



# “Failure-to-warn” when giving advice to students? No evidence for an ethnic bias among teacher students in Germany

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

Received: 26 January 2023 / Accepted: 15 February 2024 / Published online: 28 March 2024  
© The Author(s) 2024

## Abstract

Receiving appropriate, unbiased advice from their teachers is important for students’ smaller- and larger-scale educational decisions. However, teachers’ concerns about being or appearing to be prejudiced may interfere and lead them to provide encouraging advice to students belonging to negatively stereotyped groups even when it is not warranted (*failure-to-warn phenomenon*). In this experimental study, we aimed to replicate findings from the US and tested whether teacher students in Germany provided overly encouraging advice regarding the academic plans of a student with a Turkish (vs. German) name. Teacher students ( $n=174$ ) saw the overly ambitious timetable of a (supposed) student with a Turkish or German name and gave advice online on rating scales and in an open-response format. In their advice, they indicated, among others, the perceived demands of the timetable, possible affective and social consequences for the student, and the need to reconsider the timetable. Contrary to expectations and findings from the US, our analyses did not indicate differences in the advice that students with Turkish vs. German names received. Instead, teacher students warned students with a Turkish and German name equally of the difficulties associated with their potential timetable. We discuss both methodology- and theory-related potential explanations for these unexpected findings.

**Keywords** Failure-to-warn phenomenon · Advice · Ethnic bias · Positive bias · Teacher judgments

---

✉ Anna K. Nishen  
anna.nishen@tu-dortmund.de

Ursula Kessels  
ursula.kessels@fu-berlin.de

<sup>1</sup> Educational Research, Freie Universität Berlin, Berlin, Germany

## 1 Theoretical background

Teachers' jobs cover many different tasks inside as well as outside of the classroom. One such task is advising students and their parents, which can encompass advice on topics ranging from learning techniques to educational decisions (Baumert & Kunter, 2013). Advising students well is an important skill, dependent on one's ability to assess students accurately and communicate well with them. However, initial research in the United States shows that advising might be more demanding across lines of difference such as racial group membership (Crosby & Monin, 2007), as has been found for feedback (e.g., Croft & Schmader, 2012; Harber, 1998, 2004; Harber et al., 2010, 2012, 2019). Inappropriate advice—either failing to encourage or to warn someone of an intended course of action—can be problematic for students. They may invest time and energy in futile endeavors or fail to challenge themselves. Either pathway may elicit detrimental emotions, demotivate them and lead them to achieve less than their potential (Cohen & Steele, 2002; Crosby & Monin, 2007; Harber, 2023; Nishen & Kessels, 2022; Pekrun, 2006). Therefore, it is vital to understand whether advice may differ depending on students' (and teachers') group memberships.

Based on the Model of threat-infused intergroup feedback (MOTIIF; Harber, 2023), we examine how teacher students give advice to students who are vs. are not members of a negatively stereotyped group in Germany. To this end, we asked teacher students to give advice to a student with a German or Turkish name regarding the choice of courses for upper secondary school in a between-subjects experimental study. We then examined whether participants' advice, which can be thematically grouped as referring to the demands of the timetable, its possible affective and social consequences for students, and the need to reconsider the timetable, differed depending on the group membership of the student.

### 1.1 The failure-to-warn phenomenon

Communication across racial-ethnic lines can be demanding on both ethnic minority and majority interaction partners (Cohen & Steele, 2002; Harber, 2023; Trawalter et al., 2009). Two US-based experimental studies (Crosby & Monin, 2007) found that Black (vs. White) students were warned less about an overly ambitious university workload. When peer advisors were asked how they would advise a Black student on a proposed plan that was overly demanding, they indicated that they would communicate the academic difficulty to a lesser extent and insist on fewer changes. This finding was subsequently replicated with university students who were not peer advisors. Such different advice may set up students for academic difficulties and emotional duress, which they may then attribute in a disadvantageous way (Crosby & Monin, 2007; Harber, 2023). If the advice given to students in Germany differs in similar ways, this could likewise harm students' learning and future educational prospects.

### 1.2 Concerns about prejudice among members of an ethnic-racial majority

Findings that members of ethnic-racial minorities (vs. majorities) receive more positive feedback or encouraging advice can seem surprising or paradoxical given that

much research demonstrates the existence of widely-known negative stereotypes (e.g., Fiske et al., 2002) and their consequences (e.g., for expectations and judgments; Bonefeld & Dickhäuser, 2018; Glock, 2016; Tenenbaum & Ruck, 2007). Theories on interracial interactions posit that such positive biases emerge when members of ethnic-racial majorities want to demonstrate that they are unprejudiced (Bergsieker et al., 2010; Crosby & Monin, 2007; Harber, 1998, 2023; Richeson & Shelton, 2011; Shelton et al., 2006). This is because transgressing shared values such as being unprejudiced and fair (Aquino & Reed, 2002; Crandall et al., 2002; Lapsley & Lasky, 2001) can threaten one's self-image as a good person as well as one's reputation (Richeson & Shelton, 2011). As a result of what Harber (2023) has termed *racial anxiety*, people may adapt their behavior so that it is less likely to be interpreted as prejudiced. Such over-accommodation is reflected in greater affiliative behaviors such as smiles, positive judgements, and praise (e.g., Blascovich et al., 2001; Harber, 2023; Mendes & Koslov, 2013; Trawalter et al., 2009) and has been found to co-occur with a physiological threat response (Mendes & Koslov, 2013). In a feedback or advice situation, over-accommodation is theorized to take the form of increased praise or decreased criticism (Harber, 2023), and has been found in the US (Crosby & Monin 2007; Harber, 1998, 2004; Harber et al., 2010, 2012, 2019), Canada (Croft & Schmader, 2012), and Germany (Nishen & Kessels, 2022). In the context of advice, discouraging someone from pursuing a highly demanding path could be perceived as holding low expectations for them. This would be in line with negative ability-related stereotypes, and may thus be avoided by a person concerned about being or being perceived to be prejudiced (i.e., experiencing racial anxiety; Crosby & Monin, 2007; Harber, 1998, 2023).

