Check for updates

Changing preservice teacher students' stereotypes and attitudes and reducing judgment biases concerning students of different family backgrounds: Effects of a short intervention

Sabine K. Lehmann-Grube[1] · Anita Tobisch[1] · Markus Dresel[1]

Abstract
Numerous empirical findings have shown biased judgments of (future) teachers depending on students' ethnic and social background. Furthermore, research has indicated that (future) teachers' stereotypes and attitudes differ depending on students' backgrounds and appear to influence (future) teachers' judgments. Based on theories of stereotype change, attitude change, and judgment formation, a short intervention was developed to change stereotypes and attitudes and to reduce judgment biases. In an experimental study (within- and between-subject design) with $N = 215$ preservice teacher students, the effectiveness of the intervention on stereotype change, attitude change, and reduction of judgment distortions was tested. The results showed hypothesized effects of the intervention on stereotypes and attitudes towards students with an immigration background and students with low social status. Furthermore, the intervention showed effects on preservice teacher students' judgments, especially for low-status students.

Keywords Preservice teacher students · Attitudes · Stereotypes · Judgments · Students' immigration background · Students' social status

1 Introduction

"Schools have not always had the mission to support achievement for all students, and children's assignments to school and classrooms have, during many periods in history, fostered segregation rather than encouraging inclusion" (Banks et al., 2005, p. 232).

✉ Sabine K. Lehmann-Grube
sabine.lehmann-grube@uni-a.de

1 Department of Psychology, University of Augsburg, Universitätsstraße 10, 86159 Augsburg, Germany

🖄 Springer

Todays' mission of schools is to provide students equal opportunities to learn, succeed, and progress regardless of social characteristics, e.g., family background in terms of immigration or socio-economic status. However, educational (and vocational) success is still associated with family background (Organisation for Economic Co-operation and Development [OECD], 2019). In Germany, this association was higher than in many other countries, particularly concerns a lower social status and a Turkish immigration background (e.g., Weis et al., 2019), which correlate highly, and students' academic achievement only partially explained inequalities of educational success (e.g., Dumont et al., 2014; Maaz et al., 2009). These effects of family background on students' educational success even when controlling for effects of academic achievement are partially explained by parents' preferences, e.g., for secondary school tracks (Maaz et al., 2009). But they also point to social categorization processes in teacher-student interaction giving rise to differences in teachers' expectations, judgments, grading, and tracking recommendations depending on students' family background. Along with social categories, implicit and explicit social stereotypes and attitudes are always activated automatically (e.g., Bargh et al., 1992; Casper et al., 2010). However, whether or not a person, e.g., a teacher, automatically applies a social category or individuates their impression formation based on individual information depends on their motivation, personal resources, and situational resources (e.g., Brewer, 1988; Fiske & Neuberg, 1990). Hence, theories of impression formation offer three approaches to reduce social disparities in teacher judgments and improve diagnostic accuracy: (1) sensitize for influences of implicit and explicit social stereotypes and attitudes in impression formation and judgments; (2) provide conditions—and strategies—for in-depth information processing in judgment situations to inhibit these influences; (3) transform discriminatory social stereotypes and attitudes. The experimental study presented in this paper combined these three approaches in a short intervention for preservice teacher students (PTS). Participants in the control condition received a placebo-intervention. To assess effects of the intervention, PTS in both conditions rated three students with different family backgrounds (case vignettes) and stated their stereotypes about and attitudes towards students with Turkish vs. no immigration background and low vs. high social status at three time points (pre, post, and follow-up). A Turkish immigration background was chosen for the following reasons: Students with this background are the largest group of all immigration backgrounds in German schools, they are overrepresented in the lowest and underrepresented in the highest secondary school track (Blaeschke & Freitag, 2021), and this background correlates with lower socio-economic resources and lower educational levels of the parents, two important predictors of students' achievement levels (e.g., Weis et al., 2019). Consequently, many studies on educational disparities and teachers' judgment biases in the German context take this immigration background into account and provide one of the bases to test the effects of our intervention.[1]

---

[1] However, teachers have students with various immigration backgrounds and face various social and educational challenges posed by these students who themselves face differential stereotypes and attitudes of their teachers and peers.

2 Differences in (future) teachers' judgments, stereotypes, and attitudes depending on students' family background

Teachers' judgment accuracy, i.e., the correspondence of grading, ranking, and dispersion to objectively assessed student performances and characteristics (including students' self-reported motivation), has been shown to have large inter-individual differences, and seems, on average, only moderately good (e.g., Spinath, 2005; Südkamp et al., 2012). To explain inter-individual differences and quality, Südkamp et al. (2012) propose in their model of moderators of *Teacher Judgment Accuracy* to differentiate between students' characteristics and teachers' characteristics. However, they were not able to enter these differentiated characteristics into their meta-analysis of moderators due to missing data in most of the included primary studies. Concerning students' characteristics, disparities of educational success depending on family background and exceeding achievement differences (e.g., Dumont et al., 2014; Maaz et al., 2009; OECD, 2019) suggest that whether and how students' backgrounds influence teachers' judgments should be investigated. Next, if students' backgrounds are relevant, social cognition processes come into play on the side of teachers' characteristics, e.g., their stereotypes and attitudes towards different backgrounds may influence their judgments (e.g., Glock et al., 2013).

2.1 Differences in teachers' judgments depending on students' backgrounds

Studies investigating whether, how, and why teachers' judgments differ depending on students' backgrounds often employ assumptions of impression formation theories (e.g., Brewer, 1988; Fiske & Neuberg, 1990). Fiske and Neuberg's (1990) Continuum Model presupposes that automatic impression formation based on social categories is the default process in social interaction and potentially yields distorted judgments. Individuated impression formation based on a person's individual characteristics facilitates more accurate judgments but only occurs if the situation meets certain requirements: The person, e.g., a teacher, must be motivated (e.g., relevance or utility of a judgment, personal responsibility), have personal resources (e.g., knowledge, information, capacity), and situational resources (e.g., time to process information and act).

Numerous experimental studies employing case vignettes have indicated the influence of social categories and the associated stereotypes on teachers' judgments, e.g., grading, tracking recommendations, or further school relevant characteristics often including work behavior, school engagement, and motivation (e.g., Baadte, 2020; Bonefeld & Dickhäuser, 2018; Bonefeld et al., 2020; Civitillo et al., 2022; Glock, 2016; Glock & Kleen, 2023; Glock & Krolak-Schwerdt, 2013; Glock et al., 2012, 2013, 2015, 2016; Holder & Kessels, 2017; Tobisch & Dresel, 2017). For example, (future) teachers rated students without an immigration background and high social status better than students with an immigration background or low social status when only backgrounds varied (by student names) but not case content (e.g., Bonefeld & Dickhäuser, 2018; Civitillo et al., 2022; Glock et al., 2013; Tobisch & Dresel, 2017).

In some studies, biases were masked when judgments were framed by an individual reference norm (Holder & Kessels, 2017), were limited to stereotype consistent case vignettes (Glock, 2016; Glock & Krolak-Schwerdt, 2013), limited to or more pronounced for *estimations* of language proficiency than mathematics achievement (Glock, 2016; Glock & Kleen, 2023; Glock & Krolak-Schwerdt, 2013), but more pronounced for *expectations* of future achievement in math than in German exams (Tobisch & Dresel, 2017), or only occurred under low accountability conditions (e.g., Glock et al., 2012). In an experimental study manipulating gender, immigration background, and actual performance (test results of students in a virtual classroom about which participants were informed), Bonefeld et al. (2020) found a three-way interaction: PTS overestimated low test-performers and underestimated high test-performers, if they were male and had a Turkish immigration background. Likewise, Glock et al. (2015), who employed real case vignettes, found less accurate tracking recommendations for minority students when compared to majority students. Descriptively, Luxembourgish teachers and German PTS likewise underestimated above average (highest school track) and overestimated below average (lowest school track) minority students; and they felt less confident about their judgments for minority students.

Some field studies yielded similar patterns of judgment biases for students with comparable achievements. For example, teachers expected lower achievement development in the reading ability of Maori students (Rubie-Davies et al., 2006), a specifically stereotyped minority in New Zealand. Likewise, German primary school teachers expected less development in the reading ability of students with a Turkish immigration background (and some others), even when controlling for students' social status background, cognitive abilities, gender, self-reported motivation (Lorenz et al., 2016) and teachers' judgements of students' motivation (same data set, Gentrup et al., 2018). But teachers expected more development in math abilities of students with East European backgrounds. Results of further studies controlling the effects of students' immigration backgrounds on educational success for effects of (low) social status—often correlated with specific immigration backgrounds—differ when comparing similar studies in different regions and when comparing different studies (measures) in the same national context. In large scale studies on tracking decisions at the transition from primary to secondary school, social status effects only partially explained biased tracking of students with Portuguese and other immigration backgrounds in Luxembourg (Klapproth et al., 2012), but largely explained biased tracking of students with Turkish or with ethnic German repatriate immigration background in Germany (Gresch & Becker, 2010). Additionally, Stahl (2007) only found effects for parents' educational level, social status, and students' gender but not for mother-languages (German vs. non-German) on teachers' judgments of students' reading abilities.[2] In an experimental study by Tobisch and Dresel (2017) in the German context, which employed student case vignettes, the (Turkish) immigration background effect remained in teachers' judgments when compared to their

---

[2] Teachers' judgments of students' proficiency to express themselves in German were predicted by social status, language background, and gender but not controlled for students' achievements in an objective test.

judgments for students' of low social status (within-between-interaction) in contrast analyses. But Glock and Kleen (2023) found only a main effect of students' social status on PTS' judgments of target students' German language proficiency in a between subject design with four vignettes (fully crossed social status by immigration background). In their review of teacher expectancy studies published in English within the last 30 years and including experimental and field studies, but excluding studies of performance estimation in specific tasks (e.g., Hachfeld et al., 2010),[3] Wang et al. (2018) point to "a smaller number of studies, however, which showed inconsistent evidence" (p. 130), meaning studies which found no significant background biases.

The study from Glock et al. (2016) yielded first direct empirical evidence of stereotype and attitude influences on teachers' judgments of "German language proficiency, mathematical performance, intelligence, school engagement, social isolation, information processing, emotionality, and assertiveness" (p. 9): PTS with high positive implicit and explicit attitudes towards (and stereotypes about[4]) high social status judged the high-status student, presented in a case vignette, more positively and the low-status student more negatively on several of the judgment dimensions. Bonefeld and Dickhäuser (2018) included implicit attitudes concerning immigration background in their analyses and found an unexpected effect that positive (or less negative) implicit attitudes towards Turkish immigration background predicted more negative grading of a below average dictation of the student with this immigration background possibly indicating disappointment with the achievement of the implicitly preferred student group. Civitillo et al. (2022) reported effects of PTS' affective prejudices (feeling warm vs. cold towards targets) on their tracking recommendations, most strongly for self-identified Romani students and also, to a lesser degree, for self-identified Turkish students. On the other hand, Glock and Krolak-Schwerdt (2014) were able to induce stereotype activation in teachers and PTS but activated minority (ethnic as well as social status) stereotypes yielded no effects on subsequent achievement judgments for the target students.

2.2 Stereotypes about and attitudes towards students of various ethnic and social backgrounds

Hilton and Hippel (1996) define stereotypes as "beliefs about characteristics, attributes, and behaviors of members of a certain group" (p. 240). Beliefs do not per

---

[3] Hachfeld et al. (2010) reported a strong overestimation of bilingual students' performance on a linguistically demanding math text problem when compared to the lower overestimation of (immigrant and non-immigrant) monolingual students' performance on the same problem and when compared to the lowest overestimation of all students' performance on a linguistically less demanding math text problem. The latter rather speaks for teachers' general unawareness of linguistic demands in math word problems and other subject matters (c.f., Schleppegrell & O'Hallaron, 2011) than for a contrasting positive ethnic minority bias due to unawareness of students' linguistic difficulties, which teachers presumably would only consider if they were aware of the linguistic demands of tasks.

