



# Non-communicated judgements of, versus feedback on, students' essays: Is feedback inflation larger for students with a migration background?

Anna K. Nishen<sup>1</sup> · Ursula Kessels<sup>1</sup>

Received: 3 December 2020 / Accepted: 21 October 2021 / Published online: 18 November 2021  
© The Author(s) 2022

## Abstract

When providing feedback, teachers are concerned not only with the simple transmission of information, but also with motivational and interpersonal dynamics. To mitigate these concerns, teachers may inflate feedback by reducing negative or increasing positive content. The resulting difference between initial judgments and feedback may be even more drastic for ethnic minority students: In non-communicated judgments, negative stereotypes may result in more negative judgments, whereas in feedback, concerns about being or appearing prejudiced may inflate feedback towards ethnic minority students. These hypotheses were tested in a sample of 132 German teacher students in a 2 (between subjects: feedback vs. non-communicated judgment) × 2 (within subjects: target student's migration background: Turkish vs. none) design in which participants read supposed student essays and provided their written impressions to the research team or the supposed student. Findings revealed that teacher students' feedback was more positive than their non-communicated judgments on a multitude of dimensions. Contrary to expectations, these effects were not stronger when the student had a Turkish migration background. Instead, teacher students rated the essay of the student with a Turkish migration background more favorably both in the judgment and feedback conditions. Our results suggest that teachers adapt their initial judgments when giving feedback to account for interpersonal or motivational dynamics. Moreover, ethnic minority students may be especially likely to receive overly positive feedback. While the motivational/interpersonal dynamics may warrant some inflation in feedback, negative consequences of overly positive feedback, for which ethnic minority students may be especially vulnerable, are discussed.

---

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

Ursula Kessels  
ursula.kessels@fu-berlin.de

<sup>1</sup> Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, D- 14195 Berlin, Germany

**Keywords** Teacher judgment · Positive feedback bias · Feedback inflation · Ethnic bias · Migration background · Stereotypes

## 1 Introduction

*Receiving* feedback is an integral part of learning, informing students about where they stand in relation to specific criteria such as their learning goals (Hattie & Timperley, 2007). However, hearing that one's performance was substandard can be hurtful, frustrating, and threatening to one's self-concept and motivation (Ilies et al., 2010; Kerssen-Griep, 2001; Weidinger et al., 2016). With these reactions in mind, *giving* feedback to others can also be a challenge. When feedback contains criticism, people worry about demoralizing the recipient of feedback or hurting the relationship (Brown & Levinson, 1987; Rosen & Tesser, 1970). This may especially apply to teachers, as they are tasked not only with students' learning but also with their social and emotional well-being.

When giving feedback, teachers aim to enable improvement, but also to motivate and encourage students (Müller & Ditton, 2014). Students themselves stress the importance of including positive aspects in feedback (Ferguson, 2011). Therefore, teachers may adapt their initial non-communicated judgments to these motivational and interpersonal demands of the feedback situation (*feedback inflation*). Importantly, teachers may have additional concerns in giving feedback to students belonging to negatively stereotyped groups. In the US, teachers are concerned with being and/or appearing to be prejudiced towards these students (Bentley-Edwards et al., 2020; Marshall, 1996). This concern may explain the more lenient feedback for Black and Latinx students that has been found in extensive research by Harber and colleagues (*positive feedback bias*; Croft & Schmader, 2012; Harber, 1998, 2004; Harber et al., 2010, 2012, 2019). While feedback towards ethnic minority students may be more positive, research shows a consistent negative bias in judgments of students' written material submitted to third parties (e.g., Bonefeld & Dickhäuser, 2018; Holder & Kessels, 2017; Glock, 2016; Parks & Kennedy, 2007) as well as in expectations more generally (e.g., Malouff & Thorsteinsson, 2016; Wang et al., 2018). While feedback represents an impression that is communicated to the student and which may have been adapted to the social situation, a non-communicated judgment is an impression that is not shared with the student (but may be communicated to a third party, such as a research team). As a result of both negative biases in non-communicated judgments and the positive feedback bias, the difference between non-communicated judgments and feedback may be even more pronounced for ethnic minority students. In Germany, this may be the case for students with a Turkish migration background since negative stereotypes about their academic abilities are well-known among teacher students (e.g., Bonefeld & Karst, 2020; Froehlich et al., 2016).

The present research has two main aims: First, we examine the extent to which feedback is inflated (more positive) compared to judgments communicated to a third party rather than the student. Second, we aim to study both positive and negative biases towards ethnic minority students within the same study, enabling

us to determine the importance of the context—here, whether one is addressing or simply judging the student—in eliciting biases. To address these aims, we asked teacher students to document their impressions of essays written by students with a Turkish migration background and without a migration background. Half of the participants handed in a non-communicated judgment, whereas the others believed that they were providing feedback.

## 2 Feedback inflation: differences between non-communicated judgments and feedback

Feedback is understood as “information provided by an agent [...] regarding aspects of one’s performance or understanding” (Hattie & Timperley, 2007, p. 81). In addition to this aspect of information transfer, feedback has a motivational character, aiming to encourage students to improve their learning and performance (Müller & Ditton, 2014). In our research, we define feedback as including any information regarding performance that is relayed to students, unlike non-communicated judgments, which are not shared with the students themselves. In the following, when we discuss judgments, we are referring to judgments that are not communicated to the student.

Because giving feedback is inherently a social situation (Strijbos & Müller, 2014), interpersonal concerns may influence the valence of feedback. In these situations, people use self-presentation strategies to achieve or maintain being liked and/or respected (Baumeister & Leary, 1995; Brown & Levinson, 1987; Jones & Pittman, 1982; Leary & Kowalski, 1990). In order to be liked, people are theorized to deploy ingratiation techniques, i.e., ways to establish warmth and friendliness (Jones & Pittman, 1982). When giving feedback, senders may ingratiate themselves by giving feedback that eliminates negative aspects and includes additional compliments to buffer their initial judgment (Cox et al., 2011; Croft & Schmader, 2012; Jones & Pittman, 1982). Thus, the feedback teachers give to their students may be more lenient than the opinions they initially formed of students’ work.

Research has made important contributions to our understanding of feedback (e.g., Butler, 1994; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). To our knowledge, however, there is no research in the school context as to whether teachers inflate their feedback due to interpersonal and/or motivational concerns. However, studies suggest that this may be the case in the university setting (Colletti, 2000; Qadan et al., 2013). In particular, grades are more positive than judgments given to the research team when they are determined in a face-to-face conversation with the student (Colletti, 2000; Qadan et al., 2013), particularly because faculty avoided bringing up negative aspects (Qadan et al., 2013).

In organizational and social psychology, the valence of feedback in comparison to judgments has received more attention. Studies in these areas have found support for the hypothesis that people in a teaching role may be reluctant to give negative feedback and experience discomfort when they do (Ginsburg et al., 2016; Ramani et al., 2018; Waung & Highhouse, 1997). Generally speaking, people may carefully select the information they relay to others to adhere to politeness norms by omitting

negative information strategically (Bergsieker et al., 2012; Brown & Levison, 1987). This *negativity omission effect* has also been demonstrated alongside *sugarcoating* when it comes to negative feedback, particularly for early-career medical doctors who are given feedback by supervising faculty (Ginsburg et al., 2016; Qadan et al., 2013; Ramani et al., 2018). Lastly, in two experiments designed to assess prosocial lying, i.e., lies intended to benefit the other, university students gave overly lenient feedback to spare another student embarrassment or pain (Lupoli et al., 2017).

Overall, research on negativity omission suggests that people are motivated to be liked and that this may lead to more lenient feedback by excluding negative or including positive aspects (e.g., Bergsieker et al., 2012; Ramani et al., 2018). However, an important addition to the existing literature would be to estimate how much more lenient feedback is by comparing feedback to a judgment not impacted by concerns for the recipient's reaction. To our knowledge, only two studies have directly compared judgments and feedback, demonstrating feedback inflation (Lupoli et al., 2017; Qadan et al., 2013). While studies on negativity omission have looked at feedback situations at universities, there have been no studies in the school context. To understand whether and to what extent these dynamics play out in school, further studies involving teachers (in training) and students are required.

## 2.1 The impact of stereotypes in judgment and in feedback situations

Importantly, the gap between what a judgment contains and what is conveyed in feedback may be larger for those who are negatively stereotyped in the respective domain. Stereotypes are defined as “beliefs about the characteristics, attributes, and behaviors of members of certain groups” (Hilton & van Hippiel, 1996, p. 240). Based on stereotypes, varying expectations for different groups emerge, and expectations are sometimes even defined as an integral part of stereotypes (American Psychological Association, 2020; Pendry, 2014). If there exists a negative stereotype regarding academic abilities, students belonging to this negatively stereotyped group can be judged more negatively (e.g., Anderson-Clark et al., 2008; Bonefeld & Dickhäuser, 2018; Glock et al., 2013; Quinn, 2020) while at the same time being given more positive feedback than their peers not belonging to a negatively stereotyped group (Croft & Schmader, 2012; Harber, 1998, 2004; Harber et al., 2010, 2012, 2019). If either or both of these effects come into play, a greater difference between feedback and judgments is expected for negatively stereotyped students.

## 2.2 Negative bias towards ethnic minority students in judgments

A considerable amount of research has unearthed the ways in which students belonging to negatively stereotyped racial/ethnic minority groups are judged less favorably than students belonging to racial/ethnic majority groups in North America and Europe (Anderson-Clark et al., 2008; Bonefeld & Dickhäuser, 2018; Glock, 2016; Glock et al., 2013; Holder & Kessels, 2017; Parks & Kennedy, 2007; Sprietsma, 2013; but see also Baker et al., 2015; Glock & Krolak-Schwerdt, 2014). Research regarding expectations effects has found pervasive evidence that teachers have lower

expectations, e.g., regarding future performance, for students belonging to racial/ethnic minority groups compared racial/ethnic majority groups (e.g., Lorenz et al., 2016; McKown & Weinstein, 2008; Meissel et al., 2017; Quinn, 2020; Tobisch & Dresel, 2017; van den Bergh et al., 2010; Wang et al., 2018; for a meta-analysis, see Malouff & Thorsteinsson, 2016; Tenenbaum & Ruck, 2007).