Harber (2023) formalized theoretical arguments about the conditions under which racial anxiety arises within the MOTIIF. He specified that for racial anxiety to arise people must be at risk of being/appearing to be prejudiced and they must have a motive not to be/appear prejudiced. Risk is thought to arise when feedback is given to learners who are part of a group which is negatively stereotyped on the related domain. Moreover, the feedback must relate to an aspect that is sufficiently vague as to make clear criteria difficult to apply (subjective criteria), which has been found repeatedly in empirical studies (Harber, 1998, 2004; Harber et al., 2010, 2012, 2019). Secondly, motives to appear prejudiced are thought to be determined by self-image and reputational concerns, and reduced by other, competing concerns (e.g., being very accurate). Lastly, factors that buffer stress and anxiety in general can also act as buffers for racial anxiety (e.g., self-esteem, social support; Harber, 2023). Importantly, this model posits that racial anxiety arises independently of the communicator's endorsement and application of stereotypes, and instead depends on whether negative stereotypes are known to the communicator, whether there exists an ambiguity about how the feedback is interpreted (risk) and whether they are concerned about being or appearing to be prejudiced (motive).

There is reason to expect that the failure-to-warn phenomenon may also occur in Germany in relation to Turkish-Germans, who are negatively stereotyped regarding their abilities and competencies (e.g., Asbrock, 2010; Froehlich & Schulte, 2019). Teacher students are aware of these stereotypes (Bonefeld & Karst, 2020) and likely want to avoid perpetuating them in their interactions with students. Indeed, German

teacher students have been found to give more inflated praise and more dysfunctional feedback to migrant students who succeeded easily without effort (Zeeb et al., 2022). Moreover, an experimental study showed that German teacher students gave more positive feedback on the essay of a supposed student with a Turkish (vs. German) name (Nishen & Kessels, 2022). Importantly, if positive biases are due to a concern to be or appear to be prejudiced, German teacher students may also fail to warn students with a Turkish (vs. German) name to take on an overly demanding, high workload endeavor. Of course, feedback and advice do differ—whereas feedback relates to a specific performance in the past, advice is future-oriented and broader—but the motivational concern-driven mechanism should lead to a similar pattern of positive bias (Crosby & Monin, 2007; Harber, 2023).

### 1.3 Self-esteem and contingency of self-esteem as potential moderators

While the positive bias is thought to arise due to broad social norms against prejudice, some people may be more vulnerable to racial anxiety than others (Harber, 2023). Identifying more distal personality characteristics that moderate positive biases is important since it may permit us to understand who might be more prone to positive biases towards members of negatively stereotyped groups in general. In the past, self-esteem and contingency of self-esteem on the approval of others have both been examined as potential moderators of a positive bias in feedback (Nishen & Kessels, 2022). Self-esteem is defined as the “subjective assessment of one’s worth as a person” (Donnellan et al., 2015, p. 131). According to Harber (2023), self-esteem may act as a buffer for experiencing racial anxiety. Since self-esteem is associated with lower stress and anxiety and may buffer their disadvantageous effects (e.g., Boulton & Macauley, 2023; Greenberg, 2008; O’Donnell et al., 2008), it is theorized to improve resilience in a stressful situation such as when self-image or reputation may be at risk (Harber, 2023). Indeed, Nishen and Kessels (2022) found that the positive bias in open-response feedback towards Turkish (vs. German) students was especially shown by teacher students who were low in self-esteem.

Importantly, research shows that self-esteem can not only be described by its strength, but also based on the aspects of one’s life on which it depends (Crocker et al., 2003). One of these aspects, namely the approval of others, could increase the motive to be or appear to be unprejudiced. Contingency of self-esteem on the approval of others is the degree to which a person’s self-esteem varies depending on others’ positive opinions of themselves (Crocker et al., 2003; Crocker & Wolfe, 2001). Such a contingency could increase their public reputational concerns of a person because their self-esteem is more threatened if others disapprove of them. Public reputational concerns are one aspect of motives to be or appear to be unprejudiced, which are theorized to increase racial anxiety (Harber, 2023). However, to date there is limited empirical tests on this prediction, and the one study that has examined contingency of self-esteem on other’s approval did not find an effect (Nishen & Kessels, 2022).

## 1.4 Study overview and hypotheses

The present study aims to test whether the predictions of theories on interracial interactions from the US (Harber, 2023; Trawalter et al., 2009) can be applied to pedagogical situations in Germany. Specifically, we test if the US-based findings by Crosby and Monin (2007) can be replicated in advice given by teacher students to supposed 10th grade students in Germany. While the respective studies had compared reactions towards Black versus White students in the US, for a study of the phenomenon in Germany we decided to compare the group of students of German origin with students of Turkish origin, constituting a large minority group which is negatively stereotyped regarding competence (e.g., Asbrock, 2010; Froehlich & Schulte, 2019). In our study, we asked teacher students to give advice to a 10th grade student online regarding the (overly demanding) timetable he planned to take on in the last two years of high school (upper secondary school). We expected a failure-to-warn phenomenon to occur, that is, that teacher students giving advice to a student with a Turkish (vs. German) name would (1) present the timetable as less demanding, and (2) communicate that the timetable needed to be revised to a lesser extent. Adding to the original study by Crosby and Monin (2007), we expected that teacher students would also (3) warn the student less of negative affective and social consequences and highlight positive affective consequences to a greater degree. We included this latter aspect of advice to examine whether bias would also occur in more subtle ways of "warning" a student. That is, teacher students may downplay the negative affective consequences of the planned timetable out of a concern for prejudice and highlight positive affective consequences. Examining advice on the anticipated affective experience is an important addition as the emotions that students experience when encountering difficulties are theorized to have consequences for their motivation and self-regulation (Pekrun, 2006). Lastly, we assessed teacher students' self-esteem and contingency of self-esteem on others' approval following findings by Nishen and Kessels (2022). We expected that the failure-to-warn phenomenon would occur less among teacher students with higher self-esteem and lower contingency of self-esteem because this could be associated with lower racial anxiety.

## 2 Method

### 2.1 Participants

All in all, 220 participants completed the online experiment, of which we excluded those who questioned the cover story ( $n=2$ ), whose answers indicated that they were not taking the questionnaire seriously ( $n=1$ ), or who failed to type in the student's name before giving advice ( $n=9$ ). Due to the specificity of our hypotheses to those teacher students whose own ethnic group is not negatively stereotyped, we excluded teacher students who were likely the target of similar stereotypes (e.g., Turkish or Arabic migration background,  $n=20$ ), and those whose migration background was from the Global South (e.g., India,  $n=5$ ) or unknown ( $n=9$ ). Of the remaining 174 participants, 79.3% were women, which is roughly representative of teachers in Ger-

many (73.4%; Statistisches Bundesamt, 2021). The majority specialized in secondary school teaching ( $n=130$ , 74.7%; primary school teaching:  $n=42$ , 24.1%). On average, participants were in their third semester ( $M=3.3$ ,  $SD=2.2$ ,  $range=1-20$ ).

