on such subjective scales (Biernat, 2012).

Shifting standards in teacher judgments

So far, research has studied standard shifts predominantly regarding adult targets. This is surprising given the vast amount of research on ethnic and gender bias in the school context in many countries (Anderson-Clark et al., 2008; Arbuckle & Little, 2004; Glock, 2016a; Heyder & Kessels, 2017; Kessels, 2015; Kessels & Steinmayr, 2013; Krahe et al., 2007; Lorenz et al., 2016; Malouff & Thorsteinsson, 2016; Meissel et al., 2017; Parks & Kennedy, 2007; Tobisch & Dresel, 2017). In Germany, two studies applied the shifting standard paradigm to negatively stereotyped elementary school students contrasting the use of objective versus subjective scales. The use of subjective scales (e.g., characterizing a student's math competencies as “good”) allows different standards of judgment to be applied to students from different social groups and should therefore mask biases. A girl described as “really good at math” might be perceived as good only compared to other girls for whom the perceiver might have low expectations. On the other hand, the use of

objective scales (e.g., estimating the score a student has in a standardized test) does not allow for the use of different standards for different groups and should therefore reveal stereotyped expectations (Biernat, 2012). In accordance with the SSM, Holder and Kessels (2017, study 1) found girls to be judged more strictly in Math on objective scales and more leniently on subjective scales compared to boys. In parallel, teacher students judged boys with a Turkish migration background more strictly in German on objective scales and more leniently on subjective scales compared to boys without migration background (Holder & Kessels, 2017, study 2). Together, these studies revealed a standard shift in teacher judgments for the first time.

Minimum standards and confirmatory standards

Standard shifts depend not only on language, but also on contextual factors. Biernat and Kobrynowicz (1997) distinguished between minimum standards and confirmatory standards. *Minimum standards* are “expectations for a group and tend to directly reflect stereotypes” (Biernat et al., 2010, p. 855) and thus tend to be *lower* for group members stereotyped as deficient on a given attribute. These standards are used when some initial evidence for a required ability is needed, for instance for a short-list in a selection process, or when a non-exclusive reward (e.g., verbal praise) is given out. In this case, the perceived group member’s behavior is compared to the low expectations the judge has for the group he or she belongs to.

On the other hand, *confirmatory standards* are “thresholds that reflect certainty that an individual has an attribute” (Biernat et al., 2010, p. 855) and thus tend to be *higher* for group members stereotyped as deficient. These confirmatory standards are used when perceivers seek “a higher level of proof; a level that allows for the broad based inference that the attribute is ability-based” (Biernat & Kobrynowicz, 1997, p. 546). This is the case when definite proofs for a required ability are needed, when a final selection or hiring takes place, or a zero-sum reward (e.g., an award) is given out. In this case, a cross-category comparison between low expectations for the negatively stereotyped group and high expectations for the other group takes place (Biernat, 2012). For example, in one study, Black men were *suspected* of being competent based on a smaller number of behavioral examples of competence (minimum standard) but were *confirmed* to be competent based on more behavioral examples (confirmatory standard) relative to White men (Biernat & Kobrynowicz, 1997). This pattern has been repeatedly found in studies on gender and competence in workplace/masculine domains as well as on race and workplace competence (Biernat & Fuegen, 2001; Biernat & Kobrynowicz, 1997; Biernat & Ma, 2005; Biernat et al., 2008).

According to the authors, a suspicion-certainty continuum of dispositional inference is at the heart of the standards shift (Biernat et al., 2008). Since expectations are lower for members of the negatively stereotyped group, little counter-stereotypical evidence is perceived as sufficient for perceivers to suspect (minimum standard) that this group member might possess the attribute in question to some degree. However, these low expectations also imply that perceivers require more behavioral evidence to be certain (confirmatory standard) that the member of a negatively stereotyped group actually has the attribute of interest (Biernat et al., 2008).

While first studies showed the impact of subjective versus objective scales when teachers judge negatively stereotyped students (Holder & Kessels, 2017), it is unknown which role minimum and confirmatory standards may play in the educational context. If negative

stereotypes towards students of Turkish origin and towards boys in general regarding their readiness to attend the academic track do exist, the evidence needed for a recommendation for this track should differ by the standard invoked. As a result, such stereotypes should influence tracking recommendations and ultimately the likelihood of students from stereotyped groups to move into the highest track.

Ethnic bias in teachers' judgments and tracking recommendations

As research has shown, stereotypes might influence teachers' expectations for and judgments of students (e.g., Jussim & Harber, 2005), and may result in tracking recommendations influenced by stereotypes (e.g., Klapproth et al., 2018). Teacher expectations for ethnic majority students are regularly found to be higher than for negatively stereotyped ethnic minority students, especially for Black students in the USA (Jussim & Harber, 2005; McKown & Weinstein, 2008; Ready & Wright, 2011; Tenenbaum & Ruck, 2007). In Germany, many students with a migration background have a Turkish background (Rauch et al., 2016), and these students not only underperform compared to students without a migration background but also compared to students from other ethnic minorities (Stanat et al., 2010). However, research also demonstrates that differences in teacher expectations or performance judgments remain even when statistically or experimentally controlling for achievement (e.g., Glock, 2016a; Holder & Kessels, 2017; Lorenz et al., 2016).

Not only teachers' expectations but also their judgments of performance and behavior have been found to be biased by students' ethnicity (Anderson-Clark et al., 2008; Malouff & Thorsteinsson, 2016; Meissel et al., 2017; Parks & Kennedy, 2007; Quinn, 2020; but see Baker et al., 2015; Glock & Krolak-Schwerdt, 2014; Kaiser et al., 2017). Ethnic minority students' performance is often judged as deficient relative to their ethnic majority peers, in Germany most visibly in studies concerning students of Turkish origin (e.g., Bonefeld & Dickhäuser, 2018; Bonefeld et al., 2017; Sprietsma, 2013). Experimental evidence focusing on language proficiency indicated that teacher students judged a student with a Turkish migration background less favorably and as less competent in German than a student without a migration background even though their performance was identical (Glock & Krolak-Schwerdt, 2013; Glock, 2016a; Holder & Kessels, 2017). Moreover, supposed primary school students with a Turkish migration background received worse grades for the same performance (Bonefeld & Dickhäuser, 2018; Sprietsma, 2013). Since grades in German are especially relevant to teachers' tracking recommendations (Ditton et al., 2005), this bias might have lasting consequences for students.

Studies on perceptions of students' behavior revealed that when students misbehave, teachers respond more harshly to ethnic minority students, particularly teacher students who have little in-class experience (Glock, 2016b; Kleen & Glock, 2018). A relatively minor misbehavior, for example, talking to a peer, can then lead to a disproportionate negative reaction from the teacher—which may negatively affect how the teacher views the child's suitability to the academic track. Moreover, data from the USA demonstrated that elementary school teachers systematically underestimate the executive functions of ethnic minority students and students with limited English proficiency (Garcia et al., 2019). Such a bias could be highly relevant to teachers' perceptions of students' ability to succeed in the academic track.