In Germany, various experiments have explored negative bias in judgments of students with a Turkish migration background, who are stereotyped as academically less capable and motivated (Bonefeld & Karst, 2020; Froehlich & Schulte, 2019; Froehlich et al., 2016; Kahraman & Knoblich, 2000). Teachers and teacher students show negative bias in judgments of work by supposed primary school students with a Turkish migration background (Bonefeld & Dickhäuser, 2018; Glock, 2016; Holder & Kessels, 2017; Sprietsma, 2013). For example, teacher students graded essays of students with a Turkish migration background worse even though they found the same number of errors in students' work (Bonefeld & Dickhäuser, 2018).

In addition to research using materials supposedly created by students, a negative ethnic bias has also been demonstrated using experimental manipulation of vignettes (Glock, 2016; Glock et al., 2013). When a vignette described a below-average student, teachers rated ethnic minority students as less proficient in German compared to ethnic majority students (Glock, 2016). Teachers also recommended the highest secondary school track less often for students with a migration background and expected them to be less successful there compared to students without a migration background (Glock et al., 2013). Lastly, teacher students judged German proficiency scores of students with a Turkish-origin name as lower than those of students with a German-origin name (Holder & Kessels, 2017).

Overall, experimental evidence demonstrates a negative bias of teachers and teacher students towards students from non-majority ethnicities, and that it is present in grading, recommendations for further education, and expectations for their future success (Malouff & Thorsteinsson, 2016; Wang et al., 2018). Typically, these negative biases are explained by the persistent negative stereotypes that inadvertently influence people's judgments and expectations (Macrae & Bodenhausen, 2000; Macrae et al., 1994). If social categories and the corresponding stereotypes are activated in teachers, they may non-consciously adjust their judgment downwards based on negative stereotypes about students with a Turkish migration background, resulting in a negative bias.

### 2.3 Positive feedback bias towards ethnic minority students

In another strand of research, however, different results emerged, indicating not a negative but a positive bias towards ethnic minority students (e.g., Croft & Schmader, 2012; Harber, 2004). The crucial difference in their designs lies in the addressee: In experiments finding a negative bias, participants revealed their impressions to a third party (the researchers; e.g., Bonefeld & Dickhäuser, 2018), but experiments finding a positive bias asked participants to communicate their impressions to the target students, i.e., to give feedback (e.g., Harber, 2004).

In a series of experiments, Harber and colleagues compared feedback given to Black and Latinx students to feedback given to White students in the US (Harber, 1998, 2004; Harber et al., 2010, 2012, 2019). In these experiments, university students received an essay supposedly written by a fellow student who was portrayed as either Black or White. When the author was supposedly Black, participants gave more positive comments on the content of the essay and rated it higher than when the author was White. However, there were no differences regarding comments on and ratings of essay mechanics (e.g., grammar). Harber and colleagues replicated his findings in experiments involving verbal feedback (Harber, 2004) and in samples of teacher students and teachers (Harber et al., 2010, 2012). Research conducted in Canada found the same positive feedback bias when comparing feedback that university students gave to Indigenous versus White Canadian peers. Whereas Harber and colleagues investigated how positive the feedback was overall, Croft and Schmader (2012) examined whether feedback was more positive towards Indigenous Canadians because of an inflation of positive comments, a reduction of negative comments, or both. While both Indigenous and White Canadians received similar levels of praise, Indigenous Canadians received less criticism (*feedback withholding effect*). Thus, it seems that the positive feedback bias may be due to people holding back criticism rather than overemphasizing positive aspects.

Harber (1998, 2004) explains these findings using a motivational lens: Because White people aim to be and/or appear to be egalitarian and free of prejudice (Bergsieker et al., 2010; Gaertner & Dovidio, 1986; Trawalter et al., 2009), they will avoid ambiguous situations in which their behavior may be construed as prejudiced or discriminatory. However, if the situation cannot be avoided, they may overcorrect negative reactions in an attempt to demonstrate that they are unprejudiced—resulting in more positive reactions to ethnic minority than White people. Indeed, there is evidence of overcorrection in interracial interactions in general (Bergsieker et al., 2010; Littleford et al., 2005; Mendes & Koslov, 2013). Importantly, giving feedback is also an ambiguous situation in which behavior may be interpreted as prejudiced, particularly criticism. This negative feedback could be either understood as warranted or as compromised by prejudice. To avoid the latter, people may artificially “buffer” their feedback, making it more lenient. However, this may not be necessary on less ambiguous feedback tasks for which there are clear and universal rules. This may explain why a positive feedback bias was consistently found regarding feedback about essay content but not essay mechanics (e.g., Harber et al., 2012). Thus, we expect German teacher students to give less negative feedback to ethnic minority students on those feedback dimensions involving personal discretion rather than on those involving clear-cut rules.

### 3 Personal factors influencing the positive feedback bias

Not all people may show the positive feedback bias to the same extent since their motivation to not be or not appear to be prejudiced may vary. Harber et al. (2010) argue that this motivation may originate either in the desire to see oneself as unprejudiced (self-concept) or to demonstrate to others that one is unprejudiced

(self-presentation). Plant and Devine (1998) developed this distinction between internal motivation and external motivation to respond without prejudice. According to them, an internal motivation is based on the wish to be unprejudiced and to live in accordance with these values, whereas an external motivation reflects concerns of being socially excluded because of the perceived social pressure to be unprejudiced. Because internal and external motivation to respond without prejudice are theoretically and empirically independent (Plant & Devine, 1998), people may be both internally and externally motivated, motivated by only one source, or not at all. If the positive feedback bias primarily emerges due to self-presentation concerns, people who are motivated purely by external reasons may be especially likely to exhibit a positive feedback bias (Croft & Schmader, 2012). However, if self-concept concerns are at the heart of positive feedback bias, people with a high internal motivation may be more likely to give positively biased feedback (Crosby & Monin, 2007). Overall, people's motivation to respond without prejudice has been linked to a positive bias, but the results have not been consistent across studies. Since the present research focuses on a different cultural context, it is worthwhile re-examining whether a moderating effect is present.

In addition to motivation explicitly linked to responding without prejudice, people's general need to present themselves as unprejudiced might be associated with greater positive feedback bias. Importantly, people high in self-esteem may feel less pressure to engage in impression management techniques since they may already feel secure in the knowledge that they are likeable (Leary et al., 1995). However, people low in self-esteem may doubt that others view them positively, therefore being more motivated to demonstrate their goodwill and likeability—as well as their lack of prejudice in interracial/ethnic interactions. In addition to the total level of self-esteem, it may play a role how contingent a person's self-esteem is on the approval by others, i.e., how much importance they place on the opinion that others have of them (Crocker et al., 2003). People high in contingent self-esteem may be more motivated to demonstrate that they are unprejudiced, as societal norms reflect that prejudice is disapproved of (Crandall et al., 2002). To our knowledge, the moderation of the positive feedback bias by participants' trait self-esteem and contingency of self-esteem on others' approval has not yet been tested.

#### 4 Research gap and present study

Both judging students' achievements and giving feedback to students are important elements of teachers' work. Given the importance of useful and accurate feedback for students' learning (Hattie, 2008; Hattie & Timperley, 2007), the fact that to date no studies exist that compared if and how impressions differ when teachers judge a student's performance and give feedback to them is surprising. The first aim of our study is to fill in this research gap regarding feedback inflation by comparing teachers' impressions in these two situations.

So far, research on possible negative ethnic biases in judgments and research on possible positive biases when giving feedback to ethnic minority students was unconnected and conducted in separate studies. Thus, the second aim of our study



is to for the first time apply the same material and examine the same dependent variables in a study that compares situations of conveying an impression of target students' essays with and without the prospect that it will be forwarded to these students.

Conducting this experiment in a large city in Germany, we are able to test whether a positive feedback bias is present in German teacher students for the first time. While several studies from the US and Canada report these positive biases towards members of negatively stereotyped minorities (e.g., Croft & Schmader, 2012; Harber et al., 2012), it is not known if such positively biased feedback also occurs in Germany regarding students with a Turkish migration background. It is important to stress that the relationships between White and Black Americans as well as between descendants of colonialists and Indigenous peoples in Canada differ in important historical and social aspects from the relationship between Germans without a migration background and descendants of Turkish labor migrants. However, US-American teachers have shown the positive feedback bias towards Latinx Americans as well (Harber et al., 2012), which indicates that positive bias affects members of more recently migrated, negatively stereotyped groups as well. Moreover, the central tenets in Harber's (1998) reasoning are present in the German context as well. Persistent negative stereotypes about Turks and Turkish-Germans are well-known (e.g., Bonefeld & Karst, 2020) and—if applied—they are at odds with the desire to be or appear unprejudiced. German teacher students grapple with different and fair approaches to cultural heterogeneity (Hachfeld et al., 2011) and research on the differentiation between explicit and implicit attitudes towards Germans with a Turkish migration background suggests that prejudice may exist implicitly, but is not considered socially acceptable (Glock et al., 2020).

Lastly, the moderating factors of a possible positive feedback bias towards students with a Turkish migration background will be studied. Earlier studies in Canada found that a greater external motivation to respond without prejudice was associated with a greater positive feedback bias when internal motivation was low at the same time (Croft & Schmader, 2012), which is yet to be examined in Germany. Furthermore, we add to the existing literature by examining whether those with lower self-esteem and those with greater contingency of their self-esteem on the approval of others may be more prone to exhibit a positive bias in feedback.

In detail, we will test the following hypotheses: We expected that teacher students' feedback would be more positive than their judgments made to a third party, the research team (H1). Moreover, we expected that this effect would be stronger for students with than for students without a migration background due to a negative bias towards the former in the judgment condition and a positive bias in the feedback condition (H2a), with the exception of dimensions regarding writing mechanics and style (H2b). Lastly, we expected that in the feedback condition, the positive bias would be stronger in participants who had lower self-esteem, higher contingency of self-esteem on others' approval and who were highly externally motivated to respond without prejudice but simultaneously not internally motivated to do so (H3a-c).



## 5 Methods

### 5.1 Participants

In total, 160 teacher students participated in the study (74.4% female, 78.8% without a migration background). On average, they were 24.7 years old ( $SD=5.8$ ;  $range=18-50$  years). Participants were excluded from all analyses if their writing was substandard ( $n=1$ ), took less than 15 min to read each essay and write the judgments/feedback ( $n=7$ ), questioned the cover story ( $n=4$ ) or had significant trouble understanding the procedure ( $n=1$ ). Moreover, participants who could reasonably be expected to be subject to negative stereotypes based on their own migration background (e.g., Turkish)—and therefore might not be concerned about being prejudiced against this group—were excluded from the analyses, as well as those who did not fill in their migration background ( $n=15$ ). The final sample consisted of 132 participants.