## 2.2 Development of materials

In German high schools, students from the 11th grade on have some freedom to choose their basic and advanced courses. For the pilot study, we developed four timetables that were in accordance with the regulations while varying in how difficult and time-consuming they might be perceived (subjects chosen for advance placement courses, number of non-required courses chosen). In the pilot study, 40 teacher students (72.5% women) indicated how difficult and time-consuming they considered each of the timetables on two scales from 1 to 10. The timetable perceived as most difficult ( $M=8.8$ ,  $SD=1.7$ ) and most time-consuming ( $M=9.2$ ,  $SD=1.4$ ) was chosen. This timetable included mathematics and physics as the advanced placement courses as well as voluntary lessons in four subjects in addition to the required workload. All in all, if a student followed this timetable, they would have very long and demanding school days (each day had at least eight 45-minute classes).

In addition, we chose a German and a Turkish-origin name based on a dataset in which German participants rated 2000 names on multiple characteristics (Nett et al., 2020). Two names (Dominik, Kerem) that did not differ in the ratings of competence,  $t(39) < 0.01$ ,  $p = .999$ , and warmth,  $t(39) = -0.28$ ,  $p = .783$ , and whose ratings were near the theoretical mean of the scale were chosen. Please note that we chose two names equal in warmth and competence to ensure that any difference that we would find (or not find) could be attributed to the failure-to-warn phenomenon. This is because positive biases in feedback and advice are theorized to arise because of differences in the communication of perceptions, not differences in the perceptions themselves (Harber, 2023). This difference in communication should occur when racial anxiety hinders an advisor to be as frank with the ethnic minority student as with the ethnic majority student even when they are perceived as equally capable. If the perception was not held constant, differences in advice could be attributed to perception, communication, or a combination of both processes. Specifically, a non-significant difference could then be explained by the combined effects of different stereotype-based expectations and the motive to be or appear unprejudiced (compensation effect). Any student perceived as lower in competence would be warned of the timetable to a greater extent because less competent students would have a harder time following such a highly-demanding timetable. That is, if perceived competence was not held constant, advice may be similar because for the Turkish student, the advice would initially be more hesitant (due to lower perceived competence or ambition), but then it would be communicated more positively due to the concern to be or appear to be prejudiced. Such a pattern would lead to null results, but we would not be able to identify whether a failure-to-warn effect occurred. Because advice should differ by perceived competence, we included a manipulation check that the students were indeed perceived as equally gifted and ambitious (see below). Because the failure-to-warn phenomenon emerges due to a concern to be or appear to be prejudiced when giving negative advice (Harber, 2023), and not due to stereotype-congruent

lower expectations, the failure-to-warn effect is still expected when students are perceived as equally able (see also Crosby & Monin, 2007).

### 2.3 Design and procedure

Participants accessed the study via a link that they received through course instructors. Following the guidelines of the department, participants were informed that their participation was anonymous and voluntary and provided active consent. At the time of the data collection no formal approval by an institutional review board was required. As compensation, teacher students participated in a raffle for 10€ gift cards. On the first page, they read that they would provide advice to a student who was participating in a study on how students choose their course load for the last two years of school. As they proceeded to the next page, they were randomly assigned either a student with a German-origin (Dominik B.;  $n=85$ ) or a Turkish-origin name (Kerem Y.;  $n=89$ ). First, participants saw an excerpt of the student's supposed grades which indicated mediocre performance (GPA=2.5 on the German grades scale from 1 to 6, where lower values indicate better grades). Subsequently, they could examine the timetable the student supposedly planned to take. On the same page, participants saw a short note from the supposed student to bolster the cover story (two spelling/grammar mistakes were included to increase authenticity). To again make clear that participants were giving advice, they first had to type in the name of the student in a field that specified "Dear [blank]." Believing that their answers would be forwarded to the target student, participants provided their assessment of his plans and their advice on multiple items with rating scales. Each item started with "I believe..." and specifically addressed the student, e.g., "I believe, this timetable will be [...] for you" (in the original German, the sentence structure permits that the adjective is placed at the end of the sentence, i.e., the different answering options for the rating scale). They then had the option to write a text addressed to the student (open-response format of advice), above which was stated "I think you could consider the following moving forward". Most participants provided such written advice to the student (83.3%,  $n=145$ , of which  $n_{\text{Turkish name}}=73$  and  $n_{\text{German name}}=72$ ). On average, these responses were 103 words long ( $SD=68.5$ ), and this did not differ by target student's name,  $t(141.2)=0.13$ ,  $p=.896$ .

After submitting their advice, participants answered two control variables (student's perceived ambitiousness and giftedness) and provided demographic information. Lastly, participants filled in several scales, among them the moderators self-esteem and contingency of self-esteem. At the end of the data collection period, participants were informed of the cover story and told about the aims of the study as well as the reason for this deception.

### 2.4 Measures

As described above, participants provided their assessment of the timetable and their advice on both rating scales and in an open-response format. For the coding of the open-response advice, a team of four coders rated the advice on four variables (two raters/variable), namely the extensiveness of suggested changes, the degree to which

participants urged the student to make the recommended changes, and the communicated consequences for mental health as well as free time/relationships. All raters were blind regarding the condition and rated each response on scales from 1 to 7 for each variable (see below). These ratings approximate an underlying continuum (e.g., of degree of urging), which is subjective to some degree. For this reason, two raters rated each dimension for each response. After discussions, interrater-reliability on all four variables was moderate to good,  $ICC(2, 2)=0.70-0.88$ , and the two raters' scores were averaged for each construct. For each of these four rated variables, an example of a response for each code is provided in the supplementary materials (Tables S1 and S2). In the following, we describe all dependent variables, first those communications that relate to the demands of the timetable, secondly, those relating to the need to reconsider the timetable, and lastly, those relating to the affective and social consequences.

#### 2.4.1 Demands of the timetable

For their assessment of the course load, participants answered six items with rating scales related to the amount of free time the student would have, how time-consuming, difficult and demanding the course load would be, how easily the student would manage the course load and to what extent he may feel overwhelmed. As described above, each item directly addressed the student from the perspective of the participants, e.g., "I believe this timetable will be [...] for you." All items were assessed on 7-Likert rating scales where 1 represented a very positive assessment (plenty of free time/much too little time-consuming/much too little difficult/much too little demanding/ very easy to manage/not at all overwhelmed) and 7 a very negative assessment (no free time at all/much too time-consuming/much too difficult/much too demanding/very hard to manage/very much overwhelmed). This scale had high reliability (Cronbach's  $\alpha=0.85$ ). Indicating greater demands of the timetable to the student is considered warning the student to a greater extent.