[4] Some items to assess explicit attitudes asked participants for their agreement to stereotype content, e.g., what students often do, the evaluation (only) implied by social norms for behavior in schools.

se entail an evaluation, but "stereotypes about out-group members are more likely to have negative connotations than those about in-group members" (p. 240; similarly Tajfel, 1969). Depending on the context, a stereotype may devaluate outgroup members' suitability for the given objectives of a situation or prospect. Concerning achievement relevant aspects, PTS associated less positive characteristics with students with an immigration background when compared to non-immigrant students and likewise less positive characteristics with low-status when compared to high-status students (e.g., Bonefeld & Karst, 2020; Tobisch & Dresel, 2020). Hence, stereotypes often do entail attitudes—defined as mental tendencies of evaluative reactions towards objects of thought (Bohner & Wänke, 2006; Eagly & Chaiken, 1993), i.e., persons, but also abstract entities like groups or teaching students with diverse backgrounds. Attitudes can be explicit (more accessible and easier to control) or implicit (more spontaneous and less conscious; e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006; Greenwald & Banaji, 1995). Both are important for (future) teachers' judgment formation (e.g., Glock et al., 2016) and may diverge (Hofmann et al., 2005). Implicit and explicit attitudes towards students with low social status seem to align as less positive when compared with attitudes towards students with high social status (e.g., Tobisch & Dresel, 2020). Implicit and explicit attitudes towards students with an immigration background seem to diverge in German school contexts. Studies showed comparably positive explicit attitudes of (future) teachers towards students with and without an immigration background, but negative implicit attitudes towards students with immigration backgrounds, i.e., significant in-group preferences (e.g., Glock & Karbach, 2015; Kleen et al., 2019).

Summing up the findings, we do know that students with a low social status background seem to face biased judgments and discriminatory stereotypes and attitudes consistently across regions and systems, while results concerning students with ethnic minority backgrounds are inconsistent and seem to depend on an interplay of region, system (e.g., tracking), students' ethnic-cultural origin, study design, and judgment domain (e.g., math or language proficiency). Influences of stereotypes and attitudes show up more strongly or only in low accountability judgments for stereotype-consistent or content-inconsistent student profiles and when based on an objective reference norm.

3 Reducing judgment distortions and changing stereotypes and attitudes

3.1 Reducing teachers' and PTS' judgment biases regarding students' backgrounds

Judgmental biases stem from inaccurate judgments and some approaches focus on improving teachers' diagnostic competence (e.g., Böhmer et al., 2017; Klug, 2011). For example, Klug (2011) developed a training concept comprising three sessions (each three hours) with small groups to train teachers' diagnostic competence and monitored their diagnostic actions with a diary. Most training concepts focusing

on teachers' diagnostic competence showed that it is possible to improve important general competencies for the diagnostic process. However, such training is often quite extensive, very specific with regard to school subjects, or does not address or at least control for social cognition processes that are possibly important in reducing specific distortions associated with particular student backgrounds. Pit-ten Cate, Krolak-Schwerdt, Hörstermann et al., (2016) were able to raise teachers' judgment accuracy for students with immigration backgrounds to the level of accuracy for majority students through workshops, one via instruction on impression formation and training in decision strategies including feedback, and the other with training in the application of formal decision rules, which also included feedback (summary in Krolak-Schwerdt et al., 2018). These results suggest that training general diagnostic skills is effective in reducing judgment biases associated with students' backgrounds in situations of high accountability and with enough time to integrate individual information. However, general diagnostic rules and skills may not work when teachers are less accountable, e.g., when grading class tests. In such situations, but also when information is sparse, responsibility is low, or motivation to process individual information is insufficient, judgments are formed more automatically, based on memory content and readily available cues, such as social categories (e.g., Fiske & Neuberg, 1990). Furthermore, if a stereotype for a social category is salient and available, people tend to rely on this information for judgment formation (Kruglanski & Ajzen, 1983). Thus, interventions to reduce judgment biases should address social perception and social information processing which both influence more automatic judgments. Namely, interventions should address the stereotypes and attitudes activated by social categories, e.g., students' backgrounds. In the German context, Baadte (2020) seems to be the first to undertake an attempt "to reduce the impact of social stereotypes (...) on the assessment of students' achievement" (pp. 1007–1008) with a counter-stereotype training adapted from Gawronski et al. (2008). In the training conditions, addressing students with either an Arabic immigration background, low social status, or males (in a domain associated with gender stereotypes, i.e., liking to read), PTS had to identify stereotype-*inconsistent* combinations of student names and learning relevant characteristics and react with "Yes"-answers in a computer-based task. Although the training only had an effect on PTS' judgment biases along with less recall of stereotype-*consistent* information (but not more recall of stereotype-inconsistent information) concerning students in the gender condition when compared to the control condition (arithmetic task), the results indicate that an intervention attempting to change stereotypes seems fruitful to reduce judgment biases. Because stereotypes and attitudes towards social groups are strongly linked (see below), interventions to change attitudes—or address both stereotypes and attitudes—may also reduce judgment biases.

3.2 Changing stereotypes and attitudes

The long history of research on stereotype and attitude change demonstrates the rather high stability of both constructs (i.e., mental structures) and relative

resistance to modification interventions, particularly with regard to habitual biases and discrimination (e.g., Devine et al., 2012). However, research in both fields has also yielded theoretical models of how to influence stereotypes and attitudes and has identified precursors and conditions.

The Intergroup-Contact-Theory emerged from approaches to reduce prejudices toward specific outgroups and their members, which was summarized by Allport in 1954. The term prejudice denotes a judgment, i.e., an evaluation and, therefore, also an attitude towards members of a specific group. Presumably, teachers form attitudes connected to these prejudices, e.g., towards having students with specific immigration backgrounds in class, and may feel compelled to invest special effort in lesson planning and classroom management. The psychological construct prejudice also entails the cognitive component of ascribing a pattern of characteristics to groups and persons, thus, also denoting a stereotype (Petersen & Six-Materna, 2006), e.g., associating certain immigration backgrounds with a low educational level of parents, low learning support of children and, subsequently, low German literacy of students. Consequently, intergroup contact represents an approach to change attitudes and stereotypes alike. Allport (1954) derived from his summary four necessary features of the contact situation in order to reduce prejudice: equal status between the groups in the situation; common goals; intergroup cooperation; and the support of authorities, law, or custom. Equal status does not apply to educational settings, by definition (e.g., Ohlsson, 2018), even if classroom-dynamics may empower students (e.g., Hao Kuo Tai, 1998). But the extensive meta-analyses of Pettigrew and Tropp (2006, 2008) revealed the effects of intergroup contacts not meeting all features deemed to be necessary, though these effects were smaller (2006), and yielded three mediators of prejudice reduction by intergroup contact: "(1) enhancing knowledge about the outgroup, (2) reducing anxiety about intergroup contact, and (3) increasing empathy and perspective taking" (2008, p. 922). These mediators, thus, may also ensue from classroom interaction even though the power resources of teachers and students are asymmetric and dynamic. Corresponding to dual process theories of attitude change, the inverse relationship between contact and negative attitudes, stereotypes, and prejudice was stronger if contact was structured by explicit (institutionalized) measures to reduce prejudice, e.g., instructions to reflect and elaborate on the contact, the outgroup, one's feelings and expectations. Furthermore, effects of extended (indirect, e.g., cross-group contact of friends), vicarious (observing cross-group contact of others), and imagined contact were found (Miles & Crisp, 2014). According to the meta-analysis of Miles and Crisp (2014), who tested the effects of imagined positive interaction with outgroup members, "the effect was equally strong for implicit and explicit attitude measures, but was stronger on behavioral intentions than on attitudes, supporting the direct link between imagery and action proposedly underlying mental simulation effects" (p. 2). Though, they excluded studies employing (only) perspective taking, imagining the mere presence of an outgroup member, or imagining a counter-stereotypical outgroup member, the effects of simulated interactions on behavioral intentions along with attitude changes are especially promising with regard to applying intergroup contact to reduce judgment biases.

The results of the meta-analyses about effects of intergroup contact on prejudice reduction (i.e., attitude and stereotype change), as expected, correspond with *models*

of stereotype change, which, in general, recur for cognitive effects of processing stereotype inconsistent information about (members of) a group (see Hilton & Hippel, 1996, for an overview), e.g., by imagining a counter-stereotypical "exemplar" of that outgroup. Juxtaposing different out-groups, namely "skinheads" vs. "homosexuals" (p. 261), in studies with opposite results on stereotype content priming by stereotype suppression, Hilton and Hippel (1996) highlighted that participants' personal standards not to stereotype a certain out-group may motivate subjects to inhibit priming effects of stereotype activation when responding to a targeted outgroup (e.g., "homosexuals"). Against the backdrop of professional standards recommending teachers to contribute to the reduction of educational disparities (e.g., Banks et al., 2005), we may well assume that PTS are highly motivated to control their impression formation for undesired effects of—often unconscious—attitude and stereotype activation when judging "out-group"-students, e.g., with immigration or lower social backgrounds. However, Devine and colleagues (e.g., 2012) presume that an explicit motivation not to discriminate does not suffice to break prejudice habits—eventually entailing discrimination, albeit possibly subtle. They developed a systematic training regime to raise awareness of one's own implicit preferences in the first step, to inform about contexts in which prejudice habits most likely operate unnoticed in the second step, and to stipulate the application of strategies counteracting these mechanisms in the third step. This resulted in longer lasting effects on implicit attitudes and concerns about discrimination but not on explicit attitudes and motivation not to discriminate.

Dual process theories of attitude change, e.g., the *Elaboration Likelihood Model* (ELM; Petty & Cacioppo, 1986) and the *Heuristic vs. Systematic Information Processing Model* (HSM; Chaiken, 1980), explain differences in extent, strength, and persistence of possible changes with two distinct forms of information processing similar to the poles of the *Continuum Model of Impression Formation* (Fiske & Neuberg, 1990). The superficial, more automatic form rather results in less stable and short-term or no attitude changes. The in-depth, more controlled form potentially results in more stable and rather long-term attitude changes. Necessary conditions for in-depth processing of information about the attitude object are, again, individual and situational resources as well as opportunities, namely capability (e.g., prior knowledge), motivation, time to process, and comprehensibility of the message (see Bohner & Wänke, 2006, for an overview).

Interventions created to change (future) teachers' stereotypes and attitudes towards (teaching) minority students and develop their competencies to account for student diversity in their classrooms have mostly been based on models of multicultural education (see Trent et al., 2008, for an overview), culturally responsive teaching (e.g., Banks et al., 2005; Gay, 2010), or human relations (see Castro, 2010, and Sleeter, 2001, for overviews). Scopes range from several weeks of extra-curricular programs (e.g., Haberman & Post, 1992), courses embedded in regular teacher education (e.g., Wasonga, 2005), often accompanied by field experiences (e.g., Agnello & Mittag, 1999; Lucas, 2011; Martin & Koppelman, 1991), up to complex curricula (e.g., Akiba, 2011; Colby et al., 2009; Grottkau & Nickolai-May, 1989; Kumar & Hamer, 2013; Sparks & Verner, 1995). Interpreting reports on effects of these interventions is complicated by their terminological arbitrariness (interchangeably) referring to attitudes, minority

stereotypes, and prejudicial beliefs or views, perspectives, as well as understanding of diversity, even in overviews (e.g., Castro, 2010; Trent et al., 2008) with some most agreeable exceptions who clearly differentiate these terms (e.g., Sleeter, 2001). However, when accounting for item-contents and interpreting results with caution concerning methodological weaknesses, results correspond to social-psychological theories of stereotype and attitude change, confirm meta-analyses, and suggest that in ecological settings it may be specifically fruitful to integrate theoretical approaches. Most effective in changing stereotypes and attitudes and in developing knowledge about students' backgrounds seem human relations/social foundations approaches that integrate systematic instruction on societal inequalities, personal intergroup contacts in field experiences, and reflection of stereotypes, attitudes, and personal feelings (e.g., Akiba, 2011; Colby et al., 2009; Grottkau & Nickolai-May, 1989; Kumar & Hamer, 2013; Martin & Koppelman, 1991; Sparks & Verner, 1995). Instructions without personal contact may also bring about attitude and knowledge changes (Wasonga, 2005). Personal intergroup contact in field experiences without systematic instructions seems to have no effect on multicultural attitudes (Sparks & Verner, 1995) or may increase stereotyping, even if PTS feel more comfortable with (teaching) minority students (Haberman & Post, 1992). A "meta-analytic integration of over 40 years of research on diversity training evaluation" (Bezrukova et al., 2016, p. 1227), including unpublished studies, yielded high effects on participant reactions (self-reports), the lowest effects on stereotype, attitude, and prejudice changes, and medium effects on behavioral changes and knowledge gains. Two experimental studies in higher education contexts yielded positive effects of *short interventions* inducing (indirect) intergroup contact on higher education students' empathy and feelings of White guilt (Soble et al., 2011) and on attributions of warmth and intentions to join collective action against discrimination (Kotzur et al., 2019).