### 5.2 Development and pilot testing of the materials

In total, three pilot studies were conducted to ensure the validity of the materials used in the study. In the first pilot study, seventeen 11<sup>th</sup> grade students were asked to write essays on “digital media in schools.” Based on these materials, four low-quality and two high-quality essays on this topic were developed. Subsequently, 20 teacher students read each of these essays and rated style, mechanics (e.g., punctuation), and content. A high-quality buffer essay as well as the two low-quality essays which did not differ on any item apart from punctuation were selected for the study ( $M=611.3$  words,  $SD=71.1$ ,  $range=534-674$ ).

Moreover, we conducted a pilot study to select names to be attached to the essays. A separate sample of 48 teacher students indicated for 27 German- and Turkish-origin male names whether they were familiar with the name and whether they thought the name was German-origin, Turkish-origin or were uncertain. Five Turkish- and German-origin names each were chosen that participants were most familiar with (minimum: 91.6%) and for which most participants correctly indicated the origin (minimum: 88%). No Turkish-origin names were mistaken for German-origin names; participants may have been uncertain about their origin due to the similarity with Arabic-origin names.

A third pilot study with 254 teacher students was conducted to pre-test an adapted version of Sommer's (2017) translation of Plant and Devine's (1998) questionnaire on internal and external motivation to respond without prejudice. With regards to external motivation to respond without prejudice, an exploratory factor analysis revealed double factor loadings for one item regarding “political correctness” and this item was therefore excluded from the scale for the main study.

### 5.3 Design and procedure

Our experiment had a 2 (evaluation condition: non-communicated judgment vs. feedback) by 2 (target students' migration background: Turkish vs. none) design in which the former was varied between subjects and the latter varied within subjects. In the judgment condition, participants learned that the goal of the study was to find out how teacher students evaluate comprehensive essays and were explicitly informed that the student authors would not see their responses. Thus, participants expected to provide a judgment of the essays to a third party, the research team. In the feedback condition, participants heard that they would participate in an "Online-Feedback project" providing "easily accessible feedback" to students on their essays, and that they would send their feedback via e-mail. Thus, these participants expected to provide feedback to a real student, though their anonymity was still protected. All participants believed that the students were 11th grade students from the city in which the study took place.

Participants arrived at the laboratory at set time points and the evaluation condition was randomized for each timeslot (typically 2–3 participants). After the experimenter relayed the cover story, participants started working on a booklet containing (again) the cover story, the original writing prompt for the students, and three essays. A cover page indicating age, grade and name preceded each essay. The first essay was a buffer essay, supposedly written by a girl with a German name, and it was supposed to familiarize participants with the task and strengthen the cover story (not included in analyses). The next two essays were supposedly written by boys with a German- and Turkish-origin name, respectively. The order of the male students' migration background, the content of the essays, and the handwriting were all counterbalanced.

After reading each essay, participants filled in either a "judgment form" or a "feedback form" on their impression of the essay, including both an open text and rating format to cover multiple dimensions of the impression. At the top of the page, participants were prompted to type the name of the student to ensure that they were aware of his supposed migration background. In the feedback condition, participants attached their filled-in feedback form to a pre-written e-mail and sent it to the supposed e-mail address of the student before moving on to the next essay. After reading all essays, they filled in a questionnaire containing the scales on self-esteem, contingency of self-esteem, internal/external motivation to respond without prejudice and demographic data.

### 5.4 Measures

#### 5.4.1 Impressions of the essay

The first page of the judgment/feedback form consisted of an open answer format. Participants were requested to address strengths and weaknesses of the essay regarding content as well as style/mechanics. A team of four coders coded several

dimensions in each text: a) total number of negative and positive comments, b) number of positive and negative comments regarding the content of the essay, e.g., “He does not elaborate on weaknesses and risks,” c) number of positive and negative comments regarding style/mechanics of the essay, e.g., “Your greatest problem is capitalization.” d) number of suggestions for improvement, e.g., “I recommend using more paragraphs,” and e) coders’ perception of the tone of the open format answer as a whole on a scale from 1 (*The tone of the text is very cold or tough. It seems distanced, chilly, or possibly hurtful*) to 7 (*The tone of the text is very warm or considerate. It seems friendly and full of understanding*). Codes were applied to meaningful units, e.g., subclauses. Each open format answer was coded by two coders blinded to migration background. Coders used a subsample of texts for coding practice, and their subsequent inter-rater reliability was .77–.94 for all variables (ICC(1, 4)). If coders differed more than two points on a coding dimension, these deviations were discussed, resulting in maximum differences of one point. For further data analysis, the average score across the two raters was used. On average, participants provided 14.56 comments per essay ( $SD=4.3$ ;  $range=5-28$ ), and overall number of comments did not differ between the judgment and feedback conditions,  $t(130)=-0.40$ ,  $p=.691$ , or by the student’s migration background,  $t(131)=0.43$ ,  $p=.665$ .

The second page of the document consisted of five one-item scales measuring a) the grade the participant would give the essay on a scale of 1–15,<sup>1</sup> b) their impression of the student’s writing skills, of the style/mechanics of the essay, and of the content of the essay on three scales from 1 (*very bad*) to 7 (*very good*), and c) the degree to which they agreed with the statement “Based on what I have read I believe that [you/the student] could successfully master the demands of studying at a university” on a scale from 1 (*I do not agree at all*) to 7 (*I agree strongly*).

#### 5.4.2 Self-esteem and contingency of self-esteem

Self-esteem was measured using Ferring and Filipp’s (1996) translation of Rosenberg’s (1979) questionnaire. The scale consists of ten items answered on a Likert scale from 1 (*disagree strongly*) to 4 (*agree strongly*), e.g., “I fear there is not much I can be proud of.” The scale was reliable (Cronbach’s  $\alpha=.86$ ). The contingency of self-esteem on others’ approval was assessed using a 5-item scale by Schwinger et al. (2015), e.g., “I don’t mind if others have a negative opinion of me,” which are answered on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Reliability was good (Cronbach’s  $\alpha=.80$ ). After reverse-coding the relevant items, higher values indicated greater self-esteem and greater contingency of self-esteem, respectively.

#### 5.4.3 Internal/external motivation to respond without prejudice

Participants’ motivation to respond without prejudice was assessed using Plant and Devine’s (1998) questionnaire as translated into German by Sommer (2017), further

<sup>1</sup> This is the standard German grading system in the last two years of secondary school. Higher values indicate better grades.

adapting the items by substituting “people with a migration background” as the relevant target group. The final questionnaire consists of two independent subscales that measure internal and external motivation to respond without prejudice (five and four items, respectively), e.g., “It is important to me personally to act without prejudice” and “I try to give the impression of not being prejudiced so I am not rejected by others.” Participants answered the items on a scale from 1 (*I do not agree at all*) to 9 (*I agree entirely*). The subscales had adequate to good reliability, Cronbach’s  $\alpha = .75$  (internal) and  $.80$  (external).

## 6 Results

### 6.1 Analysis plan

To test whether the impressions of the essays differed based on the evaluation condition and did so differently for students with a Turkish and without a migration background, 2 (evaluation condition: non-communicated judgment vs. feedback)  $\times$  2 (target student migration background: none vs. Turkish) mixed ANOVAs were conducted. Subsequently, we tested whether the expected effects in the feedback condition were moderated by participants’ self-esteem, contingency of self-esteem and their motivation to respond without prejudice in regression analyses. For these moderator analyses in the feedback condition, the differences between the ratings of and comments about the essays of the student with a Turkish migration background and the student without a migration background were used as dependent variables. That is, these analyses predict the extent of a positive or negative bias. For each dependent variable, two regressions were conducted: the first including self-esteem and contingency of self-esteem on approval of others and the second including internal and external motivation to respond without prejudice as well as their interaction. All continuous predictors were centered at the mean.

Five participants (3.8%) had some missingness on the rating-scale dependent variables (three missed individual items and two missed the second page containing the rating scales). Due to these infrequent non-responses, the sample size varies slightly across analyses ( $n = 128$ – $132$ ).

### 6.2 Preliminary analyses

Because we randomized the order of the writer’s migration background, half of participants read an essay by a student with a German name first ( $n = 66$ ) and the other half an essay by student with a Turkish name ( $n = 66$ ). To test whether we needed to control for writer order, independent-sample *t*-tests were conducted predicting the difference scores (see 6.1) on all dependent variables. For those variables for which there was a significant difference (essay quality content, essay quality style/mechanics, total number of negative comments, negative comments on content, negative comments on style/mechanics), writer order was included as a control variable in

the analyses (reference group: student with a Turkish migration background first). No significant differences on the dependent variables by participants' gender or age emerged, and these demographic variables were therefore not included as control variables.

### 6.3 Impressions in judgment vs. feedback conditions (H1)

We hypothesized that feedback would be more positive than the judgments. Indeed, the main effect of evaluation condition was significant or marginally significant for a majority of dimensions, corroborating this hypothesis (see Table 1 for means and standard deviations). With regards to the scales, participants giving feedback tended to give higher grades in the feedback than the judgment condition ( $F(1, 126)=3.48$ ,  $p=.065$ , partial  $\eta^2=.027$ ). In the feedback condition, they rated students' writing skills as higher than in the judgment condition ( $F(1, 126)=7.30$ ,  $p=.008$ , partial  $\eta^2=.055$ ). When giving feedback, participants also rated the essay quality, both in content and style/mechanics, as higher than when providing a judgment ( $F(1, 124)_{\text{content}}=10.58$ ,  $p=.001$ , partial  $\eta^2=.079$ ;  $F(1, 124)_{\text{style/mechanics}}=11.62$ ,  $p=.001$ , partial  $\eta^2=.086$ ). Moreover, participants tended to agree more strongly that the student could be successful in higher education when giving feedback than when giving a judgment ( $F(1, 126)=3.20$ ,  $p=.076$ , partial  $\eta^2=.025$ ). In the open text format, participants also offered more suggestions for improvement in the feedback than the judgment condition ( $F(1, 130)=109.19$ ,  $p<.001$ , partial  $\eta^2=.457$ ) and wrote in a warmer and more considerate tone in their feedback than in their judgments ( $F(1, 130)=79.25$ ,  $p<.001$ , partial  $\eta^2=.379$ ). With regards to their comments, participants included more positive comments when giving feedback, both in the number of positive comments on content and total number of positive comments than when judging students' essays ( $F(1, 130)_{\text{comments content}}=9.46$ ,  $p=.003$ , partial  $\eta^2=.064$ , and  $F(1, 130)_{\text{total comments}}=6.87$ ,  $p=.010$ , partial  $\eta^2=.050$ , respectively). In tendency, participants also made less negative comments regarding content of the essay in the feedback condition than in the judgment condition ( $F(1, 128)=3.18$ ,  $p=.077$ , partial  $\eta^2=.024$ ). As a result, the percentage of positive comments was higher in the feedback condition than in the judgment condition as well ( $F(1, 130)=9.73$ ,  $p=.002$ , partial  $\eta^2=.070$ ). Unexpectedly, the number of positive and negative comments on style/mechanics as well as negative points in total did not differ by evaluation condition (see Table 1 for all results).

