The British
Psychological Society

Being smart or getting smarter: Implicit theory of intelligence moderates stereotype threat and stereotype lift effects

Laura Froehlich[1]*, Sarah E. Martiny[2], Kay Deaux[3], Thomas Goetz[4,5] and Sog Yee Mok[6]

[1]University of Hagen, Germany
[2]UiT The Arctic University of Norway, Tromsø, Norway
[3]New York University, USA
[4]University of Konstanz, Germany
[5]Thurgau University of Teacher Education, Kreuzlingen, Switzerland
[6]Technical University of Munich, Germany

This research explores implicit theory of intelligence (TOI) as a moderator of stereotype activation effects on test performance for members of negatively stereotyped and of favourably stereotyped groups. In Germany, Turkish-origin migrants are stereotyped as low in verbal ability. We predicted that on a test diagnostic of verbal intelligence, endorsement of an entity TOI predicts stereotype threat effects for Turkish-origin students and stereotype lift effects for German students. This effect could account for some of the performance gap between immigrants and host society members after stereotype activation. Study 1 (N = 107) established structural equivalence of implicit theories across the ethnic groups. In two experimental studies (Study 2: N = 182, Study 3: N = 190), we tested the moderating effect of TOI in a 2 (stereotype activation: diagnostic vs. non-diagnostic test) × 2 (ethnicity: German vs. Turkish migration background) experimental design. The results showed that when the test was described as diagnostic of verbal intelligence, higher entity theory endorsement predicted stereotype threat effects for Turkish-origin students (Study 2 and Study 3) and stereotype lift effects for German students (Study 3). The results are discussed in terms of practical implications for educational settings and theoretical implications for processes underlying stereotype activation effects.

Imagine a student taking a test that is diagnostic of verbal intelligence. Might the effects of membership in a negatively stereotyped or favourably stereotyped group on performance depend on what the student believes about the fixedness of intelligence? The following work argues that the often-observed performance gap between immigrants and members of the host society (e.g., OECD 2010, 2015) might at least partly be caused by an interaction between group members' implicit theory of intelligence (TOI) and the activation of ability-related stereotypes. More precisely, we hypothesize that (1) when negative ability-related stereotypes are activated in performance situations, an entity view

*Correspondence should be addressed to Laura Froehlich, University of Hagen, Universitaetsstr. 33, 58097 Hagen, Germany (email: laura.froehlich@fernuni-hagen.de).

of intelligence hinders performance, and (2) when favourable stereotypes are activated, an entity view boosts performance.

Stereotype threat

Stereotype threat theory (Steele & Aronson, 1995) contributes to the explanation of achievement-related group differences in performance. Ample research documents that the situation in which a test is administered can influence the performance of certain groups of test takers (e.g., Inzlicht & Schmader, 2012; Nguyen & Ryan, 2008). The activation of negative stereotypes in achievement situations leads to underperformance of members of the negatively stereotyped group. For example, Steele and Aronson (1995) showed that when African-American students took a test that was described as diagnostic of intelligence, they performed worse compared to white students, controlling for prior ability level. However, when the test was described as non-diagnostic, no ethnic performance differences emerged. Stereotype threat effects have been documented for various groups, including African-Americans (e.g., Blascovich, Spencer, Quinn, & Steele, 2001) and students of low socioeconomic status in standardized achievement tests (e.g., Croizet & Millet, 2012).

Furthermore, negative stereotypes can contribute to group differences in performance not only by reducing the performance of negatively stereotyped group members (i.e., stereotype threat) but also by enhancing the performance of group members not targeted by the negative stereotype (i.e., stereotype lift; Walton & Cohen, 2003). In comparison with a negatively stereotyped outgroup, other groups are favourably stereotyped and thus can engage in downward social comparison when stereotypes are activated. Under these conditions, they are likely to show higher performance than when no stereotypes are activated (e.g., Chatard, Selimbegović, Konan, & Mugny, 2008; Walton & Cohen, 2003).