#### 2.4.2 Need to reconsider the timetable

We examined answers on both the items with rating scales and in the open-response advice to assess how much participants advised the student to reconsider the timetable. The less participants advise the student to reconsider his timetable, the less they warn him. Three items reflected participants' advice. Participants indicated the degree to which participants advised that the student should change the course work ("I believe you should change this timetable [...]"; 1 = *not at all* to 5 = *very much*) and whether they recommended that the student should seek out further advice related to his course work as well as to the general demands of the final years of high school ("I believe you need [...] regarding the timetable [general demands of the final years of high school]"; 1 = *no further advice needed*, 5 = *a lot of further advice needed*; Cronbach's  $\alpha=0.81$ ).

To better understand the written advice, the open-response advice was rated regarding (1) their precise content (extensiveness of suggested changes), and (2) the strength or urge with which they conveyed the message (degree of urging). First,



two raters indicated how extensive the concrete suggestions for changes in the advice were (1=*none at all*, 4=*some*, e.g., *dropping two voluntary subjects*, and 7=*very high*, e.g., *dropping voluntary subjects and changing the advanced placement courses*). Greater warning is indicated by more extensive suggestions to change the course of action. The ratings of the two raters were averaged. Two additional raters indicated how strongly each response conveyed that the timetable needed to be changed (1=*not at all* to 7=*very strongly*), with stronger emphasis being indicative of greater warning. This rating reflected the tone of the advice (e.g., urgency), not the concrete suggestion for changes. Because tone is more subjective, we used anchor examples for the raters. Again, we averaged the ratings across the two raters. An example for an excerpt of a response that was rated as low in degree of urging (value of 1.5) would be ““[...] I would just like to recommend, that you could maybe (if it is possible) not take one or two voluntary subjects (depending on your opinion) because it could otherwise be a little too much [...].” In contrast, a response with the same rating on the extensiveness of suggested changes (both had a value of 4), but high in degree of urging (value of 6.5), stated “I personally believe that your time table is truly very very full. You should try to drop a couple of subjects in my opinion so that you also have time for friends/family and your hobbies. It is important that you find a balance between school and free time. Especially because in upper secondary school, you will have to do a lot for your subjects at home. But if you spent at least 8 [hours] at school a day you will barely be able to keep up with your [home work] and definitely will barely have any free time. [...].” Thus, greater values on the rating of the degree of urging reflects a more strong and urgent warning.

### 2.4.3 Affective and social consequences

Again, we examined answers on the items with rating scales and in the open-response advice. First, participants indicated how often they expected the student to experience eight positive and negative achievement-related emotions (Pekrun et al., 2011). Again, the stem of the items clearly specified the advice context (“When I look at your time table, I believe that you will feel the following [emotions] regarding school when you are in upper secondary school”; 1=*never or very seldom* to 7=*always or very often*). We added two further emotions, frustration and disappointment, that could be experienced by students with an overly ambitious workload. We averaged the ratings for the positive emotions (three items, Cronbach’s  $\alpha=0.81$ ) and for negative emotions (six items, Cronbach’s  $\alpha=0.88$ ; boredom was excluded because it may imply high ability). Participants who indicated that students would experience more positive and less negative achievement-related emotions on the items with rating scales warned the student to a lesser extent.

For the open-response advice, two coders rated how strongly students were warned about negative consequences for their mental health (e.g., stress) and negative consequences for their free time/relationships (1=*not at all* to 7=*very strongly*). Bringing up these consequences is a way of warning the student about the implications of their choice, that is, providing the reasoning behind their advice. Again, anchor examples were used and the two raters’ codes were averaged for each scale. An excerpt from a response with a relatively strong warning of the consequences for

mental health would be “This will catch up with you later during the exam period. You could save yourself a lot of [times when you are] feeling overwhelmed and stressed if you reduced [your voluntary lessons]” (value of 5). An example for a relatively strong warning of consequences for free time/relationships would be “I think it is great that you are so ambitious, but the demands for these two years are too high. [...] The work in addition to the lessons is time-consuming and I see very little free time that you allow yourself to both work on the school work as well as experience true free time“ (value of 5.5). Lastly, an excerpt from a response including strong warnings regarding both mental health and free time/relationships is “Based on my own experience I believe that less would be more here. Upper secondary school is much more strenuous than the years before. [...] It is important to have counterpart to all the studying to be happy and to be motivated to stay engaged in school. But in addition to the full timetable and all the work at home you will barely have time for friends, sports, and hobbies. That is quickly frustrating” (values of 6 and 5.5, respectively). Overall, participants who communicated in their open-response advice that students could expect more negative affective and social consequences warned the student to a greater extent.

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

As potential moderators, participants’ self-esteem and contingency of self-esteem on approval of others were assessed based on the scales developed by Rosenberg (1979; German translation by Ferring & Filipp, 1996) and Schwinger et al. (2017), respectively. Participants answered ten items on self-esteem (e.g., “I have a number of good characteristics”; 1 = *disagree strongly* to 4 = *agree strongly*) and five items on contingency of self-esteem on approval of others (e.g., “I cannot respect myself if others do not respect me”; 1 = *disagree strongly* to 5 = *agree strongly*). The scales had good or very good reliability (Cronbach’s  $\alpha_{\text{Self-esteem}} = 0.87$  and Cronbach’s  $\alpha_{\text{Contingency}} = 0.78$ ).

### 3 Results

The descriptive statistics can be found in Table 1 and correlations between all dependent variables in Table 2. To examine whether there was evidence of the failure-to-warn phenomenon, we conducted independent-sample Welch’s *t*-tests comparing the two conditions (Turkish vs. German name). Welch’s *t*-tests have the advantage that they can correct for the degree to which variances differ between two groups. When variances are equal, this test has the same results as a regular *t*-test (Delacre et al., 2017). Post-hoc sensitivity analyses using GPower (Faul et al., 2007) indicated that our analyses would be able to detect a small effect ( $d = 0.38$ ) with a power of 0.80. This is comparable to the effect sizes observed by Crosby and Monin (2007) on rating scale measures, which were  $d = |0.31| - |0.49|$  (own calculation based on assumption of equal group size). To examine whether self-esteem and contingency of self-esteem moderated this effect, regression analyses including their interactions with student name were conducted.