While bias in how students' performance and behavior is judged could influence teachers' tracking recommendations for ethnic minority and majority students, so far no consistent pattern has emerged regarding this. Data based on field studies in German primary schools did not find evidence that disadvantages exist for students with a migration background compared to students without a migration background when socioeconomic background and/or achievement indicators were accounted for (Ditton et al., 2005; Kristen, 2006; Schneider, 2011; Tiedemann & Billmann-Mahecha, 2007). Once these variables are included, some studies even demonstrated a slight advantage for students with a (specific) migration background for tracking recommendations (Bos et al., 2004; Driessen et al., 2008) and a large advantage for de facto transitions to secondary schools (Gresch & Becker, 2010). A study from the Netherlands also demonstrated substantial between-teacher variation in the degree and directionality of teacher biases towards ethnic minority students (Timmermans et al., 2015).

In contrast, experimental evidence suggests that teachers and teacher students do discriminate against them in their tracking recommendations when achievement is held constant (e.g., Sprietsma, 2013). A recent vignette study provided evidence that teachers rated a student with a Turkish migration background as less qualified for the highest school track than a student without a migration background (Tobisch & Dresel, 2017), a pattern that also extends to tracking recommendations (Glock et al., 2013; Klapproth et al., 2018; Sprietsma, 2013; but see also Glock et al., 2015).

In summary, several studies revealed that teachers hold lower expectations for ethnic minority students compared to ethnic majority students and that teacher judgments might be biased. Teachers also react to student misbehavior differently based on student ethnicity, indicating that they might perceive these students differently. In Germany, empirical evidence points to biased expectations and judgments especially regarding students with a Turkish migration background. With regard to tracking recommendations, the findings on the influence of students' ethnicity are inconclusive to date. The present study seeks to identify a contextual factor of the recommendation situation—whether the standard invoked is minimum or confirmatory—that may result in relative advantages of students with and without a migration background.

Gender bias in teachers' judgments and tracking recommendations

Gender identity is another factor related to school performance and attendance of the academic track. Though to a much smaller extent compared to students with a migration background, boys in general are underrepresented in the academic track (Spinath et al., 2014; Federal Statistical Office, 2019), and teachers are less likely to recommend a higher track for them (Bos et al., 2007; Driessen et al., 2008; Jürges & Schneider, 2011; Lehmann et al., 1997; Timmermans et al., 2015; but see also Ditton et al., 2005; Klapproth & Fischer, 2019). In part, this can be explained by their—on average—lower German grades since these are weighted more strongly for the recommendation than other grades (Ditton et al., 2005; Lehmann et al., 1997). Even when ability based on standardized tests is controlled for, girls receive better German grades than boys (Lehmann et al., 1997) and boys need better reading competence than girls to have the same chance of receiving a recommendation for the academic track (Bos et al., 2007).

Research indicates that teachers do, in addition to grades, consider motivational and social factors as is intended by the government's guidelines (Anders et al., 2010; Kaiser

et al., 2013; Sneyers et al., 2018). On an implicit level, teacher students associate girls more with positive student behaviors and boys more with negative student behaviors (Glock & Kleen, 2017). Teachers also explicitly report that girls more so than boys demonstrate specific additional skills and behaviors relevant to achievement (e.g., Heyder & Kessels, 2017; Kessels, 2015), and this might lead them to recommend the academic track more often for girls. Indeed, girls themselves indicate that they are more self-disciplined, more empathetic, and have more positive attitudes towards seeking help (Duckworth & Seligman, 2006; Kessels & Steinmayr, 2013; Krahe et al., 2007). Teachers also perceive that girls are more motivated, show more effort and behavior that fosters learning, study more independently, are less disruptive, and are more socially skilled (Anders et al., 2010; Arbuckle & Little, 2004; Heyder & Kessels, 2017; Jones & Myhill, 2004; Kessels, 2015; Kuhl & Hannover, 2012; Spinath et al., 2014; Veenstra et al., 2008). These behaviors may indicate to teachers that girls have greater suitability, affinity, and willingness to work intellectually—and thus they would be more likely to recommend them for the academic track (Anders et al., 2010).

Given the only small differences in reading competencies between boys and girls at the end of elementary school in Germany (McElvany et al., 2017), the main reason for boys being less frequently recommended to the highest track seems to be the difference in social and learning behavior. These differences lead to both teachers' perceptions of girls being more adjusted to the school demands as well as to girls' better grades (Hannover & Kessels, 2011). Research has highlighted that when controlling for students' performance *and* learning behavior (regarding homework, self-regulation), boys are in fact recommended for the highest track more often than girls (Neugebauer, 2011). If the stereotype about boys implies a lack of adaptive social and learning behaviors, tracking recommendations based on these stereotypes might be biased. In the present study on ethnic bias in tracking judgments, we therefore include gender as a second factor in order to analyze what kind of framing of the tracking judgment will result in relative advantages or disadvantages regarding both gender and minority status.

Study overview and hypotheses

The SSM has proven to be a useful framework for revealing stereotype-based judgments in many fields such as hiring, parenting, workplace competence, and personality traits such as assertiveness (Biernat, 2012). However, despite the rich body of research on gender and ethnic bias at schools (e.g., Arbuckle & Little, 2004; Heyder & Kessels, 2017; Kessels, 2015; Malouff & Thorsteinsson, 2016; Tobisch & Dresel, 2017), the SSM has only rarely been applied to school-based judgments. The present study examines for the first time if a standard shift occurs in tracking judgments, a judgment that is highly influential for students' future education and work life.

To test this, we vary the framing of the tracking judgment so that for one group of German teacher students, a minimum standard is induced, whereas a confirmatory standard is induced for the other. Following prior research, we applied the behavioral checklist paradigm (Biernat & Kobryniewicz, 1997; Biernat et al., 2008) enumerating behaviors that foster students' learning. The minimum standard was invoked by instructing participants to tick the minimum number of behaviors that are necessary to *suspect* that a student *may* be qualified to attend the highest track, while the confirmatory standard was invoked by asking them to indicate the number of behaviors that are necessary to *confirm* that a student

is qualified for the highest track. In addition, ethnicity and gender of target students were varied between participants.

If negative stereotypes about the suitability of boys and of students with a Turkish migration background for attending the highest track exist, an interaction effect should emerge (H1): In the minimum standards condition, teacher students are expected to require less evidence for the suitability of negatively stereotyped students, that is, students with a Turkish migration background and boys, than positively stereotyped students, that is, students without a migration background and girls (H1a). In the confirmatory standard condition, they are expected to require more evidence to confirm the suitability of students with a Turkish migration background and boys (H1b). Importantly, students with a Turkish background are much more underrepresented at the highest tracks than boys in general and ability stereotypes disadvantage students with a Turkish migration background (Froehlich et al., 2016) which is not the case for boys in general. Therefore, the pattern in accordance with the SSM regarding ethnicity may be more pronounced relative to gender. We will also examine the interplay of gender and ethnicity, though only in an exploratory manner since to date intersectional effects within the standards shift paradigm have not been studied in detail.

Method

Participants

Participants were recruited using the university mailing list for students at a large university in a city of Northwestern Germany. An email invited teacher students to participate in an online study on tracking recommendations. As an incentive, the possible participation in a lottery for vouchers worth 20 € was included. Notwithstanding that both the invitation email and the instruction of the online study stated clearly that only teacher students were the target group for the study, several students not in the teacher training program participated ($n=261$). These were excluded from the data analyses, as well as $n=14$ persons ticking “other” instead of the different levels of teacher training spelled out in the questionnaire. In addition, we decided to exclude those students who were enrolled in the program for teaching exclusively children with Special Educational Needs (SEN; $n=123$) as we aimed to test a sample with a comparable conception of regular students’ behaviors at grade 4 and the fictitious target student described in the vignette had no SEN. The final study sample comprised $n=280$ teacher students ($n=31$ primary school; $n=45$ secondary school/lower tracks; $n=204$ secondary school/higher track), the mean age was 24.71 ($SD=4.24$), and 95% spoke German as their mother tongue. Unfortunately, gender of participants was not collected.