A relevant variable when assessing the effects of ability stereotypes on performance is the belief in fixed versus malleable intelligence (e.g., Dweck, 1999). Although research has identified that several moderators of stereotype threat effects (for reviews, see Aronson, 2002; Martiny & Götz, 2011) and some moderators of stereotype lift effects have been suggested (Chatard *et al.*, 2008; Walton & Cohen, 2003), meta-analyses still show a considerable variation in effect sizes across studies. Therefore, further investigation of moderating variables has been called for (e.g., Nguyen & Ryan, 2008; Walton & Cohen, 2003). For this reason, we will examine the moderating role of implicit theories of intelligence for negatively stereotyped as well as for favourably stereotyped group members.

Implicit theories of intelligence

Lay conceptions about the nature of intelligence predict performance when failure occurs on difficult tasks. People hold different theories about the fixedness or malleability of intelligence (e.g., Dweck & Leggett, 1988; Molden & Dweck, 2006). If people conceive of intelligence as fixed (i.e., hold an entity theory), they prefer performance goals; if they conceive of intelligence as malleable (i.e., hold an incremental theory), they prefer learning goals and believe that they can expand intelligence by effort. After failure, entity theorists' performance is impaired, while incremental theorists' performance is not affected or even enhanced (e.g., Grant & Dweck, 2003; Levy, Stroessner, & Dweck, 1998).

The interaction between stereotype activation and implicit theory of intelligence

Ability-related stereotypes imply that the ability of one group in the target domain is lower than that of another group. Under these conditions, performance is likely to differ, depending on which implicit TOI members of negatively and favourably stereotyped groups endorse. Earlier research bringing stereotype threat theory and TOI together showed that interventions making an incremental theory salient can improve negatively stereotyped students' performance (e.g., Aronson, Fried, & Good, 2002; Blackwell, Trzesniewski, & Dweck, 2007; Good, Aronson, & Inzlicht, 2003). For example, Aronson *et al.* (2002) showed that the GPA of both black and white university students increased when they participated in an intervention promoting a malleable view of intelligence compared to two control groups. Good *et al.* (2003) showed increased standardized test performance of female, minority, and low-income students after an incremental theory intervention. Importantly, however, stereotype threat was not experimentally manipulated when considering the interventions' effect on performance in either study. Good and colleagues assumed that stereotype threat occurred during high-stakes testing, whereas Aronson and colleagues assessed only self-reported perceptions of stereotype threat unrelated to specific testing situations. Therefore, our aim in the present work was to conduct a more explicit test of the predicted relation between stereotype activation and TOI by experimentally activating ability-related stereotypes, as well as measuring TOI and performance on a standardized test within in a single study, thus allowing us to identify causal relationships. Further, rather than making a particular TOI salient by experimental instruction or intervention, we investigate the interaction of stereotype activation with the TOI that participants chronically endorse. Whereas the focus of the earlier research was the influence of different levels of TOI on performance, the present research focuses on the consequences of a particular TOI endorsement for students in achievement situations. More precisely, by assessing instead of manipulating TOI, we are able to draw conclusions concerning the effects that natural variation in this trait variable has on students' performance in achievement tests. Thus, the present research addresses the question: 'For which groups of students and under which conditions does entity theory endorsement predict performance?'

Effects of stereotype activation on negatively stereotyped group members' performance

When a test is believed to be diagnostic of intelligence (i.e., a well-established experimental activation of ability stereotypes; e.g., Steele & Aronson, 1995), entity theory endorsement should negatively predict performance for members of a negatively stereotyped group because the activation of a negative ability-related stereotype indicates that the in-group has previously failed in the target domain. Entity theorists are likely to react strongly to this past group-level failure as they will assume that it predicts their own failure in future. Importantly, because they do not believe in the value of increasing effort, they see themselves unable to do anything about the potential failure in future. Thus, they will likely become concerned about the possibility of performing poorly themselves (Brown & Josephs, 1999), which would in their eyes indicate that they are just not smart. This concern should intensify the threat experience by increasing concerns and worries occupying working-memory capacity (Schmader, Johns, & Forbes, 2008) and thus hinder performance.