**Table 1** Descriptive parameters of the dependent variables

Dependent variables	Full dataset		Students with German name		Students with Turkish name	
	<i>M (SD)</i>	<i>min-max</i>	<i>M (SD)</i>	<i>min-max</i>	<i>M (SD)</i>	<i>min-max</i>
Demands of the timetable <sup>1</sup>	5.2 (0.6)	3.3–6.7	5.2 (0.6)	3.3–6.3	5.2 (0.6)	4.0–6.7
Need to reconsider the timetable						
Advice <sup>2</sup>	2.7 (0.7)	1.0–4.3	2.8 (0.7)	1.0–4.3	2.7 (0.6)	1.3–4.0
Extensiveness suggested changes <sup>3</sup>	3.4 (1.0)	1.0–6.0	3.6 (1.2)	1.0–6.0	3.3 (0.9)	1.0–5.0
Degree of urging <sup>3</sup>	3.8 (1.4)	1.0–7.0	4.0 (1.5)	1.0–7.0	3.7 (1.4)	1.0–7.0
Affective and social consequences						
Positive achievement-related emotions <sup>4</sup>	4.2 (1.0)	2.0–7.0	4.1 (1.0)	2.0–6.3	4.3 (1.0)	2.0–7.0
Negative achievement-related emotions <sup>4</sup>	3.9 (1.0)	1.2–6.2	4.0 (1.0)	1.7–6.2	3.9 (1.0)	1.2–6.0
Consequences for mental health <sup>3</sup>	2.5 (1.6)	1.0–6.0	2.5 (1.6)	1.0–6.0	2.5 (1.6)	1.0–6.0
Consequences for free time/relationships <sup>3</sup>	2.1 (1.5)	1.0–6.5	2.2 (1.5)	1.0–6.5	2.1 (1.5)	1.0–5.5

Note. <sup>1</sup> Assessed on a rating scale from 1 to 7 (lower numbers represent more positive assessments).

<sup>2</sup> Assessed on a rating scale from 1 to 5 (lower numbers represent less change advised).

<sup>3</sup> Based on open-response advice, rated on a scale from 1 to 7 by two coders (lower numbers represent lower extensiveness/degree of urging/fewer consequences).

<sup>4</sup> Assessed on a rating scale from 1 to 7 (lower numbers indicate less frequent experience of the emotions)

**Table 2** Correlations among all dependent variables

Variable	1	2	3	4	5	6	7
1. Demands of the timetable	–						
2. Advice	0.63***	–					
3. Extensiveness suggested changes	0.35***	0.29***	–				
4. Degree of urging	0.45***	0.47***	0.61***	–			
5. Positive achievement-related emotions	–0.41***	–0.42***	–0.12	–0.29**	–		
6. Negative achievement-related emotions	0.59***	0.44***	0.18*	0.18*	–0.63***	–	
7. Consequences for mental health	0.18*	0.15+	0.28**	0.40***	–0.18*	0.13	–
8. Consequences for free time/relationships	0.17*	0.10	0.27**	0.34***	–0.21*	0.20*	0.33***

Note Coefficients based on pairwise deletion

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

### 3.1 Preliminary analyses

Before further analyses were conducted, we tested whether the students with the German and Turkish names were perceived differently regarding their ambition and giftedness, which could explain potentially varying advice (Crosby & Monin, 2007). Participants rated both students as equally ambitious,  $M_{\text{Turkish}} = 82.6$ ,  $SD = 12.7$ ;  $M_{\text{German}} = 80.3$ ,  $SD = 16.2$ ,  $t(155.8) = -1.03$ ,  $p = .304$ , and equally gifted,  $M_{\text{Turkish}} = 65.6$ ,  $SD = 11.4$ ,  $M_{\text{German}} = 64.8$ ,  $SD = 11.4$ ,  $t(169.7) = -0.49$ ,  $p = .624$ .

## 3.2 Main analyses

### 3.2.1 Demands of the timetable

The Welch's *t*-test indicated that teacher students did not communicate a different assessment of the timetable to the student with a Turkish (vs. German) name,  $t(171.9)=0.19$ ,  $p=.850$ , Cohen's  $d=-0.03$ .

### 3.2.2 Need to reconsider the timetable

Teacher students did not give the student with a Turkish (vs. German) name different advice on the items with rating scales,  $t(167.7)=1.00$ ,  $p=.319$ , Cohen's  $d=-0.15$ . Participants' open-response advice did not differ between the students with regards to how strongly they urged the student to change the plan,  $t(142.1)=1.17$ ,  $p=.244$ , Cohen's  $d=-0.20$  (degree of urging). In tendency, participants suggested less extensive changes to the student with a Turkish compared to a German name,  $t(131.2)=1.81$ ,  $p=.073$ , Cohen's  $d=-0.30$  (extensiveness of suggested changes). This latter finding hints at a failure-to-warn effect. While the differences were all in the direction that would indicate a failure-to-warn phenomenon, they did not reach traditional levels of significance.

### 3.2.3 Affective and social consequences

Participants also communicated in both conditions equally how often students would experience negative achievement-related emotions (e.g., frustration) if they followed through with their plan,  $t(171.9)=0.79$ ,  $p=.431$ , Cohen's  $d=-0.12$ . In tendency, however, participants conveyed more positive achievement-related emotions (e.g., pride) to the student with a Turkish (vs. German) name,  $t(171.8)=-1.94$ ,  $p=.054$ , Cohen's  $d=0.30$ , which could be indicative of a failure-to-warn effect. In the open-response advice, no differences emerged regarding the consequences for mental health,  $t(141.7)=-0.21$ ,  $p=.831$ , Cohen's  $d=0.04$ , or time and relationships,  $t(141)=0.34$ ,  $p=.705$ , Cohen's  $d=-0.06$ .

## 3.3 Moderator analyses: self-esteem and contingency of self-esteem on others' approval

The moderator analyses did not show the expected interaction effects between self-esteem or self-esteem contingency on other's approval and the target student's name (all  $p>.134$ , see Table 3). That is, advice given to students with a Turkish (vs. German) name was not more positive when participants were lower in self-esteem or higher in contingency of self-esteem. The hypotheses were not supported. For most variables, no main effects of self-esteem or contingency of self-esteem emerged either, and when they did, the effects did not reach the threshold of  $p<.05$  (all  $p>.051$ ). This indicates that self-esteem and contingency of self-esteem do not have consistent effects on advice.