### 6.4 Impressions of essays by students with a Turkish migration background vs. none (H2a-b)

We expected that the effect of evaluation condition would be stronger for students with a supposed Turkish migration background than for students without a migration background (H2a), with the exception of those dimensions related to essay style/mechanics (H2b). Contrary to our expectations, the interaction term was not significant for all but one rating dimension, grade ( $F(1, 126)=2.91$ ,  $p=.091$ , partial  $\eta^2=.023$ ). However, this marginally significant effect was not

**Table 1** Descriptives statistics by experimental condition and results of the mixed ANOVAs

Dependent variable	ANOVA										Interaction term			
	Feedback condition			Judgment condition			Both evaluation conditions			Student's migration background				
	All students	No MB	Turkish MB	All students	No MB	Turkish MB	No MB	Turkish MB	Turkish MB	F		Partial $\eta^2$		
<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F</i>	Partial $\eta^2$	<i>F</i>	Partial $\eta^2$		
<b>Rating scale impressions</b>														
Grade	6.85 (1.89)	6.83 (1.99)	6.89 (2.16)	6.24 (1.82)	5.94 (2.06)	6.53 (2.03)	6.41 (2.07)	6.72 (2.10)	3.94 <sup>*</sup>	.030	3.48 <sup>+</sup>	.027	2.91 <sup>+</sup>	.023
Writing skills	3.63 (0.91)	3.59 (1.05)	3.71 (1.03)	3.20 (0.89)	3.10 (1.06)	3.32 (1.02)	3.36 (1.08)	3.53 (1.04)	3.29 <sup>+</sup>	.025	7.30 <sup>**</sup>	.055	0.67	.005
Content <sup>1</sup>	3.76 (0.88)	3.69 (0.83)	3.83 (1.16)	3.23 (0.94)	3.07 (1.08)	3.40 (1.12)	3.40 (1.00)	3.63 (1.16)	6.25 <sup>*</sup>	.048	10.58 <sup>**</sup>	.079	1.17	.009
Style/mechanics <sup>1</sup>	3.37 (0.80)	3.37 (1.01)	3.38 (0.97)	2.84 (0.95)	2.75 (1.06)	2.95 (1.10)	3.08 (1.07)	3.18 (1.05)	1.34	.011	11.62 <sup>**</sup>	.086	1.34	.011
Success higher education	3.61 (1.10)	3.56 (1.21)	3.68 (1.23)	3.28 (1.01)	3.21 (1.17)	3.37 (1.13)	3.40 (1.20)	3.54 (1.19)	2.27	.018	3.20 <sup>+</sup>	.025	0.18	.001
<b>Open text impressions</b>														
<b>Positive comments</b>														
Total	4.41 (2.76)	4.10 (2.92)	4.73 (3.36)	3.22 (2.43)	2.86 (2.46)	3.58 (2.91)	3.52 (2.78)	4.19 (3.19)	7.98 <sup>**</sup>	.058	6.87 <sup>*</sup>	.050	0.04	<.001
Content	3.55 (2.29)	3.25 (2.50)	3.85 (2.83)	2.39 (2.02)	2.09 (2.00)	2.73 (2.50)	2.71 (2.34)	3.30 (2.73)	7.83 <sup>**</sup>	.057	9.46 <sup>**</sup>	.068	<0.01	<.001
Style/mechanics	0.63 (0.84)	0.57 (1.02)	0.69 (1.06)	0.70 (0.79)	0.61 (0.90)	0.79 (1.07)	0.59 (0.97)	0.74 (1.06)	2.01	.015	0.23	.002	0.08	.001
<b>Negative comments</b>														
Total <sup>1</sup>	10.29 (3.83)	10.54 (4.09)	10.05 (4.57)	11.18 (3.37)	11.44 (3.99)	10.93 (4.18)	10.96 (4.05)	10.46 (4.40)	1.98	.015	1.99	.015	<0.01	<.001
Content <sup>1</sup>	5.34 (2.73)	5.70 (3.25)	4.98 (3.09)	6.16 (2.61)	6.32 (3.11)	6.01 (2.97)	5.99 (3.19)	5.46 (3.07)	3.53 <sup>+</sup>	.027	3.18 <sup>+</sup>	.024	0.59	.005
Style/mechanics <sup>1</sup>	4.86 (2.25)	4.75 (2.12)	4.98 (3.28)	4.91 (1.84)	4.96 (2.16)	4.86 (2.49)	4.85 (2.14)	4.92 (2.93)	0.07	.001	0.01	<.001	0.42	.004
Number of suggestions	4.75 (3.28)	4.61 (3.14)	4.92 (3.94)	0.31 (0.71)	0.38 (0.92)	0.23 (0.66)	2.62 (3.17)	2.70 (3.70)	0.18	.001	109.19 <sup>***</sup>	.458	1.52	.012
Tone	5.19 (1.10)	5.15 (1.46)	5.22 (1.26)	3.59 (0.95)	3.47 (1.08)	3.71 (1.05)	4.36 (1.54)	4.51 (1.39)	1.80	.014	79.25 <sup>***</sup>	.379	0.52	.004

Table 1 (continued)

Descriptive statistics		ANOVA												
		Feedback condition			Judgment condition		Both evaluation conditions		Student's migration background		Evaluation condition		Interaction term	
		All students	No MB	Turkish MB	All students	No MB	Turkish MB	No MB	Turkish MB	Student's migration background	Student's migration background	Student's migration background	Student's migration background	Student's migration background
Dependent variable	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F</i>	Partial $\eta^2$	<i>F</i>	Partial $\eta^2$	<i>F</i>	Partial $\eta^2$
Percent positive comments	0.30 (0.15)	0.28 (0.16)	0.32 (0.18)	0.22 (0.14)	0.20 (0.16)	0.24 (0.17)	0.24 (0.16)	0.28 (0.18)	6.93**	.051	9.73**	.070	<0.01	<0.001

Sample size varied between 128 and 132 due to item-level missingness

MB Migration background

<sup>†</sup>In this analysis, we controlled for writer order (student with a Turkish migration background first vs. second) and its interactions with the other factors because *t*-tests suggested that the difference between participants' judgments/feedback differed by writer order (see Sect. 6.2 Preliminary Analyses)

\*\*\**p* < .001, \*\**p* < .01, \**p* < .05, +*p* < .10



in the expected direction. In tendency, a positive bias on grades towards students with a Turkish migration background was only present in the judgment condition, not in the feedback condition. Based on these results, hypothesis 2a was not supported. We did find support for hypothesis 2b since there was no significant effect on the scale-based ratings of essay style/mechanics and the positive and negative comments regarding style/mechanics in the open text (see Table 1).

However, a number of main effects of student migration background emerged, signaling that participants provided more positive impressions related to overall assessment, e.g., grade and writing skills, and content-oriented comments to students with a migration background in both experimental conditions. Therefore, the absence of significant interaction effects was due the lack of a negative bias in the judgment condition. In both conditions, participants gave better grades to the student with a Turkish migration background than the student without a migration background ( $F(1, 126) = 3.94, p = .049, \text{partial } \eta^2 = .030$ ), and tended to rate the writing skills of the student with a Turkish migration background as better than of the student without a migration background ( $F(1, 126) = 3.29, p = .072, \text{partial } \eta^2 = .025$ ). Moreover, they included more positive comments about the essay by the student with the Turkish migration background than the essay by the student without a migration background ( $F(1, 130) = 7.98, p = .005, \text{partial } \eta^2 = .058$ ) and had a higher percentage of positive comments in the open text about the essay by the student with a Turkish migration background than in the text about the essay by the student without a migration background ( $F(1, 130) = 6.93, p = .009, \text{partial } \eta^2 = .051$ ). In their impressions of the content of the essay, participants also exhibited a positive bias in favor of the student with a Turkish name: They rated the quality of essay content as higher for the student with a Turkish migration background than for the student without a migration background ( $F(1, 124) = 6.25, p = .014, \text{partial } \eta^2 = .048$ ). Additionally, they gave more positive comments regarding essay content of the essay by the student with a Turkish migration than the essay by the student without a migration background ( $F(1, 130) = 7.83, p = .006, \text{partial } \eta^2 = .057$ ). Lastly, participants also tended to give less negative comments regarding content on the essay by the student with a Turkish migration background than by the student without a migration background ( $F(1, 128) = 3.53, p = .063, \text{partial } \eta^2 = .027$ ). Again, no differences were found for impressions on style/mechanics (see Table 1).

Overall, the results of the mixed ANOVAs indicated that feedback was more positive than judgments for all students and that students with a Turkish migration backgrounds received higher ratings regardless of evaluation condition, with the exception of the comments and ratings regarding style and mechanics of the essays.