Negatively stereotyped group members holding an incremental view, on the other hand, should not be threatened by the activation of negative ability-related stereotypes in achievement situations. As earlier research shows, they cope better with actual failure

(e.g., Dweck, 2006) and thus should also cope better with potential failure. They should not be threatened to the same extent as entity theorists because they believe that their performance can be increased by effort. Even a failure on the task at hand does not mean that they are not smart, but only that they should invest more effort in future. Thus, the activation of negative stereotypes should not hinder incremental theorists' performance.

Effects of stereotype activation on favourably stereotyped group members' performance

Individuals who do not belong to a negatively stereotyped group are expected to do better than members of the negatively stereotyped group; thus, describing a test as diagnostic of intelligence would, for members of this favourably stereotyped group, serve to activate performance goals and the stereotype that one's group has high ability. That is, by measuring ability, the test offers the opportunity to get a favourable judgement of one's competence (Dweck & Leggett, 1988; Grant & Dweck, 2003). Entity theorists are likely to see the diagnostic test as an opportunity to get positive feedback about their ability; namely, they can prove that they are smart, and no potential threat is activated. This is in line with earlier work arguing that 'an entity view of ability should specifically boost the performance of favorably stereotyped group members by assuring them that their group's advantage. . ., is immutable'. (Mendoza-Denton, Kahn, & Chan, 2008, p. 1188). Mendoza-Denton and colleagues manipulated both stereotype activation (favourable stereotype confirmed vs. disconfirmed) and view of intelligence (entity prime vs. incremental prime). Favourably stereotyped group members performed better when the stereotype was confirmed and they were primed with an entity view. Negatively stereotyped group members showed a stereotype threat effect, but stereotypes did not interact with views of intelligence. The authors suggest that stereotype threat and lift effects may be differentially affected by relatively brief experimental manipulations of views of intelligence.

Whereas brief manipulations of views of intelligence may not have similar effects on stereotype threat and stereotype lift, we propose that the effects might emerge when TOI is measured as a dispositional trait variable, reflecting how people generally cope with performance situations. Specifically, when members of a favourably stereotyped group who endorse an entity theory are given the opportunity to show that they are smart on a diagnostic test, their performance should improve. Further, the activation of favourable stereotypes should not boost the performance of incremental theorists, as they see effort as the most important determinant of their performance rather than their in-group's assumed ability level.

Stereotype threat for immigrants in Europe

In the present work, we extend previous research conducted with different ethnic minorities in the US to the European, specifically the German, educational context. Apart from the well-established stereotype threat effects for African-Americans (e.g., Brown & Day, 2006; Steele & Aronson, 1995), relatively little work has investigated stereotype threat effects for immigrants (Appel, Weber, & Kronberger, 2015). One exception is Deaux *et al.* (2007), who investigated stereotype threat for West Indian immigrants in the United States. Although first-generation West Indian immigrants showed no stereotype threat effects, the performance of second-generation West Indian immigrants decreased under stereotype threat, similar to the patterns found for African-Americans. In the European context, Appel (2012) showed that Austrian immigrant students' intellectual

performance was impaired when they were presented with anti-immigrant right-wing political propaganda (i.e., social identity threat), but not when presented with neutral political posters.

In Germany, approximately 19% of the population has a migration background, and almost one-third of the students in the German educational system are immigrants (Bundesamt für Migration und Flüchtlinge [Federal Office for Migration and Refugees], 2010). The largest group are immigrants of Turkish origin (18.5%; Statistisches Bundesamt [Federal Statistical Office], 2012), who were recruited as guest workers between the 1950s and the 1980s (e.g., Herbert, 2001). Negative stereotypes about Turkish-origin migrants are widespread in Germany, including stereotypes about intellectual competence (e.g., Asbrock, 2010; Froehlich, Martiny, Deaux, & Mok, 2016; Kahraman & Knoblich, 2000). Turkish-origin students underperform compared to German students in standardized performance tests (e.g., Programme for International Student Assessment [PISA]: Klieme *et al.*, 2010; IGLU: Bos, 2012a; TIMSS: Bos, 2012b) and have low socioeconomic status (e.g., Ruehl & Babka von Gostomski, 2012). Stereotype threat effects have been reported for Turkish-origin high-school students in the mathematical domain (Martiny, Mok, Deaux, & Froehlich, 2015). As the underperformance of Turkish-origin migrants is an important societal issue, it is essential to investigate potential moderators that might determine the extent of the ethnic performance gap. For this reason, we focus on the role of implicit theories of intelligence in explaining performance differences between Germans and Turkish-origin students after stereotype activation.