**Table 3** Results of the regression analyses with moderator variables

Dependent variables	Intercept	Student name <sup>1</sup> <i>b</i> ( <i>SE</i> )	Self-esteem <i>b</i> ( <i>SE</i> )	Contingency of self-esteem <i>b</i> ( <i>SE</i> )	Name* self-esteem <i>b</i> ( <i>SE</i> )	Name*contingency <i>b</i> ( <i>SE</i> )	Adj. R <sup>2</sup>
Demands of the timetable	5.19***	-0.02 (0.09)	0.21 (0.13)	0.08 (0.09)	-0.13 (0.18)	-0.07 (0.13)	-0.01
Need to reconsider the timetable							
Advice	2.81***	-0.09 (0.10)	0.07 (0.15)	-0.05 (0.11)	0.06 (0.22)	-0.03 (0.16)	-0.01
Extensiveness suggested changes	3.58***	-0.31 (0.17) <sup>+</sup>	0.34 (0.24)	0.36 (0.18) <sup>+</sup>	-0.18 (0.35)	-0.31 (0.26)	0.02
Degree of urging	3.98***	-0.28 (0.24)	0.36 (0.35)	0.07 (0.27)	-0.23 (0.51)	-0.04 (0.37)	-0.02
Affective and social consequences							
Positive achievement-related emotions	4.05***	0.29 (0.15) <sup>+</sup>	0.11 (0.22)	0.05 (0.16)	-0.17 (0.31)	-0.23 (0.22)	<0.01
Negative achievement-related emotions	3.99***	-0.13 (0.15)	0.05 (0.23)	0.13 (0.17)	-0.37 (0.32)	-0.17 (0.23)	-0.01
Consequences for mental health	2.46***	0.05 (0.26)	-0.11 (0.38)	0.27 (0.29)	0.28 (0.55)	-0.21 (0.41)	-0.02
Consequences for free time/ relationships	2.21***	-0.10 (0.25)	0.6 (0.36) <sup>+</sup>	0.42 (0.27)	-0.77 (0.51)	-0.06 (0.38)	0.02

*Note.* For all dependent variables with the exception of positive achievement-related emotions, higher values indicate greater warnings. <sup>1</sup> 0=German name, 1=Turkish name

\*\*\*  $p < .001$ . <sup>+</sup>  $p < .10$

### 3.4 Exploratory analyses

We conducted several exploratory analyses to consider two different explanations for the non-significant results.

#### 3.4.1 Compensation effect

One explanation for the non-significant group differences could be a compensation effect: Participants endorsing the stereotype would underestimate the Turkish student's competence, leading to more discouraging advice, but this effect would then be compensated by the concern to be or appear to be prejudiced, which is theorized to inflate this advice. If this was the case, no group differences would be observed even though the failure-to-warn phenomenon would occur. As intended, our manipulation check indicates that participants perceived students with a Turkish and German name as similarly gifted, speaking against this interpretation. Additionally, we conducted an exploratory analysis from which we excluded participants who rated the students in relatively stereotype-congruent ways, that is, who either rated the Turkish student as relatively low in giftedness ( $-1 SD$ ,  $n=13$ ) or who rated the German student as relatively highly gifted ( $+1 SD$ ,  $n=14$ ). If these participants were responsible for the non-significant results, stronger positive biases on the advice should emerge. When the participants who judged the students in highly stereotype-congruent ways were excluded, two significant effects emerged (all other  $p>.073$ ). Specifically, participants communicated to the student with a Turkish (vs. German) name more strongly that he would experience positive achievement-related emotions,  $t(145)=-2.70$ ,  $p=.008$ , Cohen's  $d=-0.45$ ,  $M_{\text{Turkish}} = 4.44$ ,  $SD=0.95$ ;  $M_{\text{German}} = 4.01$ ,  $SD=0.98$ , an effect that had been marginally significant in the full sample. Moreover, the degree of urging in the open-response advice was lower among participants communicating with a student with a Turkish (vs. German) name,  $t(122)=2.23$ ,  $p=.028$ , Cohen's  $d=0.40$ ,  $M_{\text{Turkish}} = 3.63$ ,  $SD=1.35$ ;  $M_{\text{German}} = 4.18$ ,  $SD=1.40$ , which had not been significant in the full sample. Both of these effects could be interpreted as indicating a failure to warn the student. However, removing those participants that judged the students in the most stereotype-congruent ways, and who might thus compensate the failure-to-warn effect, did not lead to a consistent change in the pattern of results. Of course, due to the exploratory nature of these analyses, these results should be interpreted with caution.

#### 3.4.2 Insufficient manipulation using first names

Possibly, the manipulation of only first names was not targeted enough to elicit specific stereotypes about Turkish people in Germany. To examine whether the first names "Kerem" and "Dominik" were reliably associated with people from a Turkish and German background, we conducted a small study with a convenience sample. Twenty teacher students (80% female, 15% with a migration background) were presented with both names (order randomized). Participation was anonymous, voluntary, and not rewarded due to the short duration (3 min). Participants were informed that this pilot study was for us to understand the clarity of materials used in our stud-

ies. No reference to the content of these studies was made so as not to elicit desired answers. The first author presented the QR code for participation and then left the room when everyone had started the questionnaire. Participants were asked to picture a classroom in Germany and that one of the students was named Kerem [Dominik]. They then indicated (1) which country they believed the student's parents were probably from and (2) which language they believed the name came from. For the name "Kerem," 75% indicated Turkish heritage for both options, and an additional 10% indicated Arabic-speaking countries and Arabic. Nobody indicated German heritage on both prompts, though one person indicated that the name was likely Persian, but the parents were probably born in Germany. For Dominik, 55% indicated German heritage for both prompts, and an additional 40% included German in one of the two prompts (e.g., parents born in Germany, but name originally French). Though we did not conduct a manipulation check in the original study, these results indicate that the first names used in the manipulation are associated by a large majority with Turkish and German heritage, respectively. Even when "Kerem" is associated with other heritage, this name is mainly placed within the Near East/Arabic-speaking world. As a result, it is unlikely that teacher students did not perceive "Kerem" as a student facing a negative competence-related stereotype and that this would be the reason for the absence of a failure-to-warn effect.