4 Development of a short-intervention to change stereotypes and attitudes and reduce judgment biases

Summing up, research on (future) teachers' judgments of students' academic achievement and abilities has either addressed (a) the differences and the development of teachers' general diagnostic competence, i.e., accuracy, across their professional lives in the frame of teacher competence models or (b) teachers' judgment biases related to students' family backgrounds including directions, conditions, and predictors of biases in the frame of educational disparities. Studies investigating predictors of teachers' judgment biases reported influences of stereotypes and attitudes; and the effect of a counter-stereotype training on PTS' assessments of male students' academic performance (Baadte, 2020) suggests that judgment bias reduction may not only be approached by increasing diagnostic accuracy, but also by changing stereotypes and attitudes concerning students and their family backgrounds. Research on stereotype and attitude change (or prejudice reduction) in the frame of teacher education, e.g., multicultural education programs, as well as in the frame of experimental social psychology has generated models and effective interventions to change (negative) attitudes (see Bohner & Wänke, 2006, for an overview).

Integrating research on (a) (future) teachers' judgment biases, (b) influences of stereotypes and attitudes on judgment formation, and (c) stereotype and attitude modification suggests to combine the approaches in one intervention and to test for stereotype and attitude changes along with judgment changes concerning students of various backgrounds.

4.1 Overview

The short-intervention comprised three pages combining various elements derived from theories to change stereotypes, attitudes, and judgment biases: (1) At the beginning, PTS received feedback on their own attitudes, (2) followed by information in textbook format, and (3) finally strategies to implement insights in practice.

4.2 PTS' own attitudes

The intervention material started with the focus on PTS' own implicit and explicit attitudes towards students with a Turkish immigration background compared to students without an immigration background. After a short explanation of the relevant constructs, PTS saw a figure of their own implicit and explicit attitudes as measured in the pretest, because feedback plays a significant role in self-reflection and thus in changing cognitions and behavior (see Pit-ten Cate et al., 2014, for an overview).[5] The individual results were explained and interpreted with regard to possible effects for judgments of students with various backgrounds, similar to the feedback element at the beginning of the training to break prejudice habits developed by Devine et al. (2012).[6] In this context, depending on the pattern of individual results, a reference was made to the risk of either negative or positive discrimination.

4.3 Textbook information

The second section, under the heading *The influence of attitudes and stereotypes on judgments* included theoretical foundations about automatic and controlled judgment formation and empirical results about judgment distortions and biases depending on students' backgrounds (somewhat similar to information about the idea of prejudice habits in the training of Devine et al., 2012). According to Pit-ten Cate et al. (2014), it can be assumed that "in order to change teachers' cognitive processes, teachers have to be informed about the processes, which might unconsciously influence their classroom behavior and judgments" (p. 46), so it seems to be important to include

---

[5] Implicit attitudes were measured with a reaction time-based measure and explicit attitudes with a questionnaire (see Sect. 6.3 Instruments).

[6] About which, however, we learned after our study from one of the reviewers of this paper.

information about influential factors on judgments in the intervention. Therefore, the text contained information about how stereotype (in-)consistent information influences perception and judgment formation and how, consequently, differing stereotypes and attitudes associated with various student backgrounds may result in differing expectations and judgments for different students. In this context, the focus was not only negative stereotypes and judgment biases concerning minority groups, but also on the risk of positive discrimination of majority students due to overestimation stemming from particularly positive stereotypes and attitudes (we assume rather dissimilar to the information focus in the training of Devine et al., 2012). These consequences of own stereotypes and attitudes for judgments may increase PTS' accountability, subsequently increasing their attention and, thus, could lead to more accurate judgments (Pit-ten Cate, Krolak-Schwerdt, & Glock, 2014, 2016a).

4.4 Strategies for practical implementation

The last part of the intervention was more practice orientated and offered strategies on how PTS could reflect on their own stereotypes, attitudes, emotional reactions, and expectations and on how to control for stereotype and attitude influences when they judge students with differing backgrounds. For example, PTS should ask themselves whether their expectations of students are related to students' group membership. Because imagined contact (Miles & Crisp, 2014) and stereotype-inconsistent information (e.g., Baadte, 2020; Devine et al., 2012; Gawronski et al., 2008) are verified factors to reduce prejudice and negative stereotypes, in this part of the intervention PTS were also invited to think about people (they know personally or from the media) who contrast with common stereotypes and thus may also contradict their own stereotyped expectations. Furthermore, PTS were encouraged to take enough time and apply objective criteria when judging students in order to control for behavioral automatisms. Finally, participants were asked to write an individual memo sentence that would remind them to act and judge without stereotypes and prejudices, i.e., help them to process student information in a less automatic and more elaborated way.

5 Hypotheses

To begin with, combining models and methods to change stereotypes and attitudes with those that initiate cognitive control of judgments for stereotype and attitude biases requires hypotheses concerning stereotype and attitude changes, judgment changes, and associations between these changes. However, hypotheses in this paper are curtailed to (a) associations between judgments, stereotypes, and attitudes in our sample and context and to (b) effects of the intervention on judgment changes. The first serves as a control for the initial conditions concerning presuppositions of the intervention.

1. Judgments, stereotypes, and attitudes (implicit and explicit) are associated in the pretest with

    a. positive correlations between judgments, stereotypes, and explicit attitudes for target students,
    b. negative correlations between judgments for minority students and implicit attitudes (towards majority backgrounds).

2. PTS in the experimental intervention judge students with a Turkish immigration background more positively in the post-test and follow-up than in the pretest when compared to PTS in the control condition with regard to

    a. achievement expectations (future grades),
    b. eligibility for highest school track,
    c. academic capabilities, and
    d. willingness to achieve.

3. PTS in the experimental intervention judge students with low social status background more positively in the post-test and follow-up than in the pretest when compared to PTS in the control condition with regard to

    a. future grades (achievement expectations),
    b. eligibility for highest school track,
    c. academic capabilities, and
    d. willingness to achieve.

Though theories, models, and methods to induce cognitive control of judgment biases do not provide clear assumptions about long-term effects, we attempted to stimulate more stable (longer lasting) stereotype and attitude changes (see ELM, Petty & Cacioppo, 1986), possibly influencing judgments and judgment changes, which also might show up in the follow-up. Moreover, the intervention addressed both stereotype and attitude changes as well as cognitive control of stereotype and attitude influences on judgments. Therefore, judgment changes may either result from stereotype and attitude changes or from cognitive control of their influences.

6 Intervention study to reduce judgment biases and induce changes in stereotypes and attitudes

6.1 Design

We conducted an experimental online-study in a pre-post-follow-up-design with $N = 215$ PTS enrolled in primary school studies, who selected this study in the introductory psychology module to earn credit for research participation as a test person. Participants took the pretest six weeks after the start of the semester (12/2020), the intervention and the posttest eight weeks later, shortly before semester ended

(02/2021), and the follow-up six weeks later during the semester break after the exam phase (03/2021). Each time, the online platform was open for participation for one week. PTS were allocated at random to the experimental vs. the control condition. Participants' judgments for target students with different family backgrounds (based on three case vignettes), their explicit attitudes towards students with a Turkish immigration background, no immigration background, high social status background, and low social status background as well as their stereotypes concerning these student groups were collected at all measurement points. To control for possible influences of implicit attitudes regarded as rather stable and change resistant (e.g., Baron, 2015), PTS' implicit attitudes towards students with a Turkish background and towards persons of low social status were assessed once at the pretest to serve as a control variable in the analyses. At the end of the pretest, participants were asked for demographical information.

6.2 Sample

The original sample size reached $N = 221$ participants at the beginning of the study (pretest) and fell to $N = 215$ by the end (follow-up) with less than 3% dropout.[7] Analyses were carried out with participants who took part at all three measurement points ($N = 215$), most of whom were in their first semester (98.6%) and, consequently, of a young age ($M = 19.9$ years, $SD = 2.8$). Because of the vignettes (midyear 4th grade) and the judgment requests (rating students' eligibility for the highest school track in secondary education), only primary school PTS were allowed to participate with a consequently high proportion of female subjects (90.2%). To control for influences of PTS' family backgrounds, subjects were asked about their (or their families) social class and their own and their parents' country of birth. Most PTS allocated themselves (their families) as being part of the middle class (53.4%) and higher middle class (40.7%), with only 1.8% declaring being part of the upper class and 3.2% of the lower class. Concerning own immigration backgrounds, 20.8% indicated at least one non-German origin in the family. Of special importance was the distribution of a Turkish immigration background in the experimental and the control group, because vignettes included a student with a Turkish name and stereotypes and attitudes were assessed for future students with a Turkish immigration background, which would constitute an in-group for PTS with the same origin. The proportion in sum was low (3.2%). The distribution across groups, however, was uneven, $Chi^2(2) = 9.327$, $p < .01$, with all PTS of Turkish origin in the experimental

---

[7] An a-priori power analysis (G*Power 3.1.9.7) for mixed two-factorial ANOVAs (one per hypothesis 2a–3d) with three repeated measures (within-factor) of two groups (between-factor), with an assumed effect size of $\eta^2 = .10$, a significance level of $p = .05$, an intended power of $(1-\beta) = .95$, and assumed correlations among repeated measures of $r = .50$ yielded a sample size of 260. We did not fully reach this sample size because PTS were recruited in the frame of a psychology module offering competing studies to participate in but decided not to recruit PTS in other contexts. A post-hoc sensitivity analysis with the same parameters for the realized sample of $N = 215$ yielded the following critical values: $F_{2,426} = 3.017$, $\eta^2 = .11$.

group. No other demographical variable showed significant differences across conditions.

6.3 Instruments

Judgments were assessed based on case vignettes. At all measurement points PTS read midyear 4th-grade school reports for three fictive students containing numeric grades in academic subjects and a verbal description of their academic strengths and weaknesses, learning engagement, work-organization, and social behavior. To raise authenticity, numeric grades for academic subjects (ranging from 1 = *very good* to 3 = *satisfactory*) varied across vignettes with an average of 2.11 points for all academic subjects and an average of 2.33 points for main academic subjects (general studies, German, and mathematics) per report, with the latter constituting the threshold for high track secondary school recommendations. Likewise, the wording of the verbal reports varied across vignettes. However, statements were systematically balanced with rotating positive, neutral, and negative phrases about students' behavior to depict a more positive but also inconsistent evaluation of the students. The latter allowed for PTS' subjective impression formation and interpretation of the information (adapted from Tobisch & Dresel, 2017).[8] Student names in the school reports indicated family backgrounds. Because educational disparities associated with students' origin differ depending on students' gender, that is, they are stronger for male students, and testing differential effects of the intervention on 2*3 combinations of background and gender would require a much larger sample, we presented reports for male students only. To avoid memory effects across the three measurement points and control for effects of the order of presentation as well as the names as such, we rotated the order of three names for each student background (non-immigration high status: Julius, Maximilian, Justus; non-immigration low status: Justin, Kevin, Mike; Turkish immigration with presumably low status: Murat, Ayhan, Mohammed) across measurement points, so that no participant received the same report and order of backgrounds twice. Names were selected from previous studies attesting the activation of intended stereotypes (e.g., Tobisch & Dresel, 2017; Wenz & Hoenig, 2020), and our manipulation check confirmed the categorization by names.

After each vignette, PTS had to judge the student on four central aspects considered important for educational success: (1) forecast the student's next exam grade in the three main academic subjects (ranging from 1 = *very good* to 6 = *failed*), (2) decide on the student's eligibility to enter the highest secondary school track (5-point Likert-scale ranging from 1 = *not at all eligible* to 5 = *fully eligible*), (3) estimate the student's academic capabilities with four bipolar items (5-point Likert scaled, e.g., "To learn something new is for him" … 1 = *difficult* vs. 5 = *easy;* adapted from Dickhäuser et al., 2002), (4) predict the student's future willingness to achieve on five items (e.g., "He will try to do everything as good as possible",

---

[8] Further information, in particular a detailed description of the case vignettes including the construction of the verbal descriptions can be obtained from Tobisch (2019).