The present research

We conducted three studies with German and Turkish-origin high-school students. Study 1 assessed the correlational structure of TOI endorsement, goal orientation, and effort beliefs for the two ethnic groups as a preliminary study to the subsequent moderator analysis, in that we tested whether the moderator (i.e., TOI) was confounded with one of the focal independent variables of stereotype threat research (i.e., participants' ethnicity). We assessed how strongly participants endorsed an entity TOI (i.e., the more one endorses an entity view, the less one holds an incremental view) and how strongly they favoured learning goals and saw effort as a means to expand their ability. The results of Study 1 established structural equivalence of TOI between German and Turkish-origin students. In Studies 2 and 3, we investigated the interaction between TOI, group membership, and stereotype activation on verbal performance. Specifically, we hypothesized that entity TOI and performance would be negatively related after stereotype activation for members of the negatively stereotyped group (i.e., Turkish-origin migrants), consistent with stereotype threat. For members of the favourably stereotyped group (i.e., Germans), we hypothesized that entity TOI and performance would be positively related after stereotype activation, consistent with stereotype lift.

STUDY 1

Participants

Ninth-grade students from 15 classes of German middle-track high schools (Realschule) in Stuttgart participated in a questionnaire study. We focused on the middle track because there the proportions of students with migration backgrounds and Germans are approximately equal, in contrast to the upper-level (more Germans) and lower-level track (more immigrants). In each study, the participants were considered to have migration

background if they, either of their parents, or grandparents were born outside of Germany. Because we tested all students within a classroom, samples included not only German and Turkish-origin students, but also students with other migration backgrounds. We excluded these other immigrants because (1) only the Turkish-origin migrant group was large enough to conduct meaningful analyses and (2) stereotypes about different immigrant groups vary substantially (Froehlich *et al.*, 2016; Lee & Fiske, 2006), arguing against analysing a heterogeneous immigrant group. The sample of Study 1 consisted of 107 students: 61 had no migration background and 46 had a Turkish migration background; age ranged from 13 to 18 years (*M* = 15.29, *SD* = 0.77); 44 students were female.

Materials and procedure

Participants completed a questionnaire (in German language) on TOI, goal orientation, and effort beliefs (Blackwell *et al.*, 2007) in their classrooms in the presence of a German experimenter and a teacher. Three questions measured endorsement of an entity TOI ('You have a certain amount of intelligence and you really can't do much to change it', 'Your intelligence is something about you that you can't change very much', and 'You can learn new things, but you can't really change your basic intelligence'). Three questions measured learning-goal orientation ('An important reason why I do my school work is because I like to learn new things'; 'I like school work best when it makes me think hard'; and 'I like school work that I'll learn from even if I make a lot of mistakes') and three additional questions measured effort beliefs ('The harder you work at something, the better you will be at it'; 'To tell the truth, when I work hard at my schoolwork, it makes me feel like I'm not very smart'; and 'If you're not good at a subject, working hard won't make you good at it'). All items were answered on 7-point Likert scales ranging from 1 = *strongly disagree* to 7 = *strongly agree*.