#### 4 General discussion

Overall, the results of this study did not provide sufficient support for the hypothesis that advice given to a student with a Turkish as compared to a German name would be more (inappropriately) encouraging. Regardless of the ostensible origin of the student, no significant differences occurred in how demanding participants told the student his timetable was, in how urgently they advised the student to revise the timetable, or in the possible negative affective and social consequences of going through with the plan. Regarding the advice participants gave, as well as the negative achievement-related emotions, the effects were in the expected direction, but did not reach common levels of significance and their effect sizes were small ( $d < 0.20$ ). Two marginally significant differences of small effect size emerged, indicating that teacher students tended to provide less extensive suggestions for change to students with a Turkish (vs. German) name and tended to communicate that they would experience more positive achievement-related emotions. Taken together, the predictions of MOTIIF (Harber, 2023) that teacher students would warn a student with a Turkish (vs. German) name less to pursue his overambitious plan were not consistently supported. Moreover, the hypothesized interaction effects of self-esteem as well as contingency of self-esteem with the student's name did not emerge.

How might this unexpected finding be explained? Firstly, studies measuring biased judgments might be prone to social desirability, so null results could also stem from hypothesis guessing and the desire to appear unprejudiced. However, as the phenomenon of giving more positive feedback or advice to minority groups is much less debated in Germany than the occurrence of a more negative evaluation of these groups, it is highly unlikely that our participants might have suppressed their

tendency to give a less strong warning to the student with a Turkish name. Another explanation might be cultural differences between Germany and the US (e.g., differences in anti-prejudice norms regarding the respective groups). The US-American and the German cultural context look back on very different histories regarding the ethnic minorities which are typically focused when studying ethnic biases (i.e., Black and White Americans vs. people with a Turkish migration background and without a migration background). For example, past discriminatory laws in the US clearly limited the education of Black people in the US (e.g., separate-but-equal doctrine during Jim Crow). This may make educators in the US more vigilant of racial biases in educational settings— which could mean that racial anxiety is more prominent. The relationship between White and Black Americans also differs in other important aspects from the relationship between Germans without a migration background and descendants of Turkish labor migrants (e.g., perceived foreignness; Juang et al., 2021; Zou & Cheryan, 2017).

However, a positive bias in performance feedback has been found in Germany (Nishen & Kessels, 2022). This finding could therefore be better explained by considering whether differences in feedback and advice may be more meaningful than originally theorized. Feedback refers back to a past performance, whereas advice is more future-oriented (Blunden et al., 2021; Crosby & Monin, 2007). Advice is typically given on limited occasions on a highly specific question and in case of the timetable in the present study, it is a highly consequential decision. Feedback, on the other hand, occurs with greater frequency. While participants in both the present study and the Nishen and Kessels (2022) study communicated with the supposed students only once, the advice in the current study may have been perceived as carrying greater weight, impacting a highly relevant decision to be made by the student. If advice was perceived as more consequential, teacher students in the present study may have felt higher accountability, which has been shown to reduce positive biases (Ruscher et al., 2010). In terms of the model for threat-infused intergroup feedback (Harber, 2023), this would be described as a motive that competes with the motive to be or appear to be unprejudiced, which is theorized to reduce racial anxiety.

#### 4.1 Limitations

However, we cannot rule out that other specificities of our design may have led to the absence of a positive bias in this as opposed to prior research. For example, teacher students in our study participated online due to pandemic restrictions. Online participation may have reduced self-awareness relative to lab experiments, as students participated from a familiar environment without a researcher nearby. Previously, Harber (1998) has argued that reduced situational self-awareness may lead to fewer concerns about how one is or appears to others and thus to lower positive biases. Additionally, our post-hoc sensitivity analyses revealed that we would be able to detect effects as low as  $d=0.38$ . Thus, we were limited in our ability to test for small effects, and indeed, the two marginally significant effects were a little lower than this effect size from the post-hoc sensitivity analyses. Thus, the sample size may have played a role in not identifying effects because of their relatively small size.



Moreover, we were the first to test the failure-to-warn phenomenon among teacher students preparing to work in schools (rather than among peer advisors at a university; Crosby & Monin, 2007). To this end, we made use of open-response written advice in addition to the items with rating scales used in prior research. It is possible that a failure-to-warn phenomenon may be more pronounced in verbal face-to-face advice, as the person receiving the advice is immediately reacting to the information. Though Crosby and Monin (2007) examined a hypothetical situation, the peer advisors in their sample might base their answers on the advice-giving context with which they were familiar. This might be another reason why the failure-to-warn phenomenon occurred there. In schools, advice is likely given in personal encounters rather than in written form, so it would be an important test whether the failure-to-warn effect emerges in actual advice conversations. In the past, the positive feedback bias has indeed been observed in face-to-face interactions (Harber, 2004).

Lastly, we followed Crosby and Monin (2007) in considering only a timetable that was clearly overly demanding and not feasible for an average-performing student. However, teachers will be giving advice to students of different ability levels. For lower-achieving students, average-level timetables may already be considered highly demanding. Discouraging these students may communicate low expectations even more strongly, as the timetable is judged to be feasible for others. This could increase concerns about being or appearing prejudiced, and may thus facilitate the failure-to-warn phenomenon. Indeed, Harber (2023) suggests that low-achieving ethnic minority students may be most vulnerable to a positive feedback bias. Overall, research specifically targeting positive biases at different ability levels is needed. Ultimately, our research is limited to a specific advice situation and has used one specific timetable, which represents only one of many scenarios that teachers will likely encounter.

## 4.2 Implications

If our null results represent a true absence of the failure-to-warn phenomenon in the context we examined, this could inform theoretical arguments relating to differences between advice and feedback as well as between the US-American and German cultural contexts. The differences between advice and feedback regarding focus on the future, frequency of the situation, and perceived consequences of the communication (Blunden et al., 2021; Crosby & Monin, 2007) may warrant greater theoretical differentiation and will need to be established in further empirical work. In addition, research on stereotypes and biased behavior must always be understood in its specific historical, social, and cultural context. This points to an important next step in theory development: Clarifying sociocultural conditions for these psychological processes.

Alternatively, methodological specificities and circumstances may be responsible for the null result. Therefore, design features that could plausibly influence the failure-to-warn phenomenon should be examined in a next step (e.g., online vs. lab assessment). Further research may be warranted since the differences between groups were often (but not always) in the expected direction, but did not reach traditional levels of significance.

### 4.3 Conclusion

Advising students well— which includes warning them when they have taken on too much— is an important skill for teachers. Overall, our study did not support the predictions of the model of threat-infused intergroup feedback (Harber, 2023) and did not replicate prior US-based findings that ethnic minority students would be warned of an overly ambitious plan to a lesser extent. Instead, teacher students in Germany gave equivalent advice to students with German and Turkish names. Both methodological and theoretical explanations for this unexpected finding should be considered and explored in further research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11218-024-09899-y>.