**Table 2** Results of the Regression Predicting Difference Scores in the Feedback Condition by Self-Esteem Constructs

Dependent variable (difference scores)	Intercept	Self-esteem		Contingency of self-esteem		Adj. $R^2$
	<i>B</i> ( <i>SE</i> )	<i>B</i> ( <i>SE</i> )	$\beta$	<i>B</i> ( <i>SE</i> )	$\beta$	
Rating scale impressions						
Grade	0.05 (0.22)	-0.26 (0.53)	-0.07	-0.38 (0.37)	-0.16	.014
Writing skills	0.09 (0.12)	-0.19 (0.28)	-0.10	-0.03 (0.20)	-0.02	.023
Content <sup>1</sup>	0.34 <sup>+</sup> (0.18)	-0.26 (0.30)	-0.13	0.13 (0.21)	0.09	.029
Style/mechanics <sup>1</sup>	-0.38 <sup>+</sup> (0.18)	-0.11 (0.34)	-0.04	-0.08 (0.24)	-0.05	.066
Success in higher education	0.11 (0.13)	-0.30 (0.32)	-0.14	-0.09 (0.22)	-0.06	-.017
Open text impressions						
Positive comments						
Total	0.65 <sup>+</sup> (0.36)	-1.76* (0.87)	-0.29*	-0.77 (0.59)	-0.19	.030
Content	0.63* (0.31)	-2.23** (0.76)	-0.40**	-0.87 (0.52)	-0.23	.086
Style/mechanics	-0.02 (0.19)	0.58 (0.46)	0.18	0.21 (0.31)	0.10	-.006
Negative comments						
Total <sup>1</sup>	0.55 (0.66)	1.72 (1.15)	0.21	0.41 (0.78)	0.07	.076
Content <sup>1</sup>	-0.36 (0.55)	1.40 (0.95)	0.22	0.52 (0.64)	0.12	.009
Style/mechanics <sup>1</sup>	0.89 (0.54)	0.24 (0.94)	0.04	-0.12 (0.64)	-0.03	.005
Number of suggestions	0.31 (0.34)	-0.41 (0.82)	-0.07	-0.23 (0.56)	-0.06	-.026
Tone	0.08 (0.20)	-0.34 (0.48)	-0.10	-0.18 (0.33)	-0.08	-.022
Percent positive comments	0.04* (0.02)	-0.13* (0.05)	-0.36*	-0.07 <sup>+</sup> (0.03)	-0.28 <sup>+</sup>	.069

The dependent variables are difference scores (value student with Turkish migration background—value of student without a migration background). Values greater than zero indicate a positive bias towards students with a Turkish migration background and values smaller than zero a negative bias, with the exception of the negative comments, where the interpretation is reversed. Sample size varied between 67 and 70 due to item-level missingness

<sup>1</sup>In this analysis, we controlled for writer order (reference group: student with a Turkish migration background first) because *t*-tests suggested that the difference between participants' judgments/feedback differed by writer order (see Sect. 6.2 Preliminary Analyses)

\*\* $p < .01$ , \* $p < .05$ , <sup>+</sup> $p < .10$

### 6.5 Moderation effects in the feedback condition by self-esteem, contingency of self-esteem, and motivation to respond without prejudice (H3a-c)

We expected that the size of the positive bias in the feedback condition would vary with participants' characteristics and motivations. We hypothesized that lower self-esteem (H3a) and greater contingency of self-esteem (H3b) would be related to greater positive bias towards students with a Turkish migration background. Moreover, we expected that those low in internal and simultaneously high in external motivation to respond without prejudice would show a stronger positive bias (H3c). The

regressions used to test these hypotheses predicted difference scores, with the values of the student without a migration background being subtracted from the values of the student with a Turkish migration background, and thus test whether the extent of a positive or negative bias was influenced by the moderators. As expected, lower self-esteem was associated with greater positive bias towards students with a Turkish migration background (see Table 2 for all results). However, this effect was limited to dimensions indicating the general positivity of the comments in the open text, i.e., participants with a lower self-esteem differentiated more between the student with a Turkish migration background and the student without one regarding positive comments overall ( $b = -1.76$ ,  $SE = 0.87$ ,  $p = .048$ ) and positive comments related to essay content ( $b = -2.23$ ,  $SE = 0.76$ ,  $p = .005$ ). As a result, those with lower self-esteem also had a greater difference in the percentage of positive comments given to a student with a Turkish migration background and a student without a migration background ( $b = -0.13$ ,  $SE = 0.05$ ,  $p = .012$ ). Thus, hypothesis 3a was partially supported. Contrary to expectations, greater contingency of self-esteem on others' approval was not related to greater positivity of the feedback given to students with a Turkish-origin name (H3b). However, one unexpected marginally significant effect emerged: Greater contingency of self-esteem was associated with a lower percentage of positive comments for students with a Turkish compared to no migration background ( $b = -0.07$ ,  $SE = 0.03$ ,  $p = .053$ ), i.e., a more negative feedback given to them as compared to their peer without a migration background.

Regarding the effects for participants' motivation to respond without prejudice, results were mixed and provided only partial support for hypothesis 3c. Contrary to our hypothesis, those participants who were low in internal motivation and simultaneously high in external motivation did not consistently show a stronger positive bias. In tendency, when external motivation to respond without prejudice was low, internal motivation tended to have only small or no effects on participants' positive bias regarding grade ( $b_{\text{interaction}} = -0.28$ ,  $SE = 0.16$ ,  $p = .088$ ), essay style/mechanics ( $b_{\text{interaction}} = -0.18$ ,  $SE = 0.11$ ,  $p = .088$ ), total number of positive comments ( $b_{\text{interaction}} = -0.53$ ,  $SE = 0.28$ ,  $p = .063$ ), and number of positive comments on content ( $b_{\text{interaction}} = -0.44$ ,  $SE = 0.25$ ,  $p = .087$ ). Contrarily, when external motivation was high, greater internal motivation was associated with less positive bias. Thus, participants tended to show the greatest positive bias towards students with a Turkish migration background on these dimensions when they were purely motivated to do so by external reasons, i.e., when they were low in internal and simultaneously high in external motivation. There also emerged different main effects in these analyses, which are reported in the supplementary material (Online Resource 1).

## 7 Discussion

In the present research, we examined two phenomena regarding the feedback process: First, we examined whether feedback would generally be more positive than a judgment that was not communicated to the student (feedback inflation). Secondly, we tested whether the difference between feedback and non-communicated judgments would be greater for negatively stereotyped ethnic minority students,

specifically students with a Turkish migration background in Germany (positive feedback bias). To our knowledge, our study is the first to directly compare feedback and judgment conditions in a sample of aspiring teachers, making it an important addition to educational research. Moreover, we aimed to both replicate negative ethnic biases in non-communicated judgments and test whether the positive feedback bias would emerge in the German context for the first time. Lastly, we examined whether personal characteristics—in part not examined in prior literature—might be associated with the extent of a positive feedback bias.

We found that teacher students were decidedly more positive when they gave feedback to the students than when they handed in non-communicated judgments to the research team. Unexpectedly, teacher students did not exhibit a negative ethnic bias in their judgments but rather a positive bias in favor of students with a Turkish migration background both in judgments and feedback. This was the case on both rating scales similar to scales used in previous research (e.g., Sprietsma, 2013) as well as open-text feedback, which had not been examined in research on negative ethnic biases up to now. On some dimensions of feedback, this positive bias was somewhat greater for those with low self-esteem and, in tendency, those low in internal motivation to respond without prejudice, but mainly when they were highly externally motivated.

## 7.1 Feedback vs. non-communicated judgments

When teacher students provided feedback, they conveyed a more positive impression of the students' work than when they handed in their judgments to the research team. Feedback contained more positive ratings on the scales reflecting overall assessments and open comments included more positive comments and suggestions for further improvement. Overall, the tone was warmer and more considerate. These findings suggest that teacher students adapt their initial judgments to make feedback more palatable. Our results indicate that teacher students employ a strategy of increasing positive comments in concert with a somewhat attenuated strategy to reduce criticism. This corresponds to prior studies' finding that people include positive aspects in feedback (e.g., Ginsburg et al., 2016), and are reluctant to communicate negative information (Bergsieker et al., 2012; Qadan et al., 2013).

## 7.2 Positive bias towards students with a Turkish migration background

Teacher students included more positive comments and, in tendency, less negative comments on essays of students with a Turkish migration background than students without a migration background. Moreover, they rated their essays as having better content and deserving a better grade and also tended to communicate higher opinions of the writing skills of students with a Turkish migration background. Unexpectedly, this was the case both with regards to judgments and feedback. Unlike research conducted in Canada (Croft & Schmader, 2012), participants' feedback was more positive overall mainly due to an increase in positive feedback and only somewhat due to a decrease the amount of negative feedback they provided to the student.

Overall, the results do support the reasoning that a positive bias arises from concerns to be or appear to be prejudiced, though the presence of a positive bias in the judgment group challenge this interpretation.

In the feedback condition, the results found by Harber and colleagues (Croft & Schmader, 2012; Harber, 1998, 2004; Harber et al., 2010, 2012, 2019) were replicated with regards to German teacher students without a migration background addressing students with a Turkish migration background. While teacher students gave more positive feedback on overall assessments (e.g., grade) and in their comments to students with a Turkish migration background, they did not give different feedback to students with and without a migration background regarding the style and mechanics of the essay. The chief difference between overall assessments as well as impressions of essay content and essay style/mechanics is the degree to which there are objective criteria available to judge their quality. Negative feedback regarding content and overall assessments is more ambiguous and could be more easily interpreted as biased. As a result, the typical pattern in studies on the positive feedback bias emerges: a positive bias on overall assessment and content-related dimensions in favor of students belonging to a negatively stereotyped group and no bias on mechanics-related dimensions.

The positive bias in the judgment condition, however, calls into question the argument that people may give more positive feedback to negatively stereotyped people out of a concern to be or appear to be prejudiced, since this effect should be present only when interacting with another person. Moreover, this result is at odds with a host of studies conducted in Germany regarding negative biases in judgments on scales that are highly similar to the ones used in this research (e.g., Bonefeld & Dickhäuser, 2018; Sprietsma, 2013), but also judging other constructs (e.g., Glock et al., 2013). In recent years, researchers in Germany have worked on better understanding the contexts under which such negative biases occur. Some studies find negative biases only in counter-stereotypic scenarios, i.e., when an ethnic minority student is portrayed as performing well (Wenz & Hoenig, 2020). Other studies suggest the opposite: Stereotypes are most likely to be applied in stereotype-confirming scenarios (Glock, 2016; Glock & Krolak-Schwerdt, 2013). The sole study that also found a positive bias in judgments towards a student with a Turkish migration background found this positive bias when the student was performing well, but when the student was performing poorly, as the student did in the present study, a negative bias emerged (Kleen & Glock, 2018). Overall, prior research suggests that the conditions under which negative biases emerge are varied and complex (Glock, 2016; Glock & Krolak-Schwerdt, 2013; Glock et al., 2013; Holder & Kessels, 2017; Wenz & Hoenig, 2020).