Design, experimental treatment, and measures

Participants were randomly assigned to one of eight conditions in a 2 (standard: minimum vs. confirmatory) \times 2 (target student’s gender: male vs. female) \times 2 (target student’s ethnicity: no migration background vs. Turkish migration background) between-subjects design. On the first page of the online questionnaire, an “exercise in evaluating elementary school children” was announced. On the second page, participants read that teachers are requested to form an overall picture of a student when making tracking recommendations at the end

of primary school and that this is not exclusively based on students' grades. They further read that the guidelines of the KMK ask to consider the child's "suitability, affinity, and willingness [...] to work intellectually" (2015, p. 5) as well as the "general skills important for success at school" (2015, p. 6). As a basis for this, various school-related behaviors that students might exhibit could be considered. Participants were told that they would now engage in a short exercise for doing this. On the next page, they were asked to imagine one specific example of a primary student they had to judge. "[Target's name] is student of a fourth class in [city]. Considering exclusively his/her GPA, a transition to the Gymnasium would not be unambiguously justifiable. How do you proceed when judging [target's name] behavior? In the following you see ten behaviors that can be observed in fourth-graders."

Following procedures used by Biernat et al. (2010), participants in the minimum standards condition read that we were interested in "the minimum number of behaviors that are necessary to suspect that [the target student] may be qualified for the Gymnasium," while participants in the confirmatory standards condition read that we were interested in the "the total number of behaviors that are necessary to confirm that [the target student] is qualified for the Gymnasium." All participants were asked to review a list of ten behaviors and to check off "as many or as few behaviors" as the target student would need to engage in to either "give you some inkling or hint that [the target student] may be qualified for the Gymnasium" or "to confirm that [the target student] is qualified for the Gymnasium" (emphasis original). Crossed with this manipulation, participants were asked to imagine a student named "Tim Menzel" or "Anna Menzel" (German names) or "Deniz Gül" or "Selma Gül" (Turkish names).

Dependent variable

The dependent measure was the number of behaviors checked (out of ten). The ten behaviors were taken from a pilot study in which 86 teacher students (77% female; $M_{\text{age}} = 23.96$, $SD_{\text{age}} = 2.8$) rated a pool of 61 behaviors representing possible student behavior as either fostering or impeding learning. Ten behaviors that were unambiguously classified as fostering learning were used for the present study ($M < 2.0$ and $SD < 1.0$ on a 7-Likert scale where 1 = *fosters learning* and 7 = *impedes learning*). For example, the list included behaviors such as "prepares himself/herself systematically for tests" and "continuously works on the material taught in class" (see Table A1 for all ten behaviors).

Results

We conducted a Target ethnicity \times Target gender \times Standards ANOVA with number of behaviors checked as the dependent variable (see Table 1). Contrary to hypotheses, the interaction effect of gender and standard was non-significant, $F(1, 272) = 0.31$, $p = .576$, and neither was the main effect, $F(2, 272) = 0.28$, $p = .867$. Moreover, the three-way interaction effect between gender, ethnicity, and standards that was included for exploratory purposes was not significant, $F(2, 272) = 0.02$, $p = .896$. Therefore, the effect of gender was not further examined.

As expected, the ANOVA revealed a non-significant main effect of ethnicity, $F(1, 272) = 0.23$, $p = .635$, but a significant interaction between ethnicity and standard, $F(1, 272) = 6.63$, $p = .011$, partial $\eta^2 = .024$. While minimum standards were lower for a student with a Turkish name ($M = 3.82$, $SD = 1.56$, $n = 72$) than for a student with a German name

Table 1 Analysis of variance for requested number of behaviors fostering learning

Variable	$F(1, 272)$	p -value	Partial η^2
Standard	0.02	.889	<.001
Target student gender	0.28	.867	<.001
Target student ethnicity	0.23	.635	.001
Target student gender \times standard	0.31	.576	.001
Target student ethnicity \times standard	6.63*	.011	.024
Target student gender \times target student ethnicity	2.61	.108	.009
Target student gender \times target student ethnicity \times standard	0.02	.896	<.001

* $p < .05$

($M=4.34$, $SD=1.56$, $n=67$), confirmatory standards were higher for a student with a Turkish name ($M=4.28$, $SD=1.34$, $n=74$) than for a student with a German name ($M=3.93$, $SD=1.58$, $n=67$). To test whether the groups differed from one another as expected, we ran two planned contrasts, which we evaluate using one-tailed tests (Furr & Rosenthal, 2003). The first planned contrast tested whether the number of behaviors checked was higher for a student with a Turkish name compared to a German name in the minimum standards condition, and this contrast was significant, $t(276)=2.04$, $p_{\omega^2}=.021$, Cohen's $d=0.35$). The second contrasts tested whether in the confirmatory standard condition the number of behaviors checked would be greater for students with a Turkish as compared to a German name. This contrast was marginally significant, $t(276)=1.41$, $p_{\omega^2}=.080$, Cohen's $d=0.24$. Taken together, when participants were instructed to only get a vague idea whether the respective student might be suited for the academic track, participants were less strict with their expectations what kind of behaviors a student with a Turkish name should display compared to a student with a German name. At the same time, participants tended to be stricter and tended to need more proof that a Turkish student would definitively be suited for the highest track than when considering a German student.

Discussion

In the present study, we examined for the first time whether shifting standards play a role in the tracking judgments teacher students make for male and female students as well as for students with and without a migration background. We expected that the certainty with which a judgment needed to be made would influence whether judgments would be biased for or against students belonging to groups negatively stereotyped in this domain (i.e., boys, students with a Turkish migration background). When teacher students only needed to get a general idea of the students' suitability, they would require less behavioral proof of suitability for students who belong to negatively stereotyped groups than for students belonging to positively stereotyped groups. In contrast, when teacher students needed to be certain of suitability, they would require more behavioral proof of students belonging to negatively stereotyped groups. To our knowledge, this is the first study to test the difference between minimum and confirmatory standards in the German context and the first internationally to apply the SSM to tracking judgments.

We found evidence that teacher students shifted their standards for tracking judgments based on the migration background of the student. This shift was more pronounced when

teacher students were asked to indicate whether they suspected a student to be suitable to the academic track: In this case, teacher students needed to see fewer positive learning-related behaviors for students with a Turkish migration background than for students without a migration background. When teacher students needed to be certain of the suitability of a student, they tended to require more behavioral proof for the student with a Turkish migration background than the student without a migration background. These results are consistent with the predictions of the SSM (Biernat & Manis, 1994; Collins et al., 2009) and complement prior research demonstrating shifting standards more generally (e.g., Biernat & Fuegen, 2001; Biernat et al., 2008; Holder & Kessels, 2017). Thus, they are in line with the expectation that negative stereotypes about the academic abilities of Turkish-Germans (Bonefeld & Karst, 2020) result in lower expectations for this group and a subsequent shift in the amount of proof needed for a decision depending on the certainty with which the decision needs to be made.