**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). The authors have no relevant financial or non-financial interests to disclose. Open Access funding enabled and organized by Projekt DEAL.

**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

- Aquino, K., & Reed, A. I. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440. <https://doi.org/10.1037/0022-3514.83.6.1423>.
- Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology*, 41(2), 76–81. <https://doi.org/10.1027/1864-9335/a000011>.
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–48). Springer US.
- 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>.
- Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology*, 80(2), 253–267. <https://doi.org/10.1037/0022-3514.80.2.253>.
- Blunden, H., Yoon, J., Kristal, A. S., & Whillans, A. V. (2021). *Soliciting advice rather than feedback yields more developmental, critical and actionable input* (Working Paper, No. 20–021). Harvard Business School.
- Bonefante, 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., & 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 VS. <https://doi.org/10.1007/978-3-658-27275-3>
- Boulton, M. J., & Macaulay, P. J. (2023). Does authentic self-esteem buffer the negative effects of bullying victimization on social anxiety and classroom concentration? Evidence from a short-term longitudinal study with early adolescents. *British Journal of Educational Psychology*, 93(2), 500–512. <https://doi.org/10.1111/bjep.12573>.
- 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>.
- 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., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review*, 108(3), 593–623. <https://doi.org/10.1037/0033-295X.108.3.593>.
- 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>.
- 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>.
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of student's *t*-test. *International Review of Social Psychology*, 30(1), 92–101. <https://doi.org/10.5334/irsp.82>.
- Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2015). Measures of self-esteem. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 131–157). Academic Press.
- Faul, F., Erdfelder, E., Lang, A. E., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- 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.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>.
- Froehlich, L., & Schulte, I. (2019). Warmth and competence stereotypes about immigrant groups in Germany. *Plos One*, 14(9), 1–17. <https://doi.org/10.1371/journal.pone.0223103>.
- 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>.
- Greenberg, J. (2008). Understanding the vital human quest for self-esteem. *Perspectives on Psychological Science*, 3(1), 48–55. <https://doi.org/10.1111/j.1745-6916.2008.00061.x>.
- 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. (2023). The model of threat-infused intergroup feedback: Why, when, and how feedback to ethnic minority learners is positively biased. *Educational Psychologist*, 58(3), 162–177. <https://doi.org/10.1080/00461520.2023.2170377>
- Harber, K. D., Gorman, J. L., Gengaro, F. P., Butsingh, 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*(17), 1120–1242. <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/10.1348/014466609X473956>.
- Juang, L. P., Schwarzenenthal, M., Moffitt, U., & Vietze, J. (2021). No, where are you really from? Testing the foreigner objectification scale with German adolescents. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 53*(3–4), 82–93. <https://doi.org/10.1026/0049-8637/a000242>
- Kessels, U., & Nishen, A. K. (2022). Shifting Standards und positiver Feedback-Bias: Wenn negative Stereotype zu positiveren Urteilen führen [Shifting standards and the positive feedback bias: When negative stereotypes lead to more positive evaluations]. In S. Glock (Ed.), *Stereotype in der Schule II* (pp. 70–103). Springer. [https://doi.org/10.1007/978-3-658-37262-0\\_3](https://doi.org/10.1007/978-3-658-37262-0_3)
- Lapsley, D. K., & Lasky, B. (2001). Prototypic moral character. *Identity, 1*(4), 345–363. [https://doi.org/10.1207/s1532706xid0104\\_03](https://doi.org/10.1207/s1532706xid0104_03).
- 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>.
- Nett, T., Dorrrough, A., Jekel, M., & Glöckner, A. (2020). Perceived biological and social characteristics of a representative set of German first names. *Social Psychology, 51*(1), 17–34. <https://doi.org/10.1027/1864-9335/a000383>.
- Nishen, A. K., & Kessels, U. (2022). Non-communicated judgements of, vs. feedback on, students' essays: Is feedback inflation larger for students with a migration background? *Social Psychology of Education, 25*(1), 1–31. <https://doi.org/10.1007/s11218-021-09674-3>
- O'Donnell, K., Brydon, L., Wright, C. E., & Steptoe, A. (2008). Self-esteem levels and cardiovascular and inflammatory responses to acute stress. *Brain, Behavior, and Immunity, 22*(8), 1241–1247. <https://doi.org/10.1016/j.bbi.2008.06.012>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>.
- Richeson, J. A., & Shelton, J. N. (2011). Stereotype threat in interracial interactions. In M. Inzlicht, & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 583–605). Oxford University Press.
- Rosenberg, M. (1979). *Conceiving the self*. Basic Books.
- Ruscher, J. B., Wallace, D. L., Walker, K. M., & Bell, L. H. (2010). Constructive feedback in cross-race interactions. *Group Processes and Intergroup Relations, 13*(5), 603–619. <https://doi.org/10.1177/1368430210364629>.
- Schwinger, M., Schöne, C., & Otterpohl, N. (2017). 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>.
- Shelton, J. N., Richeson, J. A., & Vorauer, J. D. (2006). Threatened identities and interethnic interactions. *European Review of Social Psychology, 17*(1), 321–358. <https://doi.org/10.1080/10463280601095240>.
- Statistisches Bundesamt (2021, September). *Anteil der weiblichen Lehrkräfte an allgemeinbildenden Schulen in Deutschland im Schuljahr 2020/2021 nach Schulart* [Percentage of female teachers at schools in Germany in the 2020/2021 school year by type of school]. <https://de.statista.com/statistik/daten/studie/1129852/umfrage/frauenanteil-unter-den-lehrkraeften-in-deutschland-nach-schulart>.
- 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>.
- 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>.
- Zeeb, H., Hunecke, J., & Voss, T. (2022). Teachers' dysfunctional feedback to students from immigrant and non-immigrant backgrounds: A pilot study. *Social Psychology of Education, 25*(6), 1273–1306. <https://doi.org/10.1007/s11218-022-09725-3>

Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of Personality and Social Psychology, 112*(5), 696–717. <https://doi.org/10.1037/pspa0000080>.

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

**Anna K. Nishen** After completing her PhD at the Freie Universität Berlin, Anna K. Nishen is a post-doctoral researcher at the Department of Educational and Differential Psychology at Technische Universität Dortmund. Her interests lie in effects of inter- and intrapersonal mechanisms behind stereotype-(in) congruent educational outcomes.

**Ursula Kessels** is a psychologist and holds a professorship for heterogeneity and education at the Department of Education and Psychology at Freie Universität Berlin. She is interested in motivational and identity-related predictors of learning and educational inequality.