The most parsimonious explanation for the discrepancy between our results and prior studies may be that the participants in the judgment condition may also have adjusted their responses. For example, they may already be aware of the discourse on possible prejudice and discrimination against students with a Turkish migration background in school, or specifically writing the name of the student on the judgment form made the students migration background more salient. Thus, even participants in the judgment condition may have been motivated to demonstrate to themselves and/or the research team that they were unprejudiced and subsequently

overcorrected. Another potential explanation could be that some earlier research not finding evidence of a negative ethnic bias may be less likely to be published, resulting in a publication bias.

Alternatively, the *shifting standards theory* (Biernat & Manis, 1994; Biernat et al., 1991) could explain why a positive bias was present across evaluation conditions. This theory posits that peoples' characteristics or performance are judged relative to the expectations for a group to which they belong, which can be informed by stereotypes. The way in which judgments are assessed influences whether this group-specific shift in judgment can be expressed. If subjective scales are used, with endpoints such as "very bad" to "very good," the same endpoint can acquire different meaning for different groups—in accordance with what performance is considered "very bad" and "very good" for a member of that group ("That's very good for a student with a migration background!"). Holder and Kessels (2017) demonstrated such shifting standards in German teacher students judging students with a Turkish migration background. Since we used subjective rating scales, students with a Turkish migration background may have received more positive ratings simply because their performance is better relative to the negative stereotypes about their group (Biernat & Manis, 1994). This process would be unaffected by interactions, thus explaining why the evaluation conditions did not differ. However, all of our scales apart from the grade were subjective scales, including those related to essay mechanics. The lack of a positive bias on subjective scales related to essay mechanics speaks partly against this interpretation of our findings. Thus, our results are partially consistent with both the shifting standards and the positive feedback bias perspectives.

### **7.3 Moderating effect: Do self-concept or self-presentation concerns fuel the positive feedback bias?**

Our last aim in this research was to establish whether self-esteem, contingency of self-esteem on others' approval, and motivation to respond without prejudice were associated with the degree of positive feedback bias. This also gave us a first impression as to whether concerns about being prejudiced or appearing to be prejudiced are more central to the positive feedback bias, and we discuss our findings with regard to these competing explanations. The former perspective rests on the explanation that people are motivated to maintain a positive view of themselves (self-concept), whereas the latter posits a motivation to maintain a good reputation in the eyes of others (self-presentation; Harber et al., 2010; Plant & Devine, 1998). Our results, which established a main effect of self-esteem on three feedback dimensions, could support either a self-concept or self-presentational perspective. This main effect of self-esteem may reflect that more insecure people may use increased positivity to either convince themselves or others that they are unprejudiced, while people who are assured of themselves are both less likely to be concerned about being prejudiced and about self-presentation.

Tentative support for the self-presentational perspective can be seen in the results regarding internal and external motivation to respond without prejudice. These

results were similar to research by Croft and Schmader (2012), showing that those purely motivated by self-presentation tended to exhibit the strongest positive feedback bias. Based on the self-presentational perspective, however, we would have expected those whose self-esteem was more contingent on others' approval to be more likely to show positive bias as well. However, this is not what we found—the absence of an effect here speaks more to the relevance of one's self-concept than of one's reputation. Based on our results, people who are especially motivated to be approved of as unprejudiced exhibit a greater positive bias, but those seeking approval more generally do not. Overall, our results regarding moderation effects in the German context do not paint a clear picture as to whether positive bias towards students with a Turkish migration background is a result of self-concept or self-presentation concerns. Future experimental manipulations may be able to better differentiate between these possible motivations.

#### 7.4 Implications for the school context

Adapting feedback to take interpersonal and motivational aspects into account is not necessarily a detriment—it may well protect the relationship between a teacher and their student and balance out feedback that might otherwise be devastating (Kerssen-Griep et al., 2003, 2008). However, overly positive feedback may have negative effects on students' learning and motivation both when students accept it at face value and when they question it. When students believe that they have already understood the material well enough, they may develop unrealistic assessments of their understanding and may be less motivated to extend more effort. These unrealistic self-assessments could also relate to the attainment-aspiration gap among ethnic minority students in Germany, i.e., the finding that they have higher educational expectations than may be warranted by their academic performance (Becker & Gresch, 2016; McElvany et al., 2018). Overly positive feedback may, over time, encourage an overly optimistic self-concept and greater self-efficacy in these students—which, in turn, relate to higher aspirations (McElvany et al., 2018). While positive self-concepts in themselves have been linked to better performance (Eckert et al., 2006), the role of an overly optimistic self-concept is more ambiguous as it has been linked to both negative and positive consequences for motivation and performance (Butler, 2011; Kim et al., 2010; Praetorius et al., 2016).

Other students may question the accuracy and reliability of overly positive feedback, which undermines the credibility of the teacher and their feedback (McCroskey & Teven, 1999). Credibility is a precursor to the acceptance of feedback and students may reject feedback that lacks credibility as a basis for future learning (Finn et al., 2009; van de Ridder et al., 2015). Ethnic minority students may be particularly at risk for these negative effects of overly positive feedback since they are especially likely to question the credibility of teachers' feedback, both positive and negative (Biernat & Danaher, 2012; Crocker et al., 1991; Hoyt et al., 2007; Major et al., 2016). Receiving and discounting positive feedback is also detrimental to US-American ethnic minority students' self-esteem and their subjective and physiological stress response, especially for those sensitive to race-based rejection (Crocker



et al., 1991; Major et al., 2016; Mendes et al., 2008; Mendoza-Denton et al., 2010). Whether similar effects may occur in the German context should be examined in future research focusing on students' perception of feedback.

Overall, while adapting one's feedback to protect a student's self-efficacy and motivation may be necessary and beneficial, teachers need to be mindful in how they deliver feedback so that it remains authentic and clearly conveys which aspects need improving. In the US, researchers have identified concrete solutions that may help deliver adequate feedback: Teachers are advised to avoid giving overly positive feedback to students, especially ethnic minority students, and to incorporate other messages to reduce the ambiguity that comes with criticizing them instead (*wise feedback*; Cohen & Steele, 2002; Yeager et al., 2013). This approach could also reduce the teacher's personal concern that they might be perceived as prejudiced if they criticize negatively stereotyped students, reducing the urge to provide overly positive feedback in the first place.

## 7.5 Limitations and future research

Some limitations in the design of the present research should be taken into account when considering the results. First and foremost, we used a design in which teacher students gave feedback in written form. While this is a medium through which students receive feedback, there are many less formal ways in which teachers provide feedback, e.g., through comments in class or short face-to-face discussions. In personal conversation, feedback may actually be more positively inflated (Waung & Highhouse, 1997), and interpersonal concerns—including the concern to be or appear prejudiced—may come to the forefront more strongly. Future research should establish whether the positive bias towards students with a Turkish migration background is comparable in face-to-face interactions, as has been found in the US (Harber, 2004), or even greater. Similarly, all experimental research must consider the question of ecological validity. While experiments allow us to isolate potential causes of positive bias, teachers in classrooms have much more information about students than their name and essay. Moreover, they repeatedly interact with students—and give feedback at various times and in various forms. Because teachers must maintain a relationship with students—unlike subjects in our study—interpersonal concerns may be even more relevant to their feedback. However, as teachers are more familiar with their students and already have an established relationship, they may be less influenced by concerns that they may be or appear to be prejudiced. Thus, interactions between teachers and students familiar with one another need to be studied alongside effects in controlled environments. In addition to observational research with teachers, experiments with experienced teachers rather than teacher students may improve our understanding whether concerns about being or appearing to be prejudiced plague only inexperienced teachers.

Secondly, we varied the migration background of the target student within subjects, which has both advantages and disadvantages (Charness et al., 2012). Using a within-subjects design on this factor allowed us to account for individual differences regarding the effect of migration background, increasing power. Moreover, in this

case a within-subjects design is more closely approximating the real-life situation of giving feedback, in which teachers do not consider students' work in isolation, but one after the other. However, only one other study regarding the positive feedback bias used a within-subjects design (Croft & Schmader, 2012), while the other studies varied target student race/ethnicity between subjects. Therefore, our results are not directly comparable to these studies. Additionally, the within-subjects design might have introduced a consistency motive, such that participants aimed to judge/give feedback based on similar criteria and in a similar way. If this was the case, however, the differences by migration background might be considered a conservative test, as this would have worked to reduce any differences in judgments or feedback.

Thirdly, research on the positive feedback bias towards ethnic minority students has focused on stereotype-confirming situations, that is, feedback regarding low-quality essays. This was necessary to create a situation that would unambiguously require participants to critique the author to some extent. However, research on negative biases in judgments has found that whether a situation confirms or refutes stereotypes is impactful (Glock, 2016; Glock et al., 2013; Wenz & Hoenig, 2020). Future research may determine whether mediocre or good work is also subject to a positive feedback bias. Moreover, we kept the essay topic constant across the essays—future research could also examine possible topic effects by testing whether other, more controversial topics may lead to different results (e.g., race-related topics may lead to greater concern to be or appear to be prejudiced). Additionally, it might be of interest to test systematically whether the order of writer's ethnicity (e.g., student with a German or a Turkish name first) may affect feedback and evaluations (e.g., Dutton, 1976).

Lastly, our approach was not able to detect ways in which intersections with other identities might exacerbate or attenuate negative and positive biases towards ethnic minority students. The authors of both essays were supposedly boys, and no additional information about membership in other stereotype-relevant groups was provided (e.g., social class). Research on both feedback and judgments has established differential effects by gender (Bonefeld et al., 2020; Jampol & Zayas, 2020; Kleen & Glock, 2018). Thus, how information about different intersecting group memberships is integrated and influences biases is an important avenue for future research.

## 7.6 Conclusion

When it comes to feedback, motivational and interpersonal concerns influence how positive and negative aspects are conveyed (e.g., Lupoli et al., 2017). The tension between giving caring as well as accurate feedback reflects the multiple goals involved in giving feedback and demonstrates that feedback is more than simply a transmission of information. Simultaneously, research established that ethnic minority students—relative to their ethnic majority peers—are judged more negatively (e.g., Bonefeld & Dickhäuser, 2018), but given more positive feedback (e.g. Harber, 2004). Our research adds to this knowledge by directly comparing judgments of and feedback to German students without a migration background and with a Turkish one. We demonstrated that teacher students inflate feedback for all students by

including more positive content and, in tendency, less negative content. Moreover, teacher students showed a consistent positive bias in favor of students with a Turkish migration background in giving them more positive judgments and feedback.