5-point Likert scaled ranging from 1 = *not all true* to 5 = *fully true*; adapted from Ramm et al., 2006). Cronbach's Alpha for the scales assessing judgments of academic capabilities and willingness to achieve ranged from 0.72 up to 0.83 across student backgrounds and measurement points.

The first judgment concerns future educational success as such (representing an expectation), the second judgment concerns chances for educational success and the vignettes provided rather direct indicators, i.e., information. PTS can draw on students' grades to give a prognosis of future grades and to estimate students' eligibility for highest school track in secondary education, given the threshold of 2.33 in the main subjects, which, however, PTS in their first semester may not know (or remember). The second two judgments concern two important *prerequisites* of achievement and educational success, none of which is conceived as sufficient in itself (Walberg, 1982) and the vignettes provided rather indirect indicators. PTS had to infer estimations from intentionally inconsistent information about academic strengths and weaknesses (indicators of cognitive abilities), learning engagement (indicators of motivation) and work-organization (indicators of both). Consequently, these two judgments may be more prone to subjective interpretation and influences of stereotypes and attitudes, but also more amenable to changes either through stereotype and attitude changes and/or through cognitive control of stereotype and attitude influences on judgments. Furthermore, implicit as well as explicit judgments of these prerequisites presumably occur more often—and are required—in ongoing classroom interaction under conditions allowing less time for processing and integrating individuated information about students, e.g., when teachers (feel obliged to) give feedback or adapt instruction, and research shows effects of students' perceptions of how their teachers evaluate them on their self-concept (e.g., Dickhäuser & Stiensmeier-Pelster, 2003). Thus, assessing if such judgments were also addressed with the intervention seems promising for the reduction of disparities in the long run.

Explicit stereotypes were assessed as comparisons of students with Turkish vs. no immigration background and students with low vs. high social status on two parallelized semantic differentials (Osgood et al., 1971) presenting 15 pairs of attributes (8 negative vs. positive, 7 vice versa), e.g., "In comparison to students without migration background students with Turkish background are … < *not at all ambitious* vs. *very ambitious* >, … *not all aggressive* vs. *very aggressive*", which PTS had to rate on a six-point scale. Answers to positive–negative-polarized items were recoded, with higher values indicating positive stereotypes. Attributes were taken from an unpublished survey among PTS of the same university prior to the study to identify student characteristics generally considered learning-relevant and represent cognitive abilities, learning behavior, motivation, and social behavior ($\alpha = .89$–.94).

Explicit attitudes towards students with a Turkish and without an immigration background and towards those of low and of high social status were assessed with 4*10 items parallelized across backgrounds on bipolar six-point scales (adapted from Lehmann-Grube et al., 2022). Items represented general valence, utility, and costs of having these students in class, e.g., "For my teaching, I consider the origin of students from families with … a Turkish migration background as … < *unpleasant*

vs. *pleasant* > … < *hindering* vs. *fostering* students' learning > … < *not stressful* vs. *stressful* for the teacher >" (α = .80–.92).

Implicit attitudes towards target groups (German vs. Turkish, high vs. low social status) were measured once in advance to the intervention with two implicit association tests (IAT; e.g., Greenwald et al., 1998) assessing reaction times to combinations of target categories with negative or positive characteristics regarded as relevant for learning, achievement, and classroom interaction (e.g., diligent, lackadaisical; team-oriented, disrespectful).[9] Again, names indicated ethnicity (e.g., German: Andreas; Turkish: Mehmet), but vocations indicated social status (e.g., high status: lawyer; low status: waiter). Participants worked on blocks of stereotype consistent and stereotype inconsistent combinations in randomly assigned orders to avoid sequence effects. Based on the assumption that (strong) associations of target categories with characteristics constitute stereotypes and require less cognitive effort and less reaction time, reaction times to stereotype consistent combinations are subtracted from reaction times to stereotype inconsistent combinations. Thus, higher scores signify stronger stereotypes and implicit preferences, and in the case of negative vs. positive stereotypes, as in this study, indicate higher differences between implicit attitudes towards the target categories (Greenwald et al., 2002).

Demographical information was collected once at the end of the pretest containing questions about gender (male, female, other), age, social class (self-allocation), and countries of birth (self and parents).

7 Results

Descriptive statistics of pretest data showed relatively positive judgments for all target students (see Table 1). PTS' achievement expectations, i.e., expected grades, were equal across target students' backgrounds, $F(2,428) = 0.027$, $p = .97$, $\eta^2 = .00$, but judgments differed by students' backgrounds with regard to eligibility for high track schooling, $F(2,428) = 4.069$, $p < .05$, $\eta^2 = .02$, academic capabilities, $F(2,428) = 7.846$, $p < .001$, $\eta^2 = .04$, and willingness to achieve, $F(2,428) = 12.394$, $p < .001$, $\eta^2 = .06$. In particular, PTS judged target students with a Turkish immigration background more positively and low-status target students more negatively than target students without an immigration background and high social status. Correlations of judgments per student background were highest and consistent for target students with a Turkish immigration background. For target students without an immigration and low social status background, achievement expectations correlated lower with the other judgments, and for target students with high social status these correlations were lowest (see Table 1). Comparisons of the conditions showed in the pretest only differences with regard to judgments for target students of low social status. PTS in the experimental condition judged these students' eligibility for the highest school track, $(1,213) = 4.760$, $p < .05$, $\eta^2 = .02$, academic capabilities,

---

[9] Software Inquisit 5 Web (Millisecond Software, 2016).

Table 1 Means, standard deviations, and bivariate correlations of PTS' judgments for target students (pretest)

Judgments by backgrounds	M (SD)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Target students with Turkish immigration background (associated with low social status)												
(1) Achievement expectation [a]	2.26 (0.30)	–										
(2) El. for highest school track [b]	3.54 (0.91)	**-.42****	–									
(3) Academic capabilities [c]	4.40 (0.72)	**-.47****	**.52****	–								
(4) Willingness to achieve [c]	3.75 (0.58)	**-.41****	**.47****	**.60****	–							
Target students with low social status (and no immigration background)												
(5) Achievement expectation [a]	2.27 (0.31)	**.32****	-.15*	-.16*	-.14*	–						
(6) El. for highest school track [b]	3.34 (0.86)	-.07	**.31****	.07	.03	-.28**	–					
(7) Academic capabilities [c]	4.16 (0.74)	-.09	.14*	**.32****	.18**	-.34**	.47**	–				
(8) Willingness to achieve [c]	3.50 (0.56)	-.08	.12	.22**	**.20****	-.26**	.35**	.59**	–			
Target students with high social status (and no immigration background)												
(9) Achievement expectation [a]	2.27 (0.27)	**.26****	-.04	-.07	.04	.25**	-.11	-.05	-.05	–		
(10) El. for highest school track [b]	3.40 (0.85)	-.00	**.24****	-.04	.00	-.10	.34**	.08	.03	-.27**	–	
(11) Academic capabilities [c]	4.30 (0.73)	-.04	.01	**.12**	.10	-.07	.20**	.26**	.16*	-.26**	.57**	–
(12) Willingness to achieve [c]	3.60 (0.64)	.01	-.03	-.02	**.11**	-.10	.04	.12	.22**	-.14*	.40**	.58**

$N = 215$. Bold-faced values are of theoretical interest (expected correlations)

[a] Lower values denote higher achievement expectations (grades 1 = *very good*; 6 = *failed*). [b] Eligibility for highest school track (1 = *not at all*; 5 = *fully*), [c] (1 = *low*; 5 = *high*)

**p < .01. * *p* < .05

Table 2 *Means, standard deviations, and bivariate correlations of stereotypes about and attitudes towards target students (pretest)*

	M (SD)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stereotypes								
(1) With Turkish vs. no immigration background	3.97 (0.61)	–						
(2) With low vs. high social status background	4.03 (0.75)	**.59****	–					
Implicit attitudes								
(3) With Turkish vs. no immigration background	0.48 (0.34)	−.07	−.08	–				
(4) With low vs. high social status background	0.69 (0.38)	−.04	−.02	**.31****	–			
Explicit attitudes								
(5) With Turkish immigration background	4.39 (0.75)	**.34****	**.23****	**−.24****	−.11	–		
(6) Without immigration background	4.83 (0.60)	.02	−.07	.08	.04	**.23****	–	
(7) With low social status background	4.04 (0.85)	**.16***	**.32****	**−.14***	−.05	**.54****	**.28****	–
(8) With high social status background	4.44 (0.69)	−.05	−.09	−.10	−.03	**.28****	**.49****	**.35****

N = 215. Significant correlations are in bold

**p < .01. * *p* < .05

$F(1,213) = 6.171$, $p < .05$, $\eta^2 = .03$, and willingness to achieve, $F(1,213) = 7.391$, $p < .01$, $\eta^2 = .03$, as lower than PTS in the control condition.[10]

Table 2 shows descriptive statistics and correlations of stereotypes and attitudes in the pretest. In their stereotypes, PTS tended to rate characteristics (stereotypes) of students with a Turkish immigration background in comparison to students without an immigration background as well as of students with low social status in comparison to students with high social status slightly more positively. Both stereotype measures correlated highly. Stereotypes also correlated positively with explicit attitudes towards students with a Turkish immigration background and with attitudes towards students with low social status. In their implicit attitudes, PTS showed strong preferences for persons without an immigration background compared to persons with a Turkish immigration background and even stronger preferences for persons with high social status compared to persons with low social status. These implicit preferences correlated positively, meaning that PTS showing higher preferences for students without an immigration background also preferred persons with high social status more strongly and vice versa. Explicit attitudes were positive

---

[10] Differences between conditions in the pretest remained in control analyses excluding PTS with (Turkish) immigration background, even when controlling for implicit attitudes. Thus, the unequal distribution of participants' Turkish immigration background across conditions does not explain the different patterns.

(above scale means) for all groups but showed significant differences between student backgrounds, $F(3,642)=66.202$, $p<.001$, $\eta^2=.24$. On average, PTS valued having target students without an immigration background in their future classrooms most positively and students with low social status least positively. Explicit attitudes towards having students with a Turkish immigration background and towards having students with low social status in their classrooms correlated negatively with implicit attitudes concerning (non-)immigration background. Thus, strong implicit preferences for students without an immigration background are associated with less positive explicit attitudes towards having students with a Turkish immigration background and less positive explicit attitudes towards having students with a low social status background in classrooms. None of the explicit attitudes correlated with implicit attitudes concerning social status and none of the stereotypes correlated with any of the implicit attitudes.

Comparisons of pretest measures across conditions yielded no differences in means and variances of stereotypes, explicit attitudes, and implicit attitudes except for implicit attitudes concerning immigration backgrounds. PTS in the placebo group (control condition) exhibited stronger preferences for persons without an immigration background when compared to PTS in the experimental condition, $F(1,213)=6.789$, $p=.01$, $\eta^2=.03$. Therefore, we controlled for implicit attitudes in subsequent analyses.

7.1 Associations between judgments, stereotypes, and attitudes in the pretest

PTS' stereotypes of students with a Turkish vs. no immigration background correlated positively with judgments of these students' eligibility for the highest school track and willingness to achieve, with judgments of low-status students' eligibility for the highest school track, academic capabilities, and willingness to achieve, but not with any judgment for students of high social status (see Table 3). Similarly, PTS' stereotypes of students with low vs. high social status correlated positively with judgments of academic capabilities and willingness to achieve for students with low social status and students with a Turkish immigration background, though not with their eligibility for the highest school track, but also with judgments of high-status students' academic capabilities.