With regard to gender, we did not find the anticipated effects of the certainty with which teacher students needed to make their judgments. In contrast to research on standard shifts based on gender focusing suitability for a job (Biernat & Kobrynowicz, 1997) or competence in mathematics (Holder & Kessels, 2017), teacher students required an equal number of positive learning-related behaviors in boys and girls in order to recommend them for the academic track and did so regardless of the standard. This might be due to the content of stereotypes about boys: While they are seen as less academically minded, they are also seen as—and see themselves as—more intelligent (e.g., Steinmayr & Spinath, 2009). To some extent, this positive stereotype might counterbalance the perception that boys might be less suited to the academic track. Moreover, past research demonstrated that the direction of gender biases in teacher students' recommendations, i.e., whether they favor boys or girls, also depend on the students' achievement level and trajectory (Klapproth & Fischer, 2019).

In our study, minimum or confirmatory standards were invoked experimentally. The results of this study clearly demonstrate the need to understand whether the de facto standards applied to tracking recommendations in Germany resemble minimum or confirmatory standards. However, it is not immediately clear which standard is currently used in Germany, and as federal states vary in the laws governing the transition process, standards may also vary between them. Therefore, we first elaborate on impacts of different standards on students before discussing more generally which standard may be prevalent in Germany today.

For students, especially ethnic minority students, the difference between minimum and confirmatory standards can be highly influential for their further educational path. If confirmatory standards are applied and ethnic minority students are less likely to receive a recommendation for the academic track, they will be placed in less challenging learning environments (e.g., Retelsdorf et al., 2010; van Houtte, 2004) and may not reach their full potential. If, on the other hand, a minimum standards is used, this could reduce the underrepresentation of students with a Turkish migration background in academic tracks. However, if a student is ultimately not suited to the academic track, they may experience repeated failures, which has been theorized to be a risk factor for students' motivation, for example, due to lower self-efficacy and learned hopelessness (Au et al., 2009, 2010).

Thus, the implications for students differ depending on the standard teachers currently use in Germany for tracking recommendations—and whether they find it more problematic to send a child to a track that is below or above its abilities. If teachers are wary of underestimating children, they may use minimum standards—recommending all children who they suspect may be suited to the academic track. However, if teachers are mainly concerned

not to overestimate students, they may apply confirmatory standards—recommending only those students who they are certain are suitable to the academic track.

On a theoretical level, the SSM predicts that rewards that are considered unlimited (e.g., praise) will be given to members of negatively stereotyped groups more readily, whereas rewards that are restricted in number will be preferentially given to members of positively stereotyped groups (Biernat & Vescio, 2002). While at a given point in time there may not be an infinite number of spots available at academic track schools, the number of recommendations that can be distributed in a single elementary school classroom is not restricted. Therefore, teachers may be more likely to apply a minimum standard with regard to this in principle limitless resource. The overall increase in students being recommended for the academic track in recent decades (Dietze, 2011) suggests that teachers may be applying minimum standards more and more. Importantly, qualitative and quantitative research has shown that teachers do vary in the leniency of their recommendations (Baeriswyl et al., 2011; Maier, 2007; Pohlmann-Rother, 2010), but does not allow conclusions regarding the standard used more generally.

Additionally, structural factors may influence whether minimum or confirmatory standards are used. In some German federal states, teachers' recommendations are binding, whereas in other states, parents can ultimately decide the secondary school track for their child. If teachers' recommendations are binding, they are held accountable for their decision to a greater extent, that is, they may need to justify their decision to others (Tetlock, 1983). This may make them more cautious not to hinder students from attending more challenging tracks or not to run the risk of being in conflict with parents disagreeing with the decision. Experimental evidence suggests that greater accountability increases judgment accuracy in teachers (Pit-ten Cate et al., 2020) and reduces negative biases against ethnic minority and low-SES students (Glock et al., 2012; Krolak-Schwerdt et al., 2013; Pit-ten Cate et al., 2016), though the authors link this effect not to different standards, but less reliance on heuristics such as group membership. While large-scale studies show that binding teacher recommendations lead to lower inequality, for example, based on social class, in children's actual *enrollment* in a specific secondary track (e.g., Dollmann, 2011), no such research has been conducted regarding the effect of binding decisions on teacher *recommendations* themselves. However, Dietze (2011) showed that teachers have historically given more academic track recommendations overall in German federal states where recommendations are binding, which may be due to teachers applying minimum standards with regard to students' suitability. However, this difference has gotten smaller in the decade before Dietze's study in the subset of federal states he examined.

Overall, prior research provides some indication that the standards applied in tracking recommendations might be minimal and that greater accountability could lead to greater use of minimum standards, though neither notion has been tested directly. Thus, understanding whether teachers generally apply minimum or confirmatory standards, is central to understanding the impact of shifting standards in tracking recommendations and its implications. Moreover, different standards may be applied for different groups: The absence of a bias against ethnic minority students in de facto tracking recommendations once achievement (and sometimes social class) is accounted for speaks against the use of confirmatory standards. However, students from low-SES families still receive more lower-track recommendations once achievement is controlled (Ditton et al., 2005; Schneider, 2011), which is indicative of confirmatory standards. Thus, intersectional analyses particularly of migration background and social class are needed in the future.

Limitations

Two characteristics of the sample may limit generalizability of the results. First, due to an oversight, participants' gender was not collected. At the time the research was carried out, teacher students at the university were 65% female. As stated above, 95% of participants spoke German as a mother tongue. Overall then, the sample likely consisted mainly of female teacher students without a migration background, a group that is positively stereotyped when it comes to academic achievement. Prior research established that those who are stereotyped to possess a trait (e.g., suitability to academic pursuits) shift their evaluations the most (Biernat et al., 2008). Therefore, our sample might have been especially likely to shift standards with regard to students with a Turkish migration background.

Secondly, our sample consisted of teacher students rather than teachers practicing in schools. Teachers accumulate expertise over the years, which may improve their decision accuracy in tracking recommendations. Indeed, experienced teachers focus less on information not directly related to achievement (father's occupation, migration background, non-learning related behaviors) than teacher students when making recommendations, resulting in recommendations less biased against students with a low socioeconomic status (Böhmer et al., 2012, 2017; but see also Glock et al., 2015). Therefore, they may exhibit less bias in recommendations regardless of the implied standard. While research on shifting standards in samples with experienced teachers is warranted, the demonstration of such an effect in teacher students is highly relevant. After all, expertise is acquired over years of experience—and until novice teachers have acquired such an expertise, they will have influenced many children's lives through their recommendations.

Conclusion

The results of the present study demonstrate for the first time that the direction of ethnic biases in tracking judgments vary depending on the framing of standards as minimal or confirmatory. Specifically, students with a Turkish migration background needed to provide less behavioral proof of their suitability for the academic track when standards were minimal, but in tendency needed to provide more proof than their peer without a migration background when standards were confirmatory. Contrary to our predictions and prior empirical findings, no differential effects by the gender of the student were observed.

This research adds to our understanding of the complexity of the effects of social category membership in the school context. Rather than indicating the presence of straightforward positive and negative biases, it shows that the framing of the standard is another factor which influences teachers' evaluations of students' academic potential. Developing an evidence-based understanding of the optimal standard of certainty for unbiased tracking recommendations is a necessary precondition to further improvements both on the level of the school system and individual teachers. Potentially, creating clearer guidelines for tracking should lead to a consistent standard that neither disadvantages nor advantages students who are members of various social categories. For example, the complementary reliance on results of standardized cognitive ability tests would reduce social inequality based on parents' education and, to a lesser extent, migration background in tracking recommendations (Steinmayr et al., 2017). Though such an intervention does not reduce the existing

differences in performance between students of different backgrounds, it may ensure that they are not further exacerbated by teachers' recommendations.