Results

To investigate structural equivalence, we used a method by van de Vijver and Leung (1997). First, we conducted exploratory factor analyses with VARIMAX rotation for the samples of German and Turkish-origin students separately. Three factors were requested, based on previous work showing that the nine items assessed three distinct constructs – TOI, learning goals, and effort beliefs. Table 1 displays the items and loadings for the rotated factors. In both ethnic groups, the three items measuring an entity TOI loaded on Factor 1 ('TOI'). Factor 2 ('learning goals') consisted of the three items on learning-goal orientation and one item measuring positive effort beliefs. Factor 3 ('effort beliefs') consisted of the remaining two items reflecting a negative effort beliefs.

We then computed Tucker's Phi coefficients to assess equivalence across ethnic groups (van de Vijver & Leung, 1997). Phi greater than .90 indicates structural equivalence across samples, and the first two factors met this criterion (.93 for 'TOI' and .92 for 'learning goals'); thus, these factors can be regarded as structurally equivalent for German and Turkish-origin students. The coefficient for Factor 3 ('effort beliefs') was .89, slightly below the recommended level. In the German student sample, only the two items reflecting negative effort beliefs loaded highly on Factor 3; in the Turkish-origin student sample, all three items measuring effort beliefs loaded highly on Factor 3. Table 2 depicts mean entity theory endorsement for Germans and Turkish-origin migrants in all three studies. The reported analysis was exploratory in nature and allowed only for limited

Table 1. Loadings for the rotated factors for Germans and Turkish-origin migrants in study 1

	Germans (N = 61)			Turkish migrants (N = 46)		
	I	2	3	I	2	3
You have a certain amount of intelligence and you really can't do much to change it	0.84			0.73		
Your intelligence is something about you that you can't change very much	0.83			0.82		
You can learn new things, but you can't really change your basic intelligence	0.85			0.81		
An important reason why I do my school work is because I like to learn new things		0.63			0.72	
I like school work best when it makes me think hard		0.43			0.86	
I like school work that I'll learn from even if I make a lot of mistakes		0.72			0.86	
The harder you work at something, the better you will be at it		0.76			0.42	0.41
When I work hard at my schoolwork, it makes me feel like I'm not very smart			0.80			0.79
If you're not good at a subject, working hard won't make you good at it	0.44		0.53			0.63
Eigenvalue	2.85	1.75	1.24	2.34	2.20	1.03
% of variance	31.63	19.42	13.81	25.98	24.49	11.41

Note. Loadings < .40 are omitted.

Table 2. Mean entity theory endorsement of Germans and Turkish-origin migrants (standard deviation in parentheses)

	Ethnicity	
Entity theory endorsement	Germans	Turkish migrants
Study 1	2.83 (1.56)	3.10 (1.39)
Study 2	2.74 (1.26)	3.54 (1.38)
Study 3	2.82 (1.27)	2.73 (1.28)

conclusions about measurement invariance; therefore, we also performed confirmatory factor analysis (CFA) with multiple-group comparison. This method allows the testing of specific hypotheses of measurement invariance, ideally with large sample sizes to ensure reliable parameter estimation (e.g., Kline, 2011; Little, 2013). Because sample size in Study 1 was comparatively small, we performed CFA only as a supplementary analysis and we report the results in the Appendix A. The findings were in line with the exploratory factor analysis in that they showed metric invariance for the two factors of TOI and goal orientation.

Discussion

Study 1 investigated the equivalence of factor structures of TOI and motivational variables for German and Turkish-origin students. Showing equivalence of TOI across ethnic

groups is a requirement for the moderator analyses conducted in Studies 2 and 3. We found that TOI items formed equivalent factors for both ethnic groups, and we thus included TOI as a moderator of stereotype activation effects in the subsequent studies. In addition, the results suggest that positive and negative effort beliefs form two separate constructs for Germans, whereas they are unidimensional for Turkish-origin migrants.

STUDY 2

We hypothesized a three-way interaction between experimental condition, ethnicity, and TOI to predict verbal performance. Specifically, when stereotypes are activated, the more Turkish-origin students endorse an entity TOI, the worse they are expected to perform. In contrast, the more German students endorse an entity TOI, the better they are expected to perform.