Implicit attitudes towards students without vs. with a Turkish immigration background correlated with judgments of academic capabilities of students with a Turkish immigration background and of students with a low social status background. Explicit attitudes towards students with a Turkish immigration background correlated with only one judgment for these students (eligibility for highest school track), one judgment for students with a low social status (willingness to achieve), but with two judgments for students with a high social status (academic capabilities and willingness to achieve). Explicit attitudes towards students with a low social status correlated with no judgment for these students, only with one judgment for students with a Turkish immigration background (academic capabilities), but again with two judgments for students with a high social status (academic capabilities and

Table 3 *Bivariate correlations of judgments with stereotypes and attitudes (pretest)*

Judgments by backgrounds	Stereotypes		Implicit attitudes		Explicit attitudes			
	TIB vs. NIB	LSS vs. HSS	TIB vs. NIB	LSS vs. HSS	TIB	NIB	LSS	HSS
Target students with Turkish immigration background (associated with low social status)								
Achievement expectations [a]	−.07	−.09	−.01	.02	−.05	−.04	−.08	.01
El. for highest school track [b]	**.17***	.07	−.08	.05	**.14***	.05	.06	.01
Academic capabilities	.10	**.14***	**.14***	.05	.11	.07	**.17***	.08
Willingness to achieve	**.18****	**.16***	.06	−.03	.12	−.01	.09	.04
Target students with low social status background (without immigration background)								
Achievement expectations [a]	−.13	−.11	−.02	−.04	−.12	.04	−.07	−.02
El. for highest school track [b]	**.26****	.11	−.04	−.00	.11	−.09	.04	−.09
Academic capabilities	**.14***	**.19****	**.19****	.10	.07	−.07	.09	.06
Willingness to achieve	**.19****	**.23****	.04	−.01	**.17***	−.11	.09	.05
Target students with high social status background (without immigration background)								
Achievement expectations [a]	−.02	.07	−.04	.10	.00	−.03	.03	−.04
El. for highest school track [b]	.08	.08	−.03	.01	.11	−.03	.10	.09
Academic capabilities	.08	**.21****	.02	−.01	**.17***	.07	**.24****	**.16***
Willingness to achieve	.06	.13	.00	−.06	**.18****	.01	**.20****	.12

N=215

TIB = Turkish immigration background; NIB = No immigration background; LSS = Low social status; HSS = High social status. Significant correlations are in bold

[a]Lower values denote higher achievement expectations. [b]Eligibility for highest school track

** *p*<.01. * *p*<.05

willingness to achieve). Neither implicit attitudes towards high vs. low social status nor explicit attitudes towards having students without an immigration background in classrooms correlated with any judgment variable. Furthermore, none of the achievement expectations (grades) correlated with any of the stereotype and attitude variables, which was also the case for judgments of high-status students' eligibility for the highest school track.

7.2 Effects of the intervention on judgments, attitudes, and stereotypes

Mixed two-factorial analyses of variance (ANOVA) yielded several significant differences between the intervention and the placebo group with regard to changes in judgments, stereotypes, and attitudes (see Table 4). When controlling for effects of implicit attitudes (single measures; mixed two-factorial ANCOVA), most differences were less strong, but all had the same pattern. Means and standard deviations for direct comparisons of groups and measurement points are reported in Tables 5 and 6.

PTS' *achievement expectations* changed differently only for target students with a Turkish immigration background when comparing the intervention and the control group (see Tables 4 and 5).[11] In the post-test, participants in the intervention group expected significantly better grades from these students in future exams when compared to the pretest, $t(111) = 3.240$, $p < .001$, [0.04, 0.17], $d = 0.30$, and when compared to the control group, $t(213) = -3.103$, $p < .05$, [−0.18, −0.04], $d = 0.42$. The within-subject effect in the intervention condition decreased at the follow-up, but the difference to the pretest was still significant, $t(111) = 2.026$, $p < .025$, [0.00, 0.11], $d = 0.19$.

PTS' *ratings of eligibility for the highest school track* showed no significant differences of change between the intervention and the control group for either student background (see Tables 4 and 5). However, participants in the intervention group rated the eligibility of target students with a low social status background significantly lower in the pretest when compared to the post-test, $t(111) = -3.507$, $p = .001$, [−0.50, −0.14], $d = 0.33$, and the follow-up, $t(111) = -3.204$, $p < .001$, [−0.46, −0.11], $d = 0.30$. These changes in the intervention group may have been masked in the mixed ANOVA by their descriptively (Bonferroni-Holm corrected non-significant) lower estimations in the pretest when compared to the control group, $t(213) = 2.182$, $p = .030$, [0.48, 0.02], $d = 0.30$.

PTS' ratings of *academic capabilities* changed differently in the intervention and the control group only for target students with a low social background (see Tables 4

---

[11] The following simple effects tests were conducted with one-sided significance tests for differences between groups at post-test and follow-up as well as for within-group differences between pre-test and post-test and pre-test and follow-up (according to hypotheses) all with Bonferroni-Holm correction of significance levels (Holm, 1979) per hypothesis (2a-3d), i.e., judgment dimension by target students, for four comparisons and 95% CI. Additional t-tests to control for between group-differences at the pretest were done with two-sided significance tests (no hypotheses) and Bonferroni-Holm correction for five comparisons.

Table 4 *Effects of the intervention on judgments, stereotypes, and attitudes concerning target students—results of mixed two-factorial ANOVAs without and with covariates*

| | Implicit attitudes | | | |
| | Not controlled | | Controlled | |
	F	η^2	F	η^2
Judgments [a]				
Turkish immigration background				
Achievement expectations	**5.603****	**.03**	**5.472****	**.03**
Eligibility for highest school track	1.192	.01	1.225	.01
Academic capabilities	1.156	.01	0.816	.00
Willingness to achieve	**3.053***	**.01**	2.556[+]	.01
Low social status background				
Achievement expectations	1.251	.01	0.9796	.01
Eligibility for highest school track	2.209	.01	2.134	.01
Academic capabilities	**3.302***	**.04**	2.566[+]	.01
Willingness to achieve	**4.018***	**.02**	**3.450***	**.02**
High social status background				
Achievement expectations	0.919	.00	0.864	.00
Eligibility for highest school track	0.298	.00	0.426	.00
Academic capabilities	1.159	.01	1.409	.01
Willingness to achieve	2.439[+]	.01	2.475[+]	.01
Stereotypes [b]				
Turkish vs. no immigration background	**7.409*****	**.03**	**5.955****	**.03**
Low vs. high social status background	**6.760*****	**.03**	**6.233****	**.03**
Explicit attitudes [a]				
Turkish immigration background	**5.443****	**.03**	**5.749****	**.03**
No immigration background	1.152	.01	0.697	.00
Low social status background	**13.311*****	**.06**	**13.631*****	**.06**
High social status background	2.080	.01	1.980	.01

$N = 215$. Significant effects are in bold

[a]*df* for judgments and attitudes: Not controlled: $df = 2$; $df_{error} = 426$; Controlled: $df = 2$; $df_{error} = 422$; [b]*df* for stereotypes: Not controlled: $df = 2$; $df_{error} = 418$; Controlled: $df = 2$; $df_{error} = 414$

*** $p < .001$. ** $p < .01$. * $p < .05$. [+] $p < .10$. (Two-sided testing)

and 5). Participants in the intervention group rated these students' capabilities significantly lower in the pretest when compared to the post-test, $t(111) = -3.066$, $p < .016$, $[-0.42, -0.09]$, $d = 0.29$, and the follow-up, $t(111) = -4.081$, $p < .001$, $[-0.48, -0.17]$, $d = 0.39$, meaning that ratings in the intervention group increased. However, these within-subject increases did not result in a significant between-subject difference of intervention and control condition at post-test and follow-up.

PTS' ratings of students' *willingness to achieve* showed the same pattern again for target students with a low social background. Ratings across measurements changed differently in the groups (see Tables 4 and 5). Groups differed in the pretest,

Table 5 *Means and standard deviations (in brackets) of judgments and expectations in the experimental group (intervention condition) and the control group (placebo condition)*

Judgments by backgrounds		Pretest	Posttest	Follow up
Target students with Turkish immigration background (associated with low social status)				
Achievement expectations [a]	EG [b]	2.28 (0.29)	2.17 (0.28)	2.22 (0.21)
	CG [c]	2.25 (0.31)	2.28 (0.23)	2.23 (0.24)
Eligibility for highest school track	EG [b]	3.56 (0.81)	3.69 (0.79)	3.64 (0.78)
	CG [c]	3.51 (1.00)	3.50 (0.91)	3.66 (0.84)
Academic capabilities	EG [b]	4.38 (0.63)	4.43 (0.64)	4.40 (0.64)
	CG [c]	4.43 (0.81)	4.33 (0.64)	4.34 (0.57)
Willingness to achieve	EG [b]	3.76 (0.57)	3.82 (0.44)	3.74 (0.53)
	CG [c]	3.75 (0.59)	3.58 (0.58)	3.60 (0.57)
Target students with low social status background (without immigration background)				
Achievement expectations [a]	EG [b]	2.28 (0.34)	2.23 (0.28)	2.26 (0.20)
	CG [c]	2.25 (0.28)	2.27 (0.23)	2.27 (0.25)
Eligibility for highest school track	EG [b]	3.22 (0.80)	3.54 (0.83)	3.51 (0.74)
	CG [c]	3.48 (0.90)	3.50 (0.85)	3.62 (0.82)
Academic capabilities	EG [b]	4.04 (0.71)	4.30 (0.72)	4.37 (0.65)
	CG [c]	4.29 (0.75)	4.40 (0.67)	4.31 (0.71)
Willingness to achieve	EG [b]	3.40 (0.52)	3.62 (0.54)	3.63 (0.45)
	CG [c]	3.60 (0.58)	3.63 (0.61)	3.57 (0.59)
Target students with high social status background (without immigration background)				
Achievement expectations [a]	EG [b]	2.27 (0.28)	2.22 (0.25)	2.26 (0.29)
	CG [c]	2.26 (0.25)	2.26 (0.26)	2.29 (0.29)
Eligibility for highest school track	EG [b]	3.37 (0.86)	3.56 (0.78)	3.46 (0.90)
	CG [c]	3.44 (0.85)	3.53 (0.79)	3.45 (0.83)
Academic capabilities	EG [b]	4.23 (0.71)	4.37 (0.60)	4.27 (0.72)
	CG [c]	4.37 (0.76)	4.46 (0.61)	4.24 (0.64)
Willingness to achieve	EG [b]	3.54 (0.63)	3.62 (0.52)	3.56 (0.55)
	CG [c]	3.66 (0.65)	3.61 (0.59)	3.44 (0.59)

$N = 215$

[a] Lower values denote higher achievement expectations (grades 1 = *very good*; 6 = *failed*). On all other scales lower values denote lower judgments (1 = *not all/low*; 5 = *fully/high*). [b] Experimental group (intervention condition, $n = 112$). [c] Control group (placebo condition, $n = 103$)

$t(213) = -2.719$, $p < .01$, $[-0.35, -0.06]$, $d = 0.37$, and participants in the intervention group rated these students higher at post-test, $t(111) = -3.718$, $p < .01$, $[-0.34, -0.10]$, $d = 0.35$, and follow-up, $t(111) = -3.789$, $p < .01$, $[-0.34, -0.11]$, $d = 0.36$, when compared to the pretest. PTS' ratings of willingness to achieve also changed differently depending on conditions concerning target students with Turkish immigration background (see Tables 4 and 5). Groups did not differ in the pretest but in the post-test, $t(213) = -3.449$, $p = .01$, $[0.10, 0.38]$, $d = 0.47$. However, within-subject effects in the intervention group were not significant and descriptively ratings were even lower in the follow-up than in the pretest.

Table 6 *Means and standard deviations (in brackets) of stereotypes and explicit attitudes in the experimental group (intervention condition) and the control group (placebo condition)*

		Pretest	Posttest	Follow up
Stereotypes [a]				
Turkish vs. no immigration background	EG [c]	4.03 (0.65)	4.27 (0.69)	4.17 (0.65)
	CG [d]	3.93 (0.57)	3.91 (0.65)	3.80 (0.55)
Low vs. high social status background	EG [c]	4.05 (0.76)	4.20 (0.73)	4.13 (0.72)
	CG [d]	4.00 (0.75)	3.94 (0.70)	3.77 (0.59)
Explicit attitudes [b]				
Turkish immigration background	EG [c]	4.45 (0.71)	4.56 (0.63)	4.44 (0.69)
	CG [d]	4.32 (0.79)	4.23 (0.68)	4.04 (0.78)
No immigration background	EG [c]	4.81 (0.61)	4.68(0.56)	4.63 (0.59)
	CG [d]	4.84 (0.59)	4.67 (0.56)	4.53 (0.62)
Low social status background	EG [c]	4.04 (0.77)	4.27 (0.75)	4.20 (0.76)
	CG [d]	4.04 (0.93)	3.91 (0.90)	3.71 (0.91)
High social status background	EG [c]	4.49 (0.73)	4.45 (0.65)	4.45 (0.65)
	CG [d]	4.39 (0.66)	4.42 (0.67)	4.25 (0.68)

N=215

[a] Comparison of target students (1=*most negative rating*; 6=*most positive rating*). [b] Single target rating (1=*negative attitude*; 6=*positive attitude*). [c] Experimental group (intervention condition, *n*=112). [d] Control group (placebo condition, *n*=103)

PTS' *stereotype changes* across measurements showed considerable differences between conditions for both student background comparisons, i.e., Turkish vs. no immigration and low vs. high social status background alike (see Tables 4 and 6), with the intended positive development in the intervention group. The same holds for *changes of explicit attitudes* towards students with a Turkish immigration background and students with a low social status background (see Tables 4 and 6).