Appendix

Table A1 Complete list of example behaviors in the behavioral checklist

Continuously works on the material taught in class <i>Arbeitet kontinuierlich den durchgenommenen Stoff auf</i>
Continues to work through difficult tasks that he/she does not instantly succeed in <i>Beißt sich auch durch schwierigere Aufgaben, die nicht auf Anhieb gelingen</i>
Participates in class with appropriate verbal contributions <i>Beteiligt sich am Unterricht mit passenden Wortbeiträgen</i>
Takes good notes <i>Fertigt gute Mitschriften an</i>
Asks the teacher or fellow students (for help) if there is something he/she does not yet understand <i>Fragt die Lehrkraft oder Mitschüler, wenn er etwas nicht verstanden hat</i>
Prepares himself/herself systematically for tests <i>Geht planvoll bei der Vorbereitung von Klassenarbeiten vor</i>
Asks friends or siblings to test him/her to prepare for an exam <i>Lässt sich vor einem Test von Freunden oder Geschwistern abfragen</i>
Meets up with friends to study for school <i>Trifft sich mit Freunden zum Lernen</i>
Considers the most sensible plan and structure when working in a group <i>Überlegt sich bei Gruppenarbeiten eine sinnvolle Planung und Struktur des Vorgehen</i>
Comes up with his/her own examples to apply new vocabulary <i>Überlegt sich selbst Beispiele für die Anwendung von neuen Vokabeln</i>

Acknowledgements We would like to thank Liesel Heiermann for organizing and conducting the data collection.

Funding Open Access funding enabled and organized by Projekt DEAL. Open Access funding provided by Projekt DEAL.

Availability of data, material, and code Data, material, and code are available upon request.

Declarations

Competing interests The authors declare no competing interests.

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/>.

References

- Anders, Y., McElvany, N., & Baumert, J. (2010). Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule: Wie differenziert urteilen Lehrkräfte? [The evaluation of learning-related behaviors at the time of transition to secondary schools: How differentiated are teachers' judgments?] In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 313–330). Bundesministerium für Bildung und Forschung.
- Anderson-Clark, T. N., Green, R. J., & Henley, T. B. (2008). The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology, 27*(1), 94–99. <https://doi.org/10.1177/0261927X07309514>
- Arbuckle, C., & Little, E. (2004). Teachers' perceptions and management of disruptive classroom behaviour during the middle years (years five to nine). *Australian Journal of Educational and Developmental Psychology, 4*, 59–70.
- Au, R. C. P., Watkins, D. A., & Hattie, J. A. C. (2010). Academic risk factors and deficits of learned hopelessness: A longitudinal study of Hong Kong secondary school students. *Educational Psychology, 30*(2), 125–138. <https://doi.org/10.1080/01443410903476400>
- Au, R. C. P., Watkins, D., Hattie, J., & Alexander, P. (2009). Reformulating the depression model of learned hopelessness for academic outcomes. *Educational Research Review, 4*(2), 103–117. <https://doi.org/10.1016/j.edurev.2009.04.001>
- Baeriswyl, F., Wandeler, C., & Trautwein, U. (2011). "Auf einer anderen Schule oder bei einer anderen Lehrkraft hätte es für's Gymnasium gereicht": Eine Untersuchung zur Bedeutung von Schulen und Lehrkräften für die Übertrittsempfehlung [„At a different school or with a different teacher it would have been enough for the academic track“: A study on the importance of schools and teachers for the tracking recommendation]. *Zeitschrift Für Pädagogische Psychologie, 25*(1), 39–47. <https://doi.org/10.1024/1010-0652/a000028>
- Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis)perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of Educational Psychology, 107*(3), 805–820. <https://doi.org/10.1037/edu0000008>
- Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus [School structure and the development of differential environments for learning and development]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 95–188). VS Verlag für Sozialwissenschaften.
- Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2006). Leistungszuwachs in Mathematik: Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? [Performance gains in mathematics: Evidence of a scissor effect in the tracked school system?] *Zeitschrift für Pädagogische Psychologie, 20*(4), 233–242. <https://doi.org/10.1024/1010-0652.20.4.233>
- Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*(3), 682–699. <https://doi.org/10.1037/a0027608>
- 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*(3), 9–25. <https://doi.org/10.14712/23363177.2017.8>
- Bellenberg, G. (with Forell, M.) (2012). *Schulformwechsel in Deutschland. Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I* [Changing school tracks in Germany. Permeability and selection in the 16 school system of the federal states in middle school]. Bertelsmann Stiftung. http://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/GP_Schulformwechsel_in_Deutschland.pdf
- Biernat, M. (1995). The shifting standards model: Implications of stereotype accuracy for social judgment. In Y.-T. Lee, L. J. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 87–114). American Psychological Association. <https://doi.org/10.1037/10495-004>
- Biernat, M. (2012). Stereotypes and shifting standards: Forming, communicating, and translating person impressions. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology* (45th ed., pp. 1–60). Elsevier.
- Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57*(4), 707–724. <https://doi.org/10.1111/0022-4537.00237>

- Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin*, 36(7), 855–868. <https://doi.org/10.1177/0146167210369483>
- Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, 72(3), 544–557. <https://doi.org/10.1037//0022-3514.72.3.544>
- Biernat, M., & Ma, J. E. (2005). Stereotypes and the confirmability of trait concepts. *Personality and Social Psychology Bulletin*, 31(4), 483–495. <https://doi.org/10.1177/0146167204271712>
- Biernat, M., Ma, J. E., & Nario-Redmond, M. R. (2008). Standards to suspect and diagnose stereotypical traits. *Social Cognition*, 26(3), 288–313. <https://doi.org/10.1521/soco.2008.26.3.288>
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66(1), 5–20. <https://doi.org/10.1037/0022-3514.66.1.5>
- Biernat, M., & Vescio, T. K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. *Personality and Social Psychology Bulletin*, 28(1), 66–77. <https://doi.org/10.1177/0146167202281006>
- Böhmer, I., Gräsel, C., Hörstermann, T., & Krolak-Schwerdt, S. (2012). Die Informationssuche bei der Erstellung der Übergangsempfehlung – Die Rolle von Fallkonsistenz und Expertise [Information search during the formation of the tracking recommendation—The role of case consistency and expertise]. *Unterrichtswissenschaft: Zeitschrift für Lernforschung*, 40(2), 140–155.
- 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* (pp. 263–284). Springer Nature. https://doi.org/10.1007/978-3-319-50030-0_9
- Bonefeld, M., & Dickhäuser, O. (2018). (Biased) grading of students' performance: Students' names, performance level, and implicit attitudes. *Frontiers in Psychology*, 9, 481. <https://doi.org/10.3389/fpsyg.2018.00481>
- Bonefeld, M., Dickhäuser, O., Janke, S., Praetorius, A. K., & Dresel, M. (2017). Migrationsbedingte Disparitäten in der Notenvergabe nach dem Übergang auf das Gymnasium [Disparities in grades after the transition to the academic track based on migration background]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, 49(1), 11–23. <https://doi.org/10.1026/0049-8637/a000163>
- Bonefeld, M., & Karst, K. (2020). Döner vs. Schweinebraten – Stereotype von (angehenden) Lehrkräften über Personen deutscher und türkischer Herkunft im Vergleich [Döner vs. pig roast—Stereotypes of (preservice) teachers about people of German and Turkish descent in comparison]. In S. Glock, & H. Kleen (Eds.), *Stereotype in der Schule* (pp. 159–190). Springer VS. <https://doi.org/10.1007/978-3-658-27275-3>
- Bos, W., Hornberg, S., Arnold, K. H., Faust, G., Fried, L., Lankes, E. M., Schwippert, K., & Valtin, R. (2007). *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [IGLU 2006. International comparison of German elementary school students' reading competencies]. Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (2004). Schullaufbahneempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe [Tracking recommendations of teachers for children at the end of fourth grade]. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin, & G. Walther (Eds.), *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (pp. 191–228). Waxmann.
- Collins, E. C., Biernat, M., & Eidelman, S. (2009). Stereotypes in the communication and translation of person impressions. *Journal of Experimental Social Psychology*, 45(2), 368–374. <https://doi.org/10.1016/j.jesp.2008.09.005>
- Dietze, T. (2011). *Zum Übergang auf weiterführende Schulen: Auswertung schulstatistischer Daten aus 10 Bundesländern* [About the transition to secondary schools: Examination of school-based data in 10 federal states]. Gesellschaft zur Förderung Pädagogischer Forschung e.V..
- Ditton, H., Krüsken, J., & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule [Education inequality—The contribution of family and school]. *Zeitschrift Für Erziehungswissenschaft*, 8(2), 285–304. <https://doi.org/10.1007/s11618-005-0138-x>
- Dollmann, J. (2011). Verbindliche und unverbindliche Grundschulempfehlungen und soziale Ungleichheiten am ersten Bildungsübergang [Binding and non-binding tracking recommendations and social inequality at the transition to middle school]. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, 63(4), 595–621. <https://doi.org/10.1007/s11577-011-0148-z>