Participants

One hundred and eighty-two ninth-grade students from 13 classes of middle- and lower-track high schools (Integrierte Sekundarschule; a recent reformation had merged the lower and middle school tracks) in Berlin districts with a high percentage of immigrants participated in an experiment. Eight participants were excluded from the analyses due to the late arrival or the lack of compliance. The final sample consisted of 174 participants (69 female, 9 missing). Forty-seven students had no migration background, while 127 students had Turkish migration background. Age ranged from 14 to 21 years ($M = 14.69$, $SD = 0.82$).

Materials and procedure

The study had a 2 (stereotype activation: diagnostic vs. non-diagnostic test) × 2 (ethnicity: German vs. Turkish migration background) between-participants design. TOI was assessed as a continuous variable. Performance on the verbal ability test served as the dependent variable. Participants were randomly assigned to one of two experimental conditions of stereotype activation (diagnostic: 24 Germans and 65 Turkish-origin migrants; non-diagnostic: 23 Germans and 61 Turkish-origin migrants).

Students participated in their classrooms in the presence of a German experimenter and a teacher. Prior to a performance test, test diagnosticity was experimentally manipulated. In the diagnostic condition, instructions indicated that an upcoming performance test would measure verbal intelligence and that participants' scores would be compared to other students' scores. In the non-diagnostic condition, instructions indicated that the test was a non-diagnostic practice test for test development and that participants' scores would not be compared to other students' scores (Steele & Aronson, 1995). Students briefly summarized the instructions in writing to ensure awareness and comprehension. They then worked on a verbal performance test and filled out a questionnaire assessing the moderator variable (i.e., implicit theory of intelligence; TOI, $\alpha = .73$), control variables, and socio-demographic data (i.e., gender, age, and migration background).[1] TOI was measured by the same three items used in Study 1, with a higher

---

[1] *Additional variables were included in the questionnaire, but are not relevant to this report: self-concept clarity, self-esteem, effort, domain identification, disidentification from the domain, perceived test difficulty, ethnic group identification.*

score indicating stronger entity theory endorsement. Stereotype endorsement was assessed with the item 'How do you personally evaluate the verbal ability of Turkish-origin migrants compared to Germans?' on a scale from $-3 = worse$ to $+3 = better$. Finally, the participants were debriefed and thanked for participation. The study took 45 min.

Verbal ability test

The test consisted of three tasks administered in German language. First, the participants worked on a reading comprehension task from the PISA test ('Police', Kunter *et al.*, 2002), in which they read a text and answered four multiple-choice questions. The PISA is an international study launched by the Organization of Economic Co-Operation and Development (OECD) to evaluate education systems. Starting in 2000, 15-year olds' competencies in reading, mathematics, and science are assessed every 3 years. Participants then completed two tasks (i.e., 'word endings' and 'four-word sentences') from the verbal creativity test (Schoppe, 1975). In the task 'word endings', participants generated as many German words as possible ending with specific letters (e.g., -tung/ -der). In the task 'four-word sentences', they formed meaningful and correct sentences with four words beginning with indicated letters (e.g., T-G-F-U).

The verbal ability test consisted of three subtasks reflecting performance situations in school as well as verbal skills needed in daily life. We created a heterogeneous measure to ensure high ecological validity, despite its relatively low internal consistency ($\alpha = .53$). To ensure that participants worked on all subtasks, time limits were introduced and students were told when they should move on to the next task. Test time was 17.5 min (7 min for the PISA task, 3 min for 'word endings', and 6 min for 'four-word sentences', plus 1.5 min for reading instructions and examples). Tasks were presented in counterbalanced order.

Control variables

To assess the moderating effect of TOI on stereotype activation effects over and above other individual differences in verbal performance, we controlled for previous verbal achievement (i.e., self-reported grades in German), domain-specific self-concept (four items, e.g., 'In general, I am good at German' $\alpha = .81$; adapted from PISA assessments, Kunter *et al.*, 2002; and the Self-Description Questionnaire, Marsh, 1990), and task motivation ('I was motivated when working on the test') in the analyses. These variables were chosen because they potentially contribute to the individual differences in task performance (Hattie & Anderman, 2013).