8 Discussion

Disparities in educational success depending on students' immigration and/or social status backgrounds seem to be partially influenced by social categorization processes, i.e., teachers' judgment biases, stereotypes, and attitudes (e.g., Gentrup et al., 2018; Glock et al., 2015, 2016; Klapproth et al., 2012; Rubie-Davies et al., 2006). Against this backdrop, the aim of this experimental study was to develop and evaluate a short intervention integrating (and condensing) essential elements of theories to change PTS' stereotypes and attitudes with elements of theories to reduce judgment biases depending on students' backgrounds. To test for judgment biases, stereotypes, and attitudes at the outset as well as for effects of the intervention, all vignettes employed in this study were modeled to indicate equivalent levels of achievement, learning behavior, and motivation of target students whose backgrounds were manipulated by student names only. Analyses of pretest data revealed

biases of PTS' judgments of students' eligibility for the highest school track, academic capabilities, and willingness to achieve depending on students' backgrounds. In general, this result is in line with previous experimental research concerning judgment biases (e.g., Bonefeld & Dickhäuser, 2018; Bonefeld et al., 2020; Civitillo et al., 2022; Glock & Krolak-Schwerdt, 2013; Glock et al., 2013, 2016; Holder & Kessels, 2017; Tobisch & Dresel, 2017). However, in contrast to these previous experimental studies, PTS in our study judged (fictive) students with a Turkish immigration background the most positively. This positive bias is in line with positive secondary effects found in some field studies when controlling effects of immigration backgrounds on transitions to secondary school tracks for effects of social status, grading, and achievement in standardized tests (e.g., Gresch & Becker, 2010). It is also in line with results of some experimental studies, e.g., overestimation of below average students (Glock et al., 2015) or of (male) low test performers (Bonefeld et al., 2020) with a Turkish immigration background. These positive biases suggest that PTS may have been highly motivated not to discriminate against these target students, which was backed-up in our study by the consistently high correlations of the judgment dimensions concerning these students in the pretest. Though we did not assess social desirability, the overestimation may also indicate that (many) PTS in our study wanted to avoid the impression of discriminating against target students with a Turkish immigration background regardless of their factual impression formation. On the other hand, the lowest judgments for target students with a low social status background in the pretest reveal PTS' unawareness of judgment disparities (and discrimination) concerning social status backgrounds, with the significant difference between conditions indicating that this particularly held for PTS in the experimental condition at the outset. This result aligns with field studies yielding negative judgment biases for students of low social status (summary in Dumont et al., 2014) and, likewise, with results of experimental studies indicating effects of (negative) social categorization processes (e.g., Glock & Kleen, 2023; Glock et al., 2016; Tobisch & Dresel, 2017).

8.1 Hypotheses testing

Results on *associations between judgments, stereotypes, and attitudes* (H1) at the outset only partially supported presuppositions of the intervention. However, they confirm our considerations on why to include judgment dimensions which we assumed to allow participants a more subjective interpretation, namely the prerequisites of achievement (see Sects. 5 and 6.3). The more subjective interpretation of indicators for these dimensions may have allowed stronger influences of activated stereotypes and attitudes, and thus, judgments for achievement prerequisites showed more correlations with stereotypes and attitudes than judgments for achievement expectations for which the vignettes entailed direct indicators (i.e., grades). Pretest measures of stereotypes and attitudes correlated with estimations of target students' *eligibility for the highest school track* and with estimations of *prerequisites for achievement* (capabilities and motivation), particularly concerning students with the two minority backgrounds. They did not correlate with expected future grades of

any target student (H1a partially confirmed). With regard to future grades, a considerable number of PTS may have suppressed stereotypical expectations in their judgments, similar to teachers in the study by Glock and Krolak-Schwerdt (2014), who found stereotype activation in a memory task, but no application in the judgment task. Though overall unevenly distributed and low, the correlations of stereotypes and explicit attitudes with the other judgment dimensions at least partially correspond to results of previous experimental studies revealing influences of social categorization processes, stereotypes, and attitudes (e.g., Bonefeld & Dickhäuser, 2018; Bonefeld et al., 2020; Civitillo et al., 2022; Glock, 2016; Glock et al., 2013; Glock & Kleen, 2023; Glock & Krolak-Schwerdt, 2013; Holder & Kessels, 2017; Tobisch & Dresel, 2017). These correlations also correspond with results of some field studies assuming social categorization (and subsequent stereotyping) influences on judgment formation (e.g., Gentrup et al., 2018; Klapproth et al., 2012; Lorenz et al., 2016; Rubie-Davies et al., 2006). In particular, the five correlations regarding target students of low social status confirm previous results (Glock & Kleen, 2023; Glock et al., 2016; Stahl, 2007; Tobisch & Dresel, 2017).

The two significant correlations of implicit attitudes towards students without vs. with a Turkish immigration background with judgments of academic capabilities signify that PTS with stronger preferences for persons *without an immigration* background ascribed students *with a Turkish immigration* background and students *with low social status* higher academic capabilities than PTS with lower preferences for persons without an immigration background (H1b disconfirmed). Keeping in mind that target students of Turkish descent were judged most positively on all dimensions in the pretest, the first correlation again suggests that PTS either suppressed an activated stereotype due to (an honest) motivation not to discriminate (even before the intervention) or answered in a way they conceived as socially desirable to avoid the impression they would discriminate. However, the positive correlation of implicit attitudes concerning (no) immigration background with estimations of low-status students' academic capabilities and the missing correlations of implicit attitudes towards social status with any judgment measure, in a way, contrast the results reported from Glock et al. (2016), who found negative influences of high implicit preferences for high social status on judgments for students with low social status. Unexpectedly and also contrasting the results of Glock et al., explicit attitudes towards students with a Turkish immigration background and towards students with a low social status background correlated positively with PTS' judgments of high-status students' academic capabilities and willingness to achieve in the pretest. At least some PTS in our study may have answered items on explicit attitudes towards students with a Turkish immigration background and towards students with a low social status background in the pretest with a generally positive bias, again suggesting a social desirability motivation or a motivation not to discriminate (or both), but may not have controlled their judgments for high status students for social categorization processes concerning their in-group, the latter of which corresponds to the positive discrimination of high status students reported by Tobisch and Dresel (2017).

Mixed two-factorial analyses of variance with two between-subject and three within-subject (repeated) measures of *judgment changes by conditions* (H2, H3)

yielded some of the expected differences between conditions. The different changes showed specific patterns depending on target students' backgrounds, i.e., Turkish immigration background (associated with low social status) and social status backgrounds (without immigration background), and on judgment dimensions.

Concerning target students with a *Turkish immigration background*, PTS' achievement expectations increased markedly from pre- to post-test in the intervention group, but not in the control condition (H2a confirmed). Also, PTS' judgments of these students' willingness to achieve increased in the intervention group, but not in the control condition (H2d confirmed). However, neither recommendations for the highest secondary school track showed within- or between-subject effects (H2b rejected), nor did the ratings of these students' academic capabilities (H2c rejected). Keeping in mind that achievement expectations in the pretest were already more positive than average scores in the school reports across all student backgrounds, confirmation of hypothesis 2a indicates that the intervention induced an even stronger positive immigration background bias concerning achievement expectations. Together with more positive judgments of these target students' willingness to achieve at post-test measures, these intervention effects stand in contrast to the missing effects of Baadte's (2020) counter-stereotype training on the grading of essays of students with an Arabic immigration background. However, Baadtes' and our dependent variables (i.e., achievement ratings) involved different subject matters and abstract levels (grading of concrete essays vs. abstract school reports). Furthermore, the counter-stereotype training addressed stereotypes only but gave no information about stereotype or attitude influences on judgment distortions. Subsequently, Baadte's counter-stereotype training showed significant influences on stereotypes (less recall of stereotype-congruent information about this target student in the training condition when compared to the control condition) but not on judgment formation, with the latter possibly requiring additional cognitive control of stereotype influences.[12] Additionally, the targeted immigration backgrounds differed between Baadte's study (Arabian) and our study (Turkish). Keeping in mind the results of studies including more than one ethnic-cultural minority background (e.g., Civitillo et al., 2022; Gentrup et al., 2018; Klapproth et al., 2012; Rubie-Davies et al., 2006), which found that the extent and direction of biases depended on specific origins of students and associated stereotypes and prejudices, the differing backgrounds in Baadte's (2020) training and our intervention may also explain the different results. The increase in PTS' overestimation of students with a Turkish immigration background when compared to average scores in the school reports and to PTS' estimations of target students with a low or a high social status background surely does not accord with the results of training for teachers addressing general diagnostic skills which reduced biases of judgments for students with immigration backgrounds (Böhmer et al., 2017; Pit-ten Cate, Krolak-Schwerdt, & Glock, 2016;

---

[12] The effect of the gender specific counter-stereotype training on the grading of a male students' essay in Baadte's (2020) study may, at least partially, be explained by the methodological bias that the control condition (arithmetic task) and the dependent variable (essay) are both strongly related to gender stereotypes, but less, if at all, to social background stereotypes.

Pit-ten Cate, Krolak-Schwerdt, Hörstermann et al., 2016). While training general diagnostic skills seems to have increased *accountability for decisions*, our intervention may have only increased *sensitivity for stereotype and attitude influences* and thus may have rather stipulated a counter-stereotypical "over-reaction", again speaking either for PTS' high motivation not to discriminate against students with immigration backgrounds or a social desirability motivation to avoid the impression of discriminating against these students. Moreover, Pit-ten Cate, Krolak-Schwerdt, & Glock, 2016; Pit-ten Cate, Krolak-Schwerdt, Hörstermann et al., 2016) trained teachers, most of whom had experience in making tracking decisions for students with the prominent Portuguese immigration background in Luxembourg, which is similarly associated with low social status and lower achievement scores in general as for students with a Turkish immigration background in Germany, thus possibly confirming and consolidating teachers' stereotypes. But the accuracy scores reported by Pit-ten Cate, Krolak-Schwerdt, & Glock, (2016a) give no information on whether the participating teachers' over- or underestimated these students' eligibility for higher school tracks before the training. If the experienced teachers initially underestimated (negatively stereotyped) students with an immigration background (see Klapproth et al., 2012, for large scale analyses speaking for this possibility), both interventions induced a change to more positive judgments for students facing educational disparities, resulting in higher judgment accuracy of teachers underestimating these students initially, but higher judgment inaccuracy of PTS overestimating minority underprivileged students at the outset. These considerations, of course, do not question that Pit-ten Cate, Krolak-Schwerdt, & Glock, 2016; Pit-ten Cate, Krolak-Schwerdt, Hörstermann et al., 2016) training successfully increased teachers' judgment accuracy while our intervention increased a positive bias (i.e., decreased accuracy) for students of Turkish descent.