- Driessen, G., Slegers, P., & Smit, F. (2008). The transition from primary to secondary education: Meritocracy and ethnicity. *European Sociological Review*, 24(4), 527–542. <https://doi.org/10.1093/esr/jcn018>
- Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98(1), 198–208. <https://doi.org/10.1037/0022-0663.98.1.198>
- Federal Statistical Office (2019). *Statistisches Jahrbuch: Deutschland und Internationales* [Yearly statistical report: Germany and international information]. https://www.destatis.de/DE/Themen/Querschnitt/Jahrbuch/statistisches-jahrbuch-2019-dl.pdf?__blob=publicationFile
- Froehlich, L., Martiny, S. E., Deaux, K., & Mok, S. Y. (2016). “It’s their responsibility, not ours”: Stereotypes about competence and causal attributions for immigrants’ academic underperformance. *Social Psychology*, 47(2), 74–86. <https://doi.org/10.1027/1864-9335/a000260>
- Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics*, 2(1), 45–67. https://doi.org/10.1207/S15328031US0201_03
- Garcia, E. B., Sulik, M. J., & Obradović, J. (2019). Teachers’ perceptions of students’ executive functions: Disparities by gender, ethnicity, and ELL status. *Journal of Educational Psychology*, 111(5), 918–931. <https://doi.org/10.1037/edu0000308>
- Glock, S. (2016a). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers’ judgments of students. *Social Psychology of Education*, 19(3), 493–509. <https://doi.org/10.1007/s11218-016-9349-7>
- Glock, S. (2016b). Stop talking out of turn: The influence of students’ gender and ethnicity on preservice teachers’ intervention strategies for student misbehavior. *Teaching and Teacher Education*, 56, 106–114. <https://doi.org/10.1016/j.tate.2016.02.012>
- Glock, S., & Kleen, H. (2017). Gender and student misbehavior: Evidence from implicit and explicit measures. *Teaching and Teacher Education*, 67, 93–103. <https://doi.org/10.1016/j.tate.2017.05.015>
- Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers’ judgments. *Social Psychology of Education*, 16(1), 111–127. <https://doi.org/10.1007/s11218-012-9197-z>
- Glock, S., & Krolak-Schwerdt, S. (2014). Stereotype activation versus application: How teachers process and judge information about students from ethnic minorities and with low socioeconomic background. *Social Psychology of Education*, 17(4), 589–607. <https://doi.org/10.1007/s11218-014-9266-6>
- Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2012). Improving teachers’ judgments: Accountability affects teachers’ tracking decision. *International Journal of Technology and Inclusive Education*, 1(2), 89–98. <https://doi.org/10.20533/ijtie.2047.0533.2012.0012>
- Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013). Beyond judgment bias: How students’ ethnicity and academic profile consistency influence teachers’ tracking judgments. *Social Psychology of Education*, 16(4), 555–573. <https://doi.org/10.1007/s11218-013-9227-5>
- Glock, S., Krolak-Schwerdt, S., & Pit-ten Cate, I. M. (2015). Are school placement recommendations accurate? The effect of students’ ethnicity on teachers’ judgments and recognition memory. *European Journal of Psychology of Education*, 30(2), 169–188. <https://doi.org/10.1007/s10212-014-0237-2>
- Gresch, C., & Becker, M. (2010). Sozial- und leistungsbedingte Disparitäten im Übergangsverhalten bei türkischstämmigen Kindern und Kindern aus (Spät-)Aussiedlerfamilien [Disparities in transition behavior related to social and performance-related differences in children of Turkish descent and children from (late) remigrating families]. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule* (pp. 181–200). Bundesministerium für Bildung und Forschung.
- Hannover, B., & Kessels, U. (2011). Sind Jungen die neuen Bildungsverlierer? Empirische Evidenz für Geschlechterdisparitäten zuungunsten von Jungen und Erklärungsansätze [Are boys losing out in education? Empirical evidence of gender disparities to the disadvantage of boys and potential explanations]. *Zeitschrift Für Pädagogische Psychologie*, 25(2), 89–103. <https://doi.org/10.1024/1010-0652/a000039>
- Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, 37, 449–481. [https://doi.org/10.1016/S0883-0355\(03\)00015-6](https://doi.org/10.1016/S0883-0355(03)00015-6)
- Heyder, A., & Kessels, U. (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles*, 69(11–12), 605–617. <https://doi.org/10.1007/s11199-013-0309-9>
- Heyder, A., & Kessels, U. (2017). Boys don’t work? On the psychological benefits of showing low effort in high school. *Sex Roles*, 77(1–2), 72–85. <https://doi.org/10.1007/s11199-016-0683-1>
- Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers’ judgments: A new look from a shifting standards perspective. *Social Psychology of Education*, 20(3), 471–490. <https://doi.org/10.1007/s11218-017-9384-z>