Results

Preliminary analyses

Because there are gender stereotypes related to verbal ability (girls are assumed to have higher verbal ability than boys, e.g., Steffens & Jelenec, 2011), we tested for gender as a potential confound. The analyses revealed a main effect (i.e., girls outperformed boys on the test, $F(1, 157) = 14.19$, $p < .001$, but no interaction with test diagnosticity, $F(1, 157) = 0.30$, $p = .584$. Thus, gender was not included in the subsequent moderator analysis.

Implicit theory of intelligence was assessed after the experimental manipulation of test diagnosticity. In order to include TOI as a moderator, we checked that TOI endorsement

was not influenced by stereotype activation. An ANOVA with ethnicity (German vs. Turkish migration background) and stereotype activation (diagnostic vs. non-diagnostic test) as factors and TOI as the dependent variable showed a significant main effect of ethnicity (i.e., Turkish-origin migrants showed higher entity endorsement, $F(1, 170) = 12.51$, $p < .001$, partial $\eta^2 = .07$, see Table 2). Importantly however, there was no main effect of experimental condition, $F(1, 170) = 0.38$, $p = .541$, nor an interaction between ethnicity and condition, $F(1, 170) = 0.01$, $p = .946$. Thus, we conclude that stereotype activation did not influence TOI and included it as a moderator variable in the following analyses. However, because of the mean differences in TOI between Germans and Turkish-origin migrants in Study 2, we group-mean-centred TOI before entering it as a predictor into the regression model.

Finally, because students were nested in classes, we computed intraclass correlations (ICCs) to check whether a substantial part of the variance in the dependent variable, the covariates, and/or the moderator was located between classes. The ICCs were very low for all variables in the model, ranging between .001 and .04. Therefore, we concluded that it was not necessary to account for the multilevel data structure in the subsequent regression analysis.

Moderator analysis

We computed multiple regression analysis with verbal performance as the dependent variable, controlling for previous verbal achievement, task motivation, and domain-specific self-concept. None of the control variables correlated significantly with entity theory endorsement (all ps > .069), and their means did not vary between experimental conditions or ethnic groups (all ps > .306). The regression model simultaneously included the control variables, their interactions with experimental condition (Yzerbyt, Muller, & Judd, 2004), the simple effects (dummy-coded variables for ethnicity [0 = German, 1 = Turkish-origin] and experimental condition [0 = non-diagnostic, 1 = diagnostic], as well as the group-mean-centred measure of TOI), all two-way interactions and the three-way interaction. The model was significant, $F(13, 145) = 3.61$, $p < .001$, $R^2 = .25$ (see Table 3). Previous verbal achievement significantly predicted

Table 3. Regression results for study 2

	B	SE (B)	β	t	p
Intercept	13.32	1.26		10.61	<.001
Self-concept	−0.74	0.80	−.12	−0.93	.354
Verbal ability	1.70	0.78	.29	2.17	.031
Task motivation	1.44	0.69	.23	2.07	.040
Self-concept × diagnosticity	1.30	1.09	.16	1.20	.233
Verbal ability × diagnosticity	−0.04	1.01	−.01	−0.04	.971
Task motivation × diagnosticity	0.05	0.98	.01	0.05	.962
TOI	−2.82	0.96	−.64	−2.94	.004
Diagnosticity	0.83	1.96	.07	0.49	.627
Ethnicity	−1.22	1.46	−.09	−0.83	.407
TOI × diagnosticity	3.63	1.38	.62	2.63	.009
TOI × ethnicity	2.22	1.13	.44	1.96	.051
Diagnosticity × ethnicity	−1.26	2.00	−.10	−0.63	.530
TOI × diagnosticity × ethnicity	−4.02	1.56	−.61	−2.57	.011