Overly positive feedback could be considered a mixed blessing: While it may have motivating effects and support a more positive self-concept, a more realistic assessment of one's understanding could be beneficial. Moreover, if the intentions behind positive feedback are questioned, it may easily be discounted and the teacher-student relationship damaged. Thus, it is imperative that teachers and teacher students understand the powerful motivational consequences their feedback has and learn to balance motivational concerns with accuracy. When interacting with negatively stereotyped students, teachers may need to be particularly mindful not to let personal concerns become more influential than their concern for the students' well-being and learning.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11218-021-09674-3>.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This research was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG) allocated to the second author (KE 1412/4–1).

**Data and material availability** Data, material and code are available upon request.

## Declarations

**Conflicts of interest** The authors have no relevant financial or non-financial interests to disclose.

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

## References

- American Psychological Association (2020). *APA Dictionary of Psychology: Stereotype*. APA. <https://dictionary.apa.org/stereotype>
- Anderson-Clark, T. N., Green, R. J., & Henley, T. B. (2008). The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology, 27*(1), 94–99. <https://doi.org/10.1177/0261927X07309514>
- Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis)perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of Educational Psychology, 107*(3), 805–820. <https://doi.org/10.1037/edu0000008>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>

- Becker, B., & Gresch, C. (2016). Bildungsaspirationen in Familien mit Migrationshintergrund [Educational aspirations in families with a migration background]. In C. Diehl (Ed.), *Ethnische Ungleichheiten im Bildungsverlauf* (pp. 73–115). Springer Fachmedien. [https://doi.org/10.1007/978-3-658-04322-3\\_3](https://doi.org/10.1007/978-3-658-04322-3_3)
- Bentley-Edwards, K. L., Stevenson, H. C., Thomas, D. E., Adams-Bass, V. N., & Coleman-King, C. (2020). Teaching scared: Pre-service teacher appraisals of racial stress, socialization and classroom management self-efficacy. *Social Psychology of Education, 23*(5), 1233–1257. <https://doi.org/10.1007/s11218-020-09578-8>
- Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology, 102*(6), 1214–1238. <https://doi.org/10.1037/a0027717>
- Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals in interracial interactions. *Journal of Personality and Social Psychology, 99*(2), 248–264. <https://doi.org/10.1037/a0018474>
- Biernat, M., & Danaher, K. (2012). Interpreting and reacting to feedback in stereotype-relevant performance domains. *Journal of Experimental Social Psychology, 48*(1), 271–276. <https://doi.org/10.1016/j.jesp.2011.08.010>
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66*(1), 5–20. <https://doi.org/10.1037/0022-3514.66.1.5>
- Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgement. *Journal of Personality and Social Psychology, 60*(4), 485–499. <https://doi.org/10.1037/0022-3514.60.4.485>
- 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 [Döner vs. pig roast—stereotypes of teachers (in training) about people of German and Turkish origin in comparison]. In S. Glock & H. Kleen (Eds.), *Stereotype in der schule* (pp. 159–190). Springer. [https://doi.org/10.1007/978-3-658-27275-3\\_6](https://doi.org/10.1007/978-3-658-27275-3_6)
- Brown, P. S., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Butler, R. (1994). Teacher communications and student interpretations: Effects of teacher responses to failing students on attributional inferences in two age groups. *British Journal of Educational Psychology, 64*(2), 277–294. <https://doi.org/10.1111/j.2044-8279.1994.tb01102.x>
- Butler, R. (2011). Are positive illusions about academic competence always adaptive, under all circumstances: New results and future directions. *International Journal of Educational Research, 50*(4), 251–256. <https://doi.org/10.1016/j.ijer.2011.08.006>
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subjects and within-subjects design. *Journal of Economic Behavior & Organization, 81*(1), 1–8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- Cohen, G. L., & Steele, C. M. (2002). A barrier of mistrust: How negative stereotypes affect cross-race mentoring. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 303–327). Academic Press. <https://doi.org/10.1016/B978-012064455-1/50018-X>
- Colletti, L. M. (2000). Difficulty with negative feedback: Face-to-face evaluation of junior medical student clinical performance results in grade inflation. *Journal of Surgical Research, 90*(1), 82–87. <https://doi.org/10.1006/jsre.2000.5848>
- Cox, S. S., Marler, L. E., Simmering, M. J., & Totten, J. W. (2011). Giving feedback: Development of scales for the mum effect, discomfort giving feedback, and feedback medium preference. *Performance Improvement Quarterly, 23*(4), 49–69. <https://doi.org/10.1002/piq.20094>
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359–378. <https://doi.org/10.1037/0022-3514.82.3.359>
- Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003). Contingencies of self-worth in college students: Theory and measurement. *Journal of Personality and Social Psychology, 85*(5), 894–908. <https://doi.org/10.1037/0022-3514.85.5.894>

- Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, *60*(2), 218–228. <https://doi.org/10.1037/0022-3514.60.2.218>
- Croft, A., & Schmader, T. (2012). The feedback withholding bias: Minority students do not receive critical feedback from evaluators concerned about appearing racist. *Journal of Experimental Social Psychology*, *48*(5), 1139–1144. <https://doi.org/10.1016/j.jesp.2012.04.010>
- Crosby, J. R., & Monin, B. (2007). Failure to warn: How student race affects warnings of potential academic difficulty. *Journal of Experimental Social Psychology*, *43*(4), 663–670. <https://doi.org/10.1016/j.jesp.2006.06.007>
- Dutton, D. G. (1976). Tokenism, reverse discrimination, and egalitarianism in interracial behavior. *Journal of Social Issues*, *32*(2), 93–107.
- Eckert, C., Schilling, D., & Stiensmeier-Pelster, J. (2006). Einfluss des Fähigkeitsselbstkonzepts auf die Intelligenz- und Konzentrationsleistung [The influence of academic self-concept on performance in intelligence and concentration tests]. *Zeitschrift Für Pädagogische Psychologie*, *20*(1/2), 41–48. <https://doi.org/10.1024/1010-0652.20.1.41>
- Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, *36*(1), 51–62. <https://doi.org/10.1080/02602930903197883>
- Ferring, D., & Filipp, S. H. (1996). Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala [Measurement of self-esteem: Findings on reliability, validity, and stability of the Rosenberg Scale]. *Diagnostica*, *42*(3), 284–292.
- Finn, A. N., Schrod, P., Witt, P. L., Elledge, N., Jernberg, K. A., & Larson, L. M. (2009). A meta-analytical review of teacher credibility and its associations with teacher behaviors and student outcomes. *Communication Education*, *58*(4), 516–537. <https://doi.org/10.1080/03634520903131154>
- Froehlich, L., Martiny, S. E., Deaux, K., & Mok, S. Y. (2016). “It’s their responsibility, not ours”: Stereotypes about competence and causal attributions for immigrants’ academic underperformance. *Social Psychology*, *47*(2), 74–86. <https://doi.org/10.1027/1864-9335/a000260>
- Froehlich, L., & Schulte, I. (2019). Warmth and competence stereotypes about immigrant groups in Germany. *PLoS ONE*, *14*(9), e0223103. <https://doi.org/10.1371/journal.pone.0223103>
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Academic Press.
- Ginsburg, S., van der Vleuten, C., Eva, K. W., & Lingard, L. (2016). Hedging to save face: A linguistic analysis of written comments on in-training evaluation reports. *Advances in Health Sciences Education*, *21*(1), 175–188. <https://doi.org/10.1007/s10459-015-9622-0>
- 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., Kleen, H., Krischler, M., & Pit-ten Cate, I. (2020). Die Einstellungen von Lehrpersonen gegenüber Schüler\*innen ethnischer Minoritäten und Schüler\*innen mit sonderpädagogischem Förderbedarf: Ein Forschungsüberblick [Attitudes of teachers towards ethnic minority students and students with special educational needs: A review of the research]. In S. Glock & H. Kleen (Eds.), *Stereotype in der Schule* (pp. 225–279). Springer.
- Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers’ judgments. *Social Psychology of Education*, *16*(1), 111–127. <https://doi.org/10.1007/s11218-012-9197-z>
- Glock, S., & Krolak-Schwerdt, S. (2014). Stereotype activation versus application: How teachers process and judge information about students from ethnic minorities and with low socioeconomic background. *Social Psychology of Education*, *17*(4), 589–607. <https://doi.org/10.1007/s11218-014-9266-6>
- Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (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>
- Hachfeld, A., Hahn, A., Schroeder, S., Anders, Y., Stanat, P., & Kunter, M. (2011). Assessing teachers’ cultural and egalitarian beliefs: The teacher cultural beliefs scale. *Teaching and Teacher Education*, *27*(6), 986–996. <https://doi.org/10.1016/j.tate.2011.04.006>
- Harber, K. D. (1998). Feedback to minorities: Evidence of a positive bias. *Journal of Personality and Social Psychology*, *74*(3), 622–628. <https://doi.org/10.1037/0022-3514.74.3.622>