Concerning target students with a *low social status background*, neither achievement expectations nor recommendations for the highest school track differed in how they changed when comparing the intervention group and the control condition (H3a and H3b rejected). However, within-subject changes from pre- to post-test and to the follow-up in the intervention group concerning judgments of theses target students' eligibility for the highest school track indicate effects of the intervention in the hypothesized direction (H3b). Because at the outset PTS in the intervention group seem to have judged these target students' eligibility for the highest school track to be lower when compared to the judgments of participants in the control condition (and when compared to judgments for the other target students), we may interpret the within-subject effects as a decrease in judgment distortions in the intervention condition with regard to PTS' ratings of low-status target students. Equal patterns with significant judgment differences between conditions in the pretest (lower ratings in the intervention group) and significant differences of judgment changes across measurement points between conditions (increasingly better ratings from PTS in the intervention group) for these target students' academic capabilities and willingness to achieve (H3c and H3d confirmed) support this interpretation. Concerning target students with a low social status background, *sensitizing PTS* for judgment biases, i.e., for influences of stereotypes and attitudes on judgment formation, seems to reduce judgment biases in the same direction as interventions which

raise *teachers' accountability* for decisions and, thereby, increase judgment accuracy for students with immigration backgrounds (Pit-ten Cate, Krolak-Schwerdt, & Glock, 2016; Pit-ten Cate, Krolak-Schwerdt, Hörstermann et al., 2016). However, the effect of sensitizing for biases and stipulating cognitive control of stereotype and attitude influences on judgment formation was confined to ratings of target students' *prerequisites of achievement.* It did not impact *achievement expectations* (which were positively biased for all target students at the outset).

Differences between conditions in *changes of stereotypes and attitudes* across measurement points showed the intended effects of the intervention. Changes in stereotypes and attitudes align with the results of curricular intervention studies in teacher education contexts (e.g., Grottkau & Nickolai-May, 1989; Kumar & Hamer, 2013; Martin & Koppelman, 1991) and with results of experimental studies in college contexts (e.g., Kotzur et al., 2019; Soble et al., 2011). On the other hand, our results differ from the results of Devine et al. (2012), who found no effects of their partially similar (much more extensive) training on participants' explicit attitudes but only on implicit attitudes and concerns about discrimination. This difference may be due to the different context of Devine et al.'s study (USA) and the different ethnic minority background (Blacks) that they addressed.[13]

In summary, PTS initially showed positive biases for target students with a Turkish immigration background and negative biases for students with a low social status on three judgment variables, less positive attitudes towards both student groups, and small or no correlations of stereotypes and attitudes with judgments. The intervention increased the positive judgment bias for students with a Turkish immigration background on two judgment variables, decreased the negative bias for students with low social status on two judgment variables, and changed stereotypes and attitudes concerning both student groups to more positive ratings. Effects on judgments, stereotypes, and attitudes concerning students with a low social status background showed the highest stability until follow-up measures.

8.2 Limitations and prospects

There are, of course, limitations of this study, some pertaining to the design and some pertaining to the specificities of our sample, more precisely the accidentally unequal distribution of PTS with a Turkish immigration background across conditions. Although the proportion was low (3.2%) and excluding these cases showed no differences in the pattern of results, the unequal distribution may have influenced the coefficients concerning target students with a Turkish immigration background belonging to the in-group of these PTS. Furthermore, the results of our study at present only hold for the participating PTS, who were at the beginning of their teacher studies and members of one university. Results cannot be generalized to PTS in later phases of teacher studies or in-service student teachers or teachers—and other regions. The dissimilar result patterns across student backgrounds and judgment

---

[13] A detailed report and discussion of these results can be found in Tobisch et al. (2022).

dimensions suggest that a similar intervention for other social categories, e.g., gender or students with other immigration backgrounds or with vignettes for female students, may very likely yield other (or no) effects (e.g., Baadte, 2020; Civitillo et al., 2022) or differential effects depending on specific combinations of social categories, e.g., immigration background and social status (Glock & Kleen, 2023) or immigration background and gender (Bonefeld et al., 2020). Thus, future studies including interventions to reduce biases should increasingly differentiate social categories and combinations thereof to empirically identify specific effects. Similar to previous studies, future studies may include further judgment dimensions, e.g., social behavior (e.g., Böhmer et al., 2017; Glock, 2016; Glock et al, 2016), work behavior, school engagement, or motivation (e.g., Böhmer et al., 2017; Gentrup et al., 2018; Glock, 2016), and covariates possibly mediating or moderating intervention effects, e.g., motivation for teacher studies and professionalization, emotions associated with student backgrounds, specific capabilities of PTS, and, last but not least, social desirability tendencies. The latter may produce and increase unintended positive biases in explicit measures and judgments (i.e., concerning students with a Turkish immigration background in Germany), but may also mask persisting biases in implicit measures and judgments. The design of the study also did not allow to differentiate between the elements of the intervention in the analyses. Thus, we can make no suggestions on whether some elements were more effective than others or if the combination of elements was necessary to yield the effects. A possible solution to overcome this limitation would be a study with three or more conditions, e.g., one group with feedback on own attitudes, one group with information on judgment formation and judgment biases, and a control group. Thereby, judgment changes may be more easily attributed to changes in stereotypes and attitudes or, on the other hand, to more cognitive control of inadvertent (unconscious) influences of stereotypes and attitudes.

8.3 Conclusions

We cannot prevent teachers from implicitly or explicitly categorizing their students by family background or other characteristics (e.g., gender). However, we may and should strive to reduce judgment biases depending on these social categories. Because teachers often act and judge under pressure, e.g., in ongoing classroom interaction when judgments are more likely to be formed by stereotypes, we should also strive to change their possibly negative social stereotypes and attitudes, to furnish them with strategies to control inadvertent (unconscious) influences of these stereotypes and attitudes on their judgments of students' achievements, classroom behavior, and academic potential, and to motivate them to apply these strategies. The short intervention presented in this paper offers ideas to induce changes in judgments, stereotypes, and attitudes and it can be implemented in regular university courses for PTS and in training sessions for teachers. It requires less time and fewer resources than curricular concepts. But the differential and partially unintended effects of our intervention implicate, that such interventions may focus not only, or less, on those immigration backgrounds prominent in the regionally specific discourses on

discrimination and disparities, but also, or even stronger, on social status background, which is likewise stereotyped–and associated with the former. Moreover, social status background includes a broader range of students persistently facing biases and disparities and it is more often than not associated with further immigration backgrounds and religious orientations possibly facing stronger biases and prejudices, e.g., Arab, Muslim, or Romani. Furthermore, replication is needed and it remains open whether (experienced) teachers also benefit from a similar intervention. Last but not least, keeping in mind that social disparities in educational and vocational success at least partially ensue from positive biases for members (i.e., students) with a majority (no immigration and high social status) background, positive biases for students with an immigration (and often lower social status) background may contribute to a long-term reduction of social disparities in a society with high immigration rates.

Declarations

Conflict of interest We have no known conflict of interest to disclose.

Consent for publication The authors are fully responsible for all contents of this publication.

References

Agnello, M. F., & Mittag, K. C. (1999). *Comparing pre-service teachers' attitudes toward diversity: Internship and student teaching Experiences* (Paper presented at the Annual Meeting of the Southwest Educational Research Association in San Antonio, TX). https://eric.ed.gov/?q=ED430970

Akiba, M. (2011). Identifying program characteristics for preparing pre-service teachers for diversity. *Teachers College Record, 113*(3), 658–697. https://doi.org/10.1177/016146811111300304

Allport, G. (1954). *The nature of prejudice*. Addison-Wesley.

Baadte, C. (2020). Effects of counterstereotypic training on preservice teachers' assessment of students' academic performance. *Social Psychology of Education, 23*(4), 1003–1022. https://doi.org/10.1007/s11218-020-09562-2

Banks, J., Cochran-Smith, M., Moll, L., Richert, A., Zeichner, K., LePage, P., Darling-Hammond, L., & Duffy, H. (2005). Teaching diverse learners. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 232–274). Jossey-Bass.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*(6), 893–912. https://doi.org/10.1037//0022-3514.62.6.893

Baron, A. S. (2015). Constraints on the development of implicit intergroup attitudes. *Child Development Perspectives, 9*(1), 50–54. https://doi.org/10.1111/cdep.12105

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin, 142*(11), 1227–1274. https://doi.org/10.1037/bul0000067

Blaeschke, F., & Freitag, H.-W. (2021). Bildungsbeteiligung, Bildungsniveau und Bildungsbudget. [Educational participation, educational level, and educational budget]. In Statistisches Bundesamt, Wissenschaftszentrum Berlin für Sozialforschung & Bundesinstitut für Bevölkerungsforschung (Eds.), *Datenreport 2021. Ein Sozialbericht für die Bundesrepublik Deutschland* (pp. 101–120). Bundeszentrale für politische Bildung.

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, & D. Leutner (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 131–147). Springer International Publishing. https://doi.org/10.1007/978-3-319-50030-0_9

Bohner, G., & Wänke, M. (2006). *Attitudes and attitude change (repr.)*. Psychology Press.

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., & Karst, K. (2020). Do preservice teachers' judgments and judgment accuracy depend on students' characteristics? The effect of gender and immigration background. *Social Psychology of Education, 23*(1), 189–216. https://doi.org/10.1007/s11218-019-09533-2

Bonefeld, M., & Karst, K. (2020). Döner vs. Schweinebraten: Stereotype von (angehenden) Lehrkräften über Personen deutscher und türkischer Herkunft im Vergleich. [Doener vs. pork roast: Stereotypes of future teachers about persons of German and Turkish origin in comparison]. In S. Glock & H. Kleen (Eds.), *Stereotype in der Schule* (pp. 159–190). Springer.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition: A dual process model of impression formation* (Vol. 1, pp. 1–36). Lawrence Erlbaum Associates, Inc.

Casper, C., Rothermund, K., & Wentura, D. (2010). Automatic stereotype activation is context dependent. *Social Psychology, 41*(3), 131–136. https://doi.org/10.1027/1864-9335/a000019

Castro, A. J. (2010). Themes in the research on preservice teachers' views of cultural diversity. *Educational Researcher, 39*(3), 198–210. https://doi.org/10.3102/0013189X10363819

Civitillo, S., Ialuna, F., Lieck, D., & Jugert, P. (2022). Do infrahumanization or affective prejudice drive teacher discrimination against Romani students? A conceptual replication of Bruneau et al. (2020) in Germany. *Peace and Conflict: Journal of Peace Psychology, 28*(3), 340–344. https://doi.org/10.1037/pac0000609

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Colby, S., Bercaw, L., Clark, A. M., & Galiardi, S. (2009). From community service to service-learning leadership: A program perspective. *New Horizons in Education, 57*(3), 20–31.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Dickhäuser, O., Schöne, C., Spinath, B., & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept. [The Academic Self Concept Scales: Construction and evaluation of a new Instrument]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 23*(4), 393–405. https://doi.org/10.1024/0170-1789.23.4.393

Dickhäuser, O., & Stiensmeier-Pelster, J. (2003). Wahrgenommene Lehrereinschätzungen und das Fähigkeitsselbstkonzept von Jungen und Mädchen in der Grundschule. [Perceived teachers' ability evaluation and boys' and girls' concepts of their mathematical ability in elementary school]. *Psychologie in Erziehung und Unterricht, 50*, 182–190.

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. *Zeitschrift für Erziehungswissenschaft, 17*(Suppl. 24), 141–165. https://doi.org/10.25656/01:12370

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition, 25*(5), 603–637. https://doi.org/10.1521/soco.2007.25.5.603

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. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Academic Press.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "Just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology, 44*(2), 370–377. https://doi.org/10.1016/j.jesp.2006.12.004

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice* (2nd ed.). Teachers College Press.

Gentrup, S., Rjosk, C., Stanat, P., & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen. [Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations]. *Zeitschrift für Erziehungswissenschaft, 21*(4), 867–891. https://doi.org/10.1007/s11618-018-0806-2

Glock, S. (2016). 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., & Karbach, J. (2015). Preservice teachers' implicit attitudes towards racial minority students: Evidence from three implicit measures. *Studies in Educational Evaluation, 45*, 55–61. https://doi.org/10.1016/j.stueduc.2015.03.006

Glock, S., & Kleen, H. (2023). Disentangling the confounding effects of students' ethnicity and socioeconomic background on preservice teachers' judgments. *International Journal of Intercultural Relations, 95*, 101830. https://doi.org/10.1016/j.ijintrel.2023.101830

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., & Hörstermann, T. (2016). The higher the SES the better? Implicit and explicit attitudes influence preservice teachers' judgment of students. In C. Fields (Ed.), *Stereotypes and stereotyping: Misperceptions, perspectives and role of social media* (pp. 1–20). Nova.