- Jones, S., & Myhill, D. (2004). "Troublesome boys" and "compliant girls": Gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education*, 25(5), 547–561. <https://doi.org/10.1080/0142569042000252044>
- Jürges, H., & Schneider, K. (2011). Why young boys stumble: Early tracking, age and gender bias in the German school system. *German Economic Review*, 12(4), 371–394. <https://doi.org/10.1111/j.1468-0475.2011.00533.x>
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155.
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, 28, 73–84. <https://doi.org/10.1016/j.learninstruc.2013.06.001>
- Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, 109(6), 871–888. <https://doi.org/10.1037/edu0000156>
- Kessels, U. (2015). Bridging the gap by enhancing the fit: How stereotypes about STEM clash with stereotypes about girls. *International Journal of Gender, Science and Technology*, 7(2), 280–296.
- Kessels, U., & Steinmayr, R. (2013). Macho-man in school: Toward the role of gender role self-concepts and help seeking in school performance. *Learning and Individual Differences*, 23, 234–240. <https://doi.org/10.1016/j.lindif.2012.09.013>
- 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., Kärchner, H., & Glock, S. (2018). Do students' religion and school absences moderate the effect of ethnic stereotypes on school-placement recommendations? *The Journal of Experimental Education*, 86(2), 173–194. <https://doi.org/10.1080/00220973.2017.1293602>
- Kleen, H., & Glock, S. (2018). Good classroom management for ethnic minority students? A comparison of the strategies that preservice and inservice teachers use to respond to student misbehavior. *Journal of Classroom Interaction*, 53(2), 4–20.
- Krahé, B., Berger, A., & Möller, I. (2007). Entwicklung und Validierung eines Inventars zur Erfassung des Geschlechtsrollen-Selbstkonzepts im Jugendalter [Development and validation of an inventory to assess the gender role self-concept in adolescence]. *Zeitschrift Für Sozialpsychologie*, 38(3), 195–208. <https://doi.org/10.1024/0044-3514.38.3.195>
- KMK Secretary of the Standing Conference of the Ministers of Education and Cultural Affairs (2015). *Übergang von der Grundschule in Schulen des Sekundarbereichs I und Förderung, Beobachtung und Orientierung in den Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe)* [Transition from elementary school to secondary schools and facilitation, observation and orientation in fifth and sixth grade]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_02_19-Uebergang_Grundschule-SI-Orientierungsstufe.pdf
- Kristen, C. (2006). Ethnische Diskriminierung in der Grundschule? Die Vergabe von Noten und Bildungsempfehlungen [Ethnic discrimination in elementary school? Grading and tracking recommendations]. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, 58(1), 79–97. <https://doi.org/10.1007/s11575-006-0004-y>
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: A social cognitive analysis. *Social Psychology of Education*, 16(2), 215–239. <https://doi.org/10.1007/s11218-013-9215-9>
- Kuhl, P., & Hannover, B. (2012). Differenzielle Benotungen von Mädchen und Jungen: Der Einfluss der von der Lehrkraft eingeschätzten Kompetenz zum selbstgesteuerten Lernen [Differential grading of boys and girls: The influence of teacher-perceived competence for self-regulated learning]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, 44(3), 153–162. <https://doi.org/10.1026/0049-8637/a000066>
- Lehmann, R. H., Peek, R., & Gänfuß, R. (1997). *Aspekte der Lernausgangslage und der Lernentwicklung: Bericht über die Erhebung im September 1996 (LAU 5)*. [Aspects of the learning baseline and learning development: Report on the data collection in September 1996]. Amt für Schule, Behörde für Schule, Jugend und Berufsbildung.
- Lorenz, G., Gentrup, S., Kristen, C., Stanat, P., & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen [Stereotypes in teachers? A study on systematically biased teacher expectations]. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, 68, 89–111. <https://doi.org/10.1007/s11577-015-0352-3>

- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, 2(2), 99–106. <https://doi.org/10.1111/j.1750-8606.2008.00048.x>
- Maier, U. (2007). Systematische Lehrerereffekte bei Übergangsquoten auf weiterführende Schulen [Systematic teacher effects on transition rates for secondary schools]. *Zeitschrift Für Erziehungswissenschaft*, 10(2), 271–284. <https://doi.org/10.1007/s11618-007-0031-x>
- Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education*, 60(3), 245–256. <https://doi.org/10.1177/0004944116664618>
- McElvany, N., Kessels, U., Schwabe, F., & Kasper, D. (2017). Geschlecht und Lesekompetenz [Gender and reading competency]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 177–194). Waxmann.
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, 46(3), 235–261. <https://doi.org/10.1016/j.jsp.2007.05.001>
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>
- Neugebauer, M. (2011). Werden Jungen von Lehrerinnen bei den Übergangsempfehlungen für das Gymnasium benachteiligt? Eine Analyse auf Basis der IGLU-Daten [Are boys disadvantaged by female teachers' tracking recommendations for the academic track?]. In A. Hadjar (Ed.), *Geschlechtsspezifische Bildungsungleichheiten* (pp. 235–260). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92779-4_10
- Parks, F. R., & Kennedy, J. H. (2007). The impact of race, physical attractiveness, and gender on education majors' and teachers' perceptions of student competence. *Journal of Black Studies*, 37(6), 936–943. <https://doi.org/10.1177/0021934705285955>
- Pit-ten Cate, I. M., Hörstermann, T., Krolak-Schwerdt, S., Gräsel, C., Böhmer, I., & Glock, S. (2020). Teachers' information processing and judgement accuracy: Effects of information consistency and accountability. *European Journal of Psychology of Education*, 35, 675–702. <https://doi.org/10.1007/s10212-019-00436-6>
- Pit-ten Cate, I. M., Krolak-Schwerdt, S., & Glock, S. (2016). Accuracy of teachers' tracking decisions: Short- and long-term effects of accountability. *European Journal of Psychology of Education*, 31(2), 225–243. <https://doi.org/10.1007/s10212-015-0259-4>
- Pohlmann-Rother, S. (2010). Die Herausbildung der Übergangsempfehlung am Ende der Grundschulzeit [The formation of the tracking recommendation at the end of primary school]. *Zeitschrift Für Grundschulforschung*, 3(2), 136–148.
- Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3), 375–392. <https://doi.org/10.3102/0162373720932188>
- Rauch, D., Mang, J., Härtig, H., & Haag, N. (2016). Naturwissenschaftliche Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund [Science competency of students with a migration background]. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Köller (Eds.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation* (pp. 317–347). Waxmann.
- Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, 48(2), 335–360. <https://doi.org/10.3102/0002831210374874>
- Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, 82(4), 647–671. <https://doi.org/10.1111/j.2044-8279.2011.02051.x>
- Retelsdorf, J., Butler, R., Streblov, L., & Schiefele, U. (2010). Teachers' goal orientations for teaching: Associations with instructional practices, interest in teaching, and burnout. *Learning and Instruction*, 20(1), 30–46. <https://doi.org/10.1016/j.learninstruc.2009.01.001>
- Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung [The relevance of social class and migration background for teacher judgments exemplified by the tracking recommendation]. *Zeitschrift Für Erziehungswissenschaft*, 14(3), 371–396. <https://doi.org/10.1007/s11618-011-0221-4>
- Sneyers, E., Vanhoof, J., & Mahieu, P. (2018). Primary teachers' perceptions that impact upon track recommendations regarding pupils' enrolment in secondary education: A path analysis. *Social Psychology of Education*, 21(5), 1153–1173. <https://doi.org/10.1007/s11218-018-9458-6>

- Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence, personality and motivation? *Educational Research*, *56*(2), 230–243. <https://doi.org/10.1080/00131881.2014.898917>
- Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, *45*(1), 523–538. <https://doi.org/10.1007/s00181-012-0609-x>
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund [Students with a migration background]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller & M. Prenzel (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 200–230). Waxmann.
- Steinmayr, R., Michels, J., & Weidinger, A. F. (2017). *FA(IR)BULOUS – FAIRe Beurteilung des Leistungs-potenzials von Schülerinnen Und Schülern* [Fa(ir)bulous—fair assessment of performance potential of students]. TU Dortmund University.
- Steinmayr, R., & Spinath, B. (2009). What explains boys' stronger confidence in their intelligence? *Sex Roles*, *61*(9–10), 736–749. <https://doi.org/10.1007/s1199-009-9675-8>
- Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A Meta-Analysis. *Journal of Educational Psychology*, *99*(2), 253–273. <https://doi.org/10.1037/0022-0663.99.2.253>
- Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, *45*(1), 74–83. <https://doi.org/10.1037/0022-3514.45.1.74>
- Tiedemann, J., & Billmann-Mahecha, E. (2007). Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I [On the influence of migration background and social class membership on the tracking recommendation for middle school]. *Zeitschrift Für Erziehungswissenschaft*, *10*(1), 108–120. <https://doi.org/10.1007/s11618-007-0009-8>
- 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*(4), 459–478. <https://doi.org/10.1111/bjep.12087>
- Tobisch, A., & Dresel, M. (2017). Negatively or positively biased? Dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds. *Social Psychology of Education*, *20*(4), 731–752. <https://doi.org/10.1007/s11218-017-9392-z>
- Van Houtte, M. (2004). Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff. *American Journal of Education*, *110*(4), 354–388. <https://doi.org/10.1086/422790>
- Veenstra, R., Lindenberg, S., Oldehinkel, A. J., De Winter, A. F., Verhulst, F. C., & Ormel, J. (2008). Prosocial and antisocial behavior in preadolescence: Teachers' and parents' perceptions of the behavior of girls and boys. *International Journal of Behavioral Development*, *32*(3), 243–251. <https://doi.org/10.1177/0165025408089274>

Authors themes of research and publications **Anna K. Nishen**. Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195, Berlin, Germany. E-mail address: anna.nishen@fu-berlin.de

Current themes of research:

Current research encompasses evaluations of and feedback given to ethnic minority and majority students as well as teacher students' concerns relating to interethnic interactions.

Most relevant publications in the field of Psychology of Education:

Nishen, A. K. & Kessels, U. (2021). Non-communicated judgments of, vs. feedback on, students' essays: Is feedback inflation larger for students with a migration background? *Social Psychology of Education* (Advance online publication). <https://doi.org/10.1007/s11218-021-09674-3>

Marksteiner, T., **Nishen, A. K.**, & Dickhäuser, O. (2021). Students' perception of teachers' reference norm orientation and cheating in the classroom. *Frontiers in Psychology*, *12*, Article 614199. <https://doi.org/10.3389/fpsyg.2021.614199>

Katja Corcoran. Institute of Psychology, Karl-Franzens-Universität Graz, Universitätsplatz 2, 8010, Graz, Austria

Current themes of research:

Research on effects of social comparison on self-evaluation, motivation, goals, and emotions. Furthermore, she studies behavior change techniques and user-participation in smart energy systems.

Most relevant publications in the field of Psychology of Education:

Corcoran, K., Kedia, G., Illemann, R., & Innerhofer, H. (2020). Affective consequences of social comparisons by women with breast cancer: An experiment. *Frontiers in Psychology*, *11*, Article 1234. <https://doi.org/10.3389/fpsyg.2020.01234>

Corcoran, K. (2013). The efficiency of similarity-focused comparisons in person perception. *The Journal of Social Psychology*, *153*(2), 127–130. <https://doi.org/10.1080/00224545.2012.725111>

- Corcoran, K.**, Crusius, J., & Mussweiler, T. (2011). Social comparison: Motives, standards, and mechanisms. In D. Chadee (Ed.), *Theories in social psychology* (pp. 119-139). Wiley-Blackwell.
- Corcoran, K.**, Hundhammer, T., & Mussweiler, T. (2009). A tool for thought! When comparative thinking reduces stereotyping effects. *Journal of Experimental Social Psychology*, *45*, 1008-1011. <https://doi.org/10.1016/j.jesp.2009.04.015>
- Corcoran, K.**, & Mussweiler, T. (2009). Comparative thinking styles in group and person perception: One mechanism - many effects. *Social and Personality Psychology Compass*, *3*, 244-259. <https://doi.org/10.1111/j.1751-9004.2009.00173.x>
- Mussweiler, T., **Rüter, K.**, & Epstude, K. (2004). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology*, *87*(6), 832-844. <https://doi.org/10.1037/0022-3514.87.6.832>
- Katharina Holder**. Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195, Berlin, Germany
Current themes of research:
 Research interests focus on data-based decisions in school and the compatibility of inclusion and educational standards.
Most relevant publications in the field of Psychology of Education:
- Holder, K.**, & Kessels, U. (2019). Unterrichtsgestaltung und Leistungsbeurteilung im inklusiven und standardorientierten Unterricht aus der Sicht von Lehrkräften [Teachers' perspectives on instructional quality and performance evaluation in the context of inclusion and educational standards]. *Zeitschrift für Erziehungswissenschaft*, *22*, 325-346. <https://doi.org/10.1007/s11618-018-0854-7>
- Holder, K.**, & Kessels, U. (2018). Lehrkräfte zwischen Bildungsstandards und Inklusion: Eine experimentelle Studie zum Einfluss von „Standardisierung“ und „Individualisierung“ auf die Bezugsnormorientierung [Between educational standards and inclusion: An experimental study on the effects of “standardization” and “individualization” on teachers' reference norm orientations]. *Unterrichtswissenschaft*, *46*, 87-104. <https://doi.org/10.1007/s42010-018-0013-0>
- Holder, K.**, & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards' perspective. *Social Psychology of Education*, *20*, 471-490. <https://doi.org/10.1007/s11218-017-9384-z>
- Ursula Kessels**. Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195, Berlin, Germany
Current themes of research:
 Research interests encompass the role of gender and ethnic stereotypes for students' self-concept, academic outcomes, and evaluations; epistemological beliefs and beliefs about science; STEM subjects; school-burnout.
Most relevant publications in the field of Psychology of Education:
- Kessels, U.**, & Heyder, A. (2020). Not stupid, but lazy? Psychological benefits of disruptive classroom behavior from an attributional perspective. *Social Psychology of Education*, *23*, 583-613. <https://doi.org/10.1007/s11218-020-09550-6>
- Herrmann, J., Koeppe, K., & **Kessels, U.** (2019). Do girls take school too seriously? Investigating gender differences in school burnout from a self-worth perspective. *Learning and Individual Differences*, *69*, 150-161. <https://doi.org/10.1016/j.lindif.2018.11.011>
- Kessels, U.**, Heyder, A., Latsch, M., & Hannover, B. (2014). How gender differences in academic engagement relate to students' gender identity. *Educational Research*, *56*(2), 219-228. <http://dx.doi.org/10.1080/00131881.2014.898916>
- Heyder, A., & **Kessels, U.** (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles*, *69* (11-12), 605-617. <https://doi.org/10.1007/s11199-013-0309-9>
- Kessels, U.**, Rau, M., & Hannover, B. (2006). What goes well with physics? Measuring and altering the image of science. *British Journal of Educational Psychology*, *74* (4), 761-780. <https://doi.org/10.1348/000709905X59961>
- Kessels, U.** (2005). Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education*, *20*, 309-323. <https://doi.org/10.1007/BF03173559>