- Harber, K. D. (2004). The positive feedback bias as a response to out-group unfriendliness. *Journal of Applied Social Psychology*, 34(11), 2272–2297. <https://doi.org/10.1111/j.1559-1816.2004.tb01977.x>
- Harber, K. D., Gorman, J. L., Gengaro, F. P., Butisingh, S., Tsang, W., & Ouellette, R. (2012). Students' race and teachers' social support affect the positive feedback bias in public schools. *Journal of Educational Psychology*, 104(4), 1149–1161. <https://doi.org/10.1037/a0028110>
- Harber, K. D., Reeves, S., Gorman, J. L., Williams, C. H., Malin, J., & Pennebaker, J. W. (2019). The conflicted language of interracial feedback. *Journal of Educational Psychology*, 111(7), 1220–1242. <https://doi.org/10.1037/edu0000326>
- Harber, K. D., Stafford, R., & Kennedy, K. A. (2010). The positive feedback bias as a response to self-image threat. *British Journal of Social Psychology*, 49(1), 207–218. <https://doi.org/10.1348/01446609X473956>
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- 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>
- 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>
- Hoyt, C. L., Aguilar, L., Kaiser, C. R., Blascovich, J., & Lee, K. (2007). The self-protective and undermining effects of attributional ambiguity. *Journal of Experimental Social Psychology*, 43(6), 884–893. <https://doi.org/10.1016/j.jesp.2006.10.013>
- Ilies, R., Judge, T. A., & Wagner, D. T. (2010). The influence of cognitive and affective reactions to feedback on subsequent goals. *European Psychologist*, 15(2), 121–131. <https://doi.org/10.1027/1016-9040/a0000011>
- Jampol, L., & Zayas, V. (2020). Gendered white lies: Women are given inflated performance feedback compared to men. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/0146167220916622>
- Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (pp. 231–262). Lawrence Erlbaum Associates.
- Kahraman, B., & Knoblich, G. (2000). «Stechen statt Sprechen»: Valenz und Aktivierbarkeit von Stereotypen über Türken [“Stabbing, not speaking”]: Valence and ease of activation of stereotypes about Turks]. *Zeitschrift Für Sozialpsychologie*, 31(1), 31–43. <https://doi.org/10.1024/0044-3514.31.1.31>
- Kerssen-Griep, J. (2001). Teacher communication activities relevant to student motivation: Classroom facework and instructional communication competence. *Communication Education*, 50(3), 256–273. <https://doi.org/10.1080/03634520109379252>
- Kerssen-Griep, J., Hess, J. A., & Trees, A. R. (2003). Sustaining the desire to learn: Dimensions of perceived instructional facework related to student involvement and motivation to learn. *Western Journal of Communication*, 67(4), 357–381. <https://doi.org/10.1080/10570310309374779>
- Kerssen-Griep, J., Trees, A. R., & Hess, J. A. (2008). Attentive facework during instructional feedback: Key to perceiving mentorship and an optimal learning environment. *Communication Education*, 57(3), 312–332. <https://doi.org/10.1080/03634520802027347>
- Kim, Y. H., Chiu, C. Y., & Zou, Z. (2010). Know thyself: Misperceptions of actual performance undermine achievement motivation, future performance, and subjective well-being. *Journal of Personality and Social Psychology*, 99(3), 395–409. <https://doi.org/10.1037/a0020555>
- Kleen, H., & Glock, S. (2018). A further look into ethnicity: The impact of stereotypical expectations on teachers' judgments of female ethnic minority students. *Social Psychology of Education*, 21(4), 759–773. <https://doi.org/10.1007/s11218-018-9451-0>
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107(1), 34–47. <https://doi.org/10.1037/0033-2909.107.1.34>



- Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology*, 68(3), 518–530. <https://doi.org/10.1037/0022-3514.68.3.518>
- Littleford, L. N., Wright, M. O. D., & Sayoc-Parial, M. (2005). White students' intergroup anxiety during same-race and interracial interactions: A multimethod approach. *Basic and Applied Social Psychology*, 27(1), 85–94. [https://doi.org/10.1207/s15324834basp2701\\_9](https://doi.org/10.1207/s15324834basp2701_9)
- 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]. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, 68(1), 89–111. <https://doi.org/10.1007/s11577-015-0352-3>
- Lupoli, M. J., Jampol, L., & Oveis, C. (2017). Lying because we care: Compassion increases prosocial lying. *Journal of Experimental Psychology: General*, 146(7), 1026–1042. <https://doi.org/10.1037/xge0000315>
- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51, 93–120. <https://doi.org/10.1146/annurev.psych.51.1.93>
- Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, 66(1), 37–47. <https://doi.org/10.1037/0022-3514.66.1.37>
- Major, B., Kunstman, J. W., Malta, B. D., Sawyer, P. J., Townsend, S. S., & Mendes, W. B. (2016). Suspicion of motives predicts minorities' responses to positive feedback in interracial interactions. *Journal of Experimental Social Psychology*, 62, 75–88. <https://doi.org/10.1016/j.jesp.2015.10.007>
- Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education*, 60(3), 245–256. <https://doi.org/10.1177/0004944116664618>
- Marshall, P. L. (1996). Multicultural teacher concerns: New dimensions in the area of teacher concerns research? *The Journal of Educational Research*, 89(6), 371–379. <https://doi.org/10.1080/00220671.1996.9941341>
- McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communications Monographs*, 66(1), 90–103. <https://doi.org/10.1080/03637759909376464>
- McElvany, N., Ferdinand, H. D., Gebauer, M. M., Bos, W., Huelmann, T., Köller, O., & Schöber, C. (2018). Attainment-aspiration gap in students with a migration background: The role of self-efficacy. *Learning and Individual Differences*, 65, 159–166. <https://doi.org/10.1016/j.lindif.2018.05.002>
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, 46(3), 235–261. <https://doi.org/10.1016/j.jsp.2007.05.001>
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>
- Mendes, W. B., & Koslov, K. (2013). Brittle smiles: Positive biases toward stigmatized and outgroup targets. *Journal of Experimental Psychology: General*, 142(3), 923–933. <https://doi.org/10.1037/a0029663>
- Mendes, W. B., Major, B., McCoy, S., & Blascovich, J. (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *Journal of Personality and Social Psychology*, 94(2), 278–291. <https://doi.org/10.1037/0022-3514.94.2.278>
- Mendoza-Denton, R., Goldman-Flythe, M., Pietrzak, J., Downey, G., & Aceves, M. J. (2010). Group-value ambiguity: Understanding the effects of academic feedback on minority students' self-esteem. *Social Psychological and Personality Science*, 1(2), 127–135. <https://doi.org/10.1177/1948550609357796>
- Müller, A., & Ditton, H. (2014). Feedback Begriff, Formen und Funktionen feedback concept, form and functions. In H. Ditton & A. Müller (Eds.), *Feedback und Rückmeldungen Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 11–28). Waxmann Verlag.
- Parks, F. R., & Kennedy, J. H. (2007). The impact of race, physical attractiveness, and gender on education majors' and teachers' perceptions of student competence. *Journal of Black Studies*, 37(6), 936–943. <https://doi.org/10.1177/0021934705285955>
- Pendry, L. (2014). Soziale Kognition Social cognition. In K. Jonas, W. Stroebe, & M. Hewstone (Eds.), *Sozialpsychologie* (6th ed., pp. 107–140). Springer-Verlag.



- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811–832. <https://doi.org/10.1037/0022-3514.75.3.811>
- Praetorius, A. K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbstschätzungen die Nase vorn? Zusammenhänge zwischen optimistischen, realistischen und pessimistischen Selbstkonzepten und der Leistungsentwicklung von Grundschulkindern [Are students with optimistic self-concepts one step ahead? Relations between optimistic, realistic and pessimistic self-concepts and the achievement development of primary school children]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, 48(1), 14–26. <https://doi.org/10.1026/0049-8637/a000140>
- Qadan, L., Al-Ozairi, E., Ayed, A., & Huang, G. (2013). Avoiding honest feedback: Discordance between formal evaluations and candid assessments of Kuwaiti PBL students. *Medical Teacher*, 35(6), 459–464. <https://doi.org/10.3109/0142159X.2013.774337>
- Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3), 375–392. <https://doi.org/10.3102/0162373720932188>
- Ramani, S., Könings, K. D., Mann, K. V., Pisarski, E. E., & van der Vleuten, C. P. (2018). About politeness, face, and feedback: Exploring resident and faculty perceptions of how institutional feedback culture influences feedback practices. *Academic Medicine*, 93(9), 1348–1358. <https://doi.org/10.1097/ACM.0000000000002193>
- Rosen, S., & Tesser, A. (1970). On reluctance to communicate undesirable information: The MUM effect. *Sociometry*, 33(3), 253–263. <https://doi.org/10.2307/2786156>
- Rosenberg, M. (1979). *Conceiving the self*. Basic Books.
- Schwinger, M., Schöne, C., & Otterpohl, N. (2015). Structure of contingent self-esteem: Global, domain-specific, or hierarchical construct? *European Journal of Psychological Assessment*, 33(5), 388–397. <https://doi.org/10.1027/1015-5759/a000296>
- Sommer, K. (2017). *Stereotype und die Wahrnehmung von Medienwirkungen* [Stereotypes and the perception of effectiveness of media; Doctoral Dissertation; University of Zürich]. Springer VS. doi: <https://doi.org/10.1007/978-3-658-18518-3>
- Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, 45(1), 523–538. <https://doi.org/10.1007/s00181-012-0609-x>
- Strijbos, J. W., & Müller, A. (2014). Personale Faktoren im Feedbackprozess Personal factors in the feedback process. In H. Ditton & A. Müller (Eds.), *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 83–134). Waxmann Verlag.
- Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A Meta-Analysis. *Journal of Educational Psychology*, 99(2), 253–273. <https://doi.org/10.1037/0022-0663.99.2.253>
- 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>
- Trawalter, S., Richeson, J. A., & Shelton, J. N. (2009). Predicting behavior during interracial interactions: A stress and coping approach. *Personality and Social Psychology Review*, 13(4), 243–268. <https://doi.org/10.1177/1088868309345850>
- van de Ridder, J. M., Berk, F. C. J., Stokking, K. M., & ten Cate, O. T. J. (2015). Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher*, 37(8), 767–774. <https://doi.org/10.3109/0142159X.2014.970617>
- van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47(2), 497–527. <https://doi.org/10.3102/0002831209353594>
- 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>
- Waung, M., & Highhouse, S. (1997). Fear of conflict and empathic buffering: Two explanations for the inflation of performance feedback. *Organizational Behavior and Human Decision Processes*, 71(1), 37–54. <https://doi.org/10.1006/obhd.1997.2711>
- Weidinger, A. F., Spinath, B., & Steinmayr, R. (2016). Why does intrinsic motivation decline following negative feedback? The mediating role of ability self-concept and its moderation by goal

- orientations. *Learning and Individual Differences*, 47, 117–128. <https://doi.org/10.1016/j.lindif.2016.01.003>
- Wenz, S. E., & Hoinig, 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>
- Yeager, D. S., Purdie-Vaughns, V., Garcia, J., Apfel, N., Brzustoski, P., Master, A., Hessert, W. T., Williams, M. E., & Cohen, G. L. (2013). Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. *Journal of Experimental Psychology: General*, 143(2), 804–824. <https://doi.org/10.1037/a0033906>

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

**Anna K. Nishen** is currently a doctoral student in the department of Education Research/Heterogeneity at the Freie Universität Berlin. Her research focusses on positive biases towards members of negatively stereotyped groups with a focus on pedagogical contexts.

**Ursula Kessels** earned her doctoral degree in psychology in 2001 and held the chair for Educational Psychology at the University of Cologne from 2009 to 2013. She now holds the chair for Educational Research/Heterogeneity at the Freie Universität Berlin. Her research interests include the role of stereotypes for students' academic outcomes, focusing on the impact of students' gender and migration background for their motivation, self-concepts and evaluation by their teachers.