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

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition. Attitudes, self-esteem, and stereotypes. *Psychological Bulletin, 102*(1), 4–27. https://doi.org/10.1037//0033-295x.102.1.4

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*(1), 3–25. https://doi.org/10.1037//0033-295X.109.1.3

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Gresch, C. & Becker, M. (2010). Sozial- und leistungsbedingte Disparitäten im Übergangsverhalten bei türkischstämmigen Kindern und Kindern aus (Spät-)Aussiedlerfamilien. [Social and achievement dependent disparities in the transition behavior of students with Turkish and German repatriate origin]. 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. 181–200). BMBF.

Grottkau, B. J., & Nickolai-May, S. (1989). An empirical analysis of a multicultural education paradigm for preservice teachers. *Educational Research Quarterly, 13*(4), 27–33.

Haberman, M., & Post, L. (1992). Does direct experience change education students' perceptions of low-income minority children. *Midwestern Educational Researcher, 5*(2), 29–31.

Hachfeld, A., Anders, Y., Schroeder, S., Stanat, P., & Kunter, M. (2010). Does immigration background matter? How teachers' predictions of students' performance relate to student background. *International Journal of Educational Research, 49*(2–3), 78–91. https://doi.org/10.1016/j.ijer.2010.09.002

Hao Kuo Tai, B. (1998). Editor's review: Power dynamics in the classroom. *Harvard Educational Review, 1998*(Fall). https://www.hepg.org/her-home/issues/harvard-educational-review-volume-68-issue-3/herarticle/_187

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, 237–271. https://doi.org/10.1146/annurev.psych.47.1.237

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

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

Holm, S. (1979). Simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Klapproth, F., Glock, S., Böhmer, M., Krolak-Schwerdt, S., & Martin, R. (2012). School placement decisions in Luxembourg: Do teachers meet the education ministry's standards? *Literacy Information and Computer Education Journal, Special Issue, 1*(1), 856–862. https://doi.org/10.20533/LICEJ.2040.2589.2012.0113

Kleen, H., Bonefeld, M., Glock, S., & Dickhäuser, O. (2019). Implicit and explicit attitudes toward Turkish students in Germany as a function of teachers' ethnicity. *Social Psychology of Education, 22*(4), 883–899. https://doi.org/10.1007/s11218-019-09502-9

Klug, J. (2011). *Modeling and training a new concept of teachers' diagnostic competence* [Doctoral dissertation, Technische Universität Darmstadt]. https://tuprints.ulb.tu-darmstadt.de/2838/

Kotzur, P. F., Schäfer, S. J., & Wagner, U. (2019). Meeting a nice asylum seeker: Intergroup contact changes stereotype content perceptions and associated emotional prejudices, and encourages solidarity-based collective action intentions. *British Journal of Social Psychology, 58*(3), 668–690. https://doi.org/10.1111/bjso.12304

Krolak-Schwerdt, S., Pit-ten Cate, I. M., & Hörstermann, T. (2018). Teachers' judgments and decision-making. Studies concerning the transition from primary to secondary education and their implications for teacher education. In O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, C. Lautenbach, & C. Kuhn (Eds.), *Assessment of learning outcomes in higher education* (pp. 73–101). Springer International Publishing. https://doi.org/10.1007/978-3-319-74338-7_5

Kruglanski, A. W., & Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology, 13*(1), 1–44. https://doi.org/10.1002/ejsp.2420130102

Kumar, R., & Hamer, L. (2013). Preservice teachers' attitudes and beliefs toward student diversity and proposed instructional practices: A sequential design study. *Journal of Teacher Education, 64*(2), 162–177. https://doi.org/10.1177/0022487112466899

Lehmann-Grube, S. K., Hartinger, A., Grassinger, R., Brandl-Bredenbeck, H. P., Ohl, U., Riegger, M., & Dresel, M. (2022). Einstellungen Lehramtsstudierender zu Heterogenität: Modell, Messinstrument und Ergebnisse der Validierung. [Preservice teacher students' attitudes towards diversity: Model, assessment, and results of a validation study]. In A. Hartinger, M. Dresel, E. Matthes, K. Peuschel, & U. Nett (Eds.), *Lehrkräfteprofessionalität im Umgang mit Heterogenität: Theoretische Konzepte, Förderansätze, empirische Befunde* (pp. 87–102). Waxmann.

Lorenz, G., Gentrup, S., Kristen, C., Stanat, P., & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen [Stereotypes among teachers? A study of systematic bias in teacher expectations]. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 68*(1), 89–111. https://doi.org/10.1007/s11577-015-0352-3

Lucas, D. J. (2011). *The effects of a service-learning introductory diversity course on pre-service teachers' attitudes toward teaching diverse student populations* [Doctoral dissertation, Liberty University, Lynchberg, VA]. https://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=1513&context=doctoral

Maaz, K., Baumert, J., & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert sich soziale Ungleichheit. [Formation of social inequality in the institutional context of schools: Where does social inequality emerge and increase]. In J. Baumert, K. Maaz, & U. Trautwein (Eds.), *Bildungsentscheidungen* (Zeitschrift für Erziehungswissenschaft, Sonderheft 12, pp. 11–46). VS.

Martin, R. J., & Koppelman, K. (1991). The impact of a human relations/multicultural education course on the attitudes of prospective teachers. *The Journal of Intergroup Relations, 18*(1), 16–27.

Miles, E., & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes & Intergroup Relations, 17*(1), 3–26. https://doi.org/10.1177/1368430213510573

Millisecond Software. (2016) Inquisit 5 Web (Windows) [Computer software]: Millisecond Software. Retrieved from http://www.millisecond.com/

Ohlsson, R. (2018). Student-teacher conferencing in Swedish upper secondary school: Dimensions of dominance and relations between perspectives in institutional discourse. *Learning, Culture and Social Interaction, 18*, 113–123. https://doi.org/10.1016/j.lcsi.2018.03.010

Organisation for Economic Co-operation and Development (OECD). (2019). *PISA 2018 results (Volume 2): Where all students can succeed*. OECD Publishing.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1971). *The measurement of meaning* (8th ed.). University Press.

Petersen, L.-E. & Six-Materna, I. (2006). Stereotype. [Stereotypes]. In H.-W. Bierhoff (Ed.), *Handbuch der Psychologie: Vol. 3. Handbuch der Sozialpsychologie und Kommunikationspsychologie* (pp. 430–436). Hogrefe.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*(6), 922–934. https://doi.org/10.1002/ejsp.504

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 123–205). Academic Press.

Pit-ten Cate, I., Krolak-Schwerdt, S., Glock, S., & Markova, M. (2014). Improving teachers' judgments. In S. Krolak-Schwerdt, S. Glock, & M. Böhmer (Eds.), *Teachers' professional development* (pp. 45–61). Sense Publishers. https://doi.org/10.1007/978-94-6209-536-6_4

Pit-ten Cate, I. M., Krolak-Schwerdt, S., & Glock, S. (2016a). 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

Pit-ten Cate, I. M., Krolak-Schwerdt, S., Hörstermann, T., & Glock, S. (2016b). *Theoretical knowledge and formal decision rules: Can we reduce bias in orientation decisions?* Paper presented at the 4. Tagung der Gesellschaft für Empirische Bildungsforschung [4th Meeting of the Association for Empirical Educational Research]. Berlin. https://www.gebf-ev.de/deutsch/tagungen-der-gebf/2016-4-gebf-tagung-berlin/

Ramm, G. & Adamsen, C. (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. [PISA 2003: Documentation of assessment instruments]. Waxmann.

Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *The British Journal of Educational Psychology, 76*(3), 429–444. https://doi.org/10.1348/000709905X53589

Schleppegrell, M. J., & O'Hallaron, C. L. (2011). Teaching academic language in L2 secondary settings. *Annual Review of Applied Linguistics, 31*, 3–18. https://doi.org/10.1017/S0267190511000067

Sleeter, C. E. (2001). Preparing teachers for culturally diverse schools: Research and the overwhelming presence of whiteness. *Journal of Teacher Education, 52*(2), 94–106. https://doi.org/10.1177/0022487101052002002

Soble, J. R., Spanierman, L. B., & Liao, H.-Y. (2011). Effects of a brief video intervention on White university students' racial attitudes. *Journal of Counseling Psychology, 58*(1), 151–157. https://doi.org/10.1037/a0021158

Sparks, W. G., & Verner, M. E. (1995). Intervention strategies in multicultural education. *The Physical Educator, 52*(4), 170–180.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. [Teachers' judgment accuracy of student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie, 19*(1/2), 85–95. https://doi.org/10.1024/1010-0652.19.12.85

Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte. [Teachers' perceptions and judgments of students]. In H. Ditton (Ed.), *Kompetenzaufbau und Laufbahnen im Schulsystem: Ergebnisse einer Längsschnittuntersuchung an Grundschulen* (pp. 171–198). Waxmann.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–762. https://doi.org/10.1037/a0027627

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues, 25*(4), 79–97. https://doi.org/10.1111/j.1540-4560.1969.tb00620.x

Tobisch, A. (2019). *Herkunftsassoziierte Lehrkrafturteile und -erwartungen: Soziale Kognitionen und Urteilsbildungsprozesse im Kontext einer ethnisch und sozial heterogenen Schülerschaft.* [Teacher judgments and expectations associated with students' backgrounds: Social cognition and judgment formation processes in the context of an ethnically and socially heterogeneous student population]. [Doctoral dissertation, University of Augsburg]. urn:nbn:de:bvb:384-opus4-638356

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

Tobisch, A., & Dresel, M. (2020). Fleißig oder faul? Welche Einstellungen haben angehende Lehrkräfte gegenüber Schüler*innen aus unterschiedlichen sozialen Schichten. [Diligent or lackadaisical? Which attitudes towards students of different social stratums do future teachers hold?]. In S. Glock & H. Kleen (Eds.), *Stereotype in der Schule* (pp. 133–158). Springer.

Tobisch, A., Lehmann-Grube, S. K., Shames, L., & Dresel, M. (2022). Herkunftsbezogene Einstellungen und Stereotype Lehramtsstudierender: Eine Interventionsstudie. [Pre-service teacher students' attitudes and stereotypes: An intervention study]. In S. Glock (Ed.), *Stereotype in der Schule II: Ursachen und Möglichkeiten der Intervention* (pp. 301–340). Springer VS.

Trent, S. C., Kea, C. D., & Oh, K. (2008). Preparing preservice educators for cultural diversity. How far have we come? *Exceptional Children, 74*(3), 328–350. https://doi.org/10.1177/001440290807400

Walberg, H. J. (1982). Educational productivity: Theory, evidence, and prospects. *Australian Journal of Education, 26*(2), 115–122. https://doi.org/10.1177/000494418202600202

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation, 24*(3–5), 124–179. https://doi.org/10.1080/13803611.2018.1548798

Wasonga, T. A. (2005). Multicultural education knowledgebase, attitudes and preparedness for diversity. *International Journal of Educational Management, 19*(1), 67–74. https://doi.org/10.1108/09513540510574966

Wenz, S. E., & Hoenig, K. (2020). Ethnic and social class discrimination in education: Experimental evidence from Germany. *Research in Social Stratification and Mobility, 65*, 100461. https://doi.org/10.1016/j.rssm.2019.100461

Weis, M., Müller, K., Mang, J., Heine, J.-H., Mahler, N., & Reiss, K. (2019). Soziale Herkunft, Zuwanderungshintergrund und Lesekompetenz. [Social background, immigration background and reading competence]. In K. Reiss, M. Weis, E. Klieme, & O. Köller (Eds.), *PISA 2018: Grundbildung im internationalen Vergleich* (pp. 129–162). Waxmann.

Sabine K. Lehmann-Grube, Dr. phil., is research assistant at the University of Augsburg, Department of Psychology and coordinator of evaluation studies in the project "Förderung der Lehrerprofessionalität im Umgang mit Heterogenität". Her research interests focus social cognition and action in complex situations including socially constructed knowledge bases and mixed methods methodology.

Anita Tobisch, Dr. phil., is research assistant at the University of Augsburg, Department of Psychology. Her research interests include diversity in educational contexts as well as teachers' judgments and diagnostic competence.

Markus Dresel, Prof. Dr., is a full professor in Psychology at the University of Augsburg, Department of Psychology. His research interests include motivation and self-regulated learning in school and higher education contexts, associations of academic achievement with gender and ethnic and social backgrounds as well as teacher training and teachers' professional development.