performance, $b = 1.70$, $t(145) = 2.17$, $SE = 0.78$, $p = .031$, and task motivation was also a significant predictor, $b = 1.44$, $t(145) = 2.07$, $SE = 0.69$, $p = .040$, as well as TOI, $b = -2.82$, $t(145) = -2.94$, $SE = 0.96$, $p = .004$. There was a significant two-way interaction between TOI and experimental condition, $b = 3.63$, $t(145) = 2.63$, $SE = 1.38$, $p = .01$, on performance. These effects were qualified by a significant three-way interaction between ethnicity, experimental condition, and TOI, $b = -4.02$, $t(145) = -2.57$, $SE = 1.56$, $p = .011$. In the non-diagnostic condition, the two-way interaction between TOI and ethnicity was significant, $b = 2.22$, $t(145) = 1.96$, $SE = 1.13$, $p = .051$; in the diagnostic condition, it was non-significant, $b = -1.80$, $t(145) = -1.66$, $SE = 1.08$, $p = .098$. Figure 1 depicts the interactive effect of the three predictors on test performance. Simple slopes analyses (Aiken & West, 1991; Preacher, Curran, & Bauer, 2006) revealed that in the diagnostic condition, TOI significantly predicted the performance of Turkish-origin students, $b = -0.99$, $t(145) = -2.03$, $SE = 0.49$, $p = .045$, but not of German students, $b = 0.81$, $t(145) = 0.82$, $SE = 0.99$, $p = .416$. As predicted, the more strongly Turkish-origin students endorsed an entity TOI, the lower was their performance. In contrast, we did not find the hypothesized positive relationship between entity TOI endorsement and performance after stereotype activation for the favourably stereotyped group of German students, although the trend
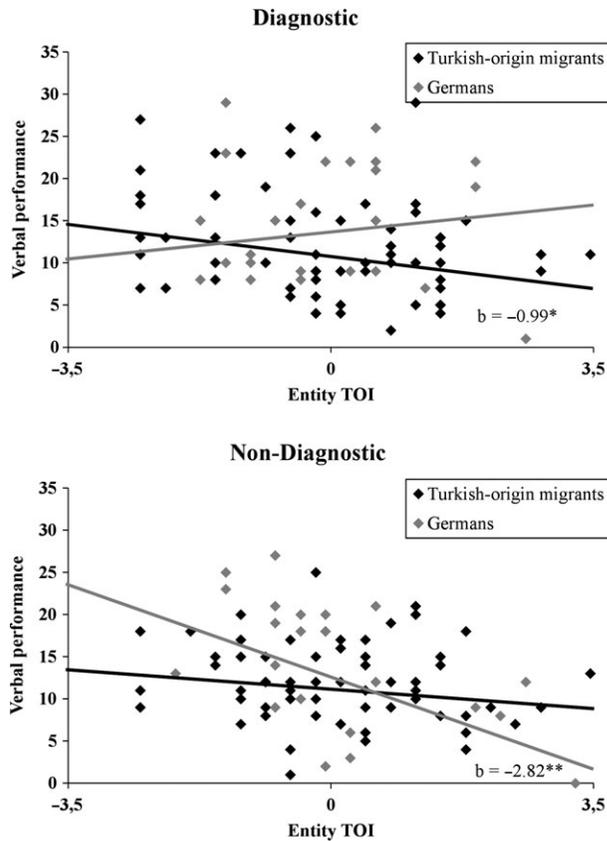

Figure 1. Implicit theory of intelligence moderates the effects of stereotype activation in Study 2 ($N = 173$). *$p \leq .05$; **$p \leq .01$.

was in that direction. Unexpectedly, in the non-diagnostic condition, TOI significantly predicted the performance of German students, $b = -2.82$, $t(145) = -2.94$, $SE = 0.96$, $p = .004$. The more strongly they endorsed an entity TOI, the lower was their performance. A visual inspection of the scatter plot of TOI and test performance showed that this effect was mainly driven by one participant who showed high entity endorsement and received zero points in the performance test. Inspection of the original test materials revealed that the student quit working on the test after the first two subtasks. When the participant was excluded from the analysis, the simple slope for Germans in the non-diagnostic condition failed to reach the significance level of 5%, $b = -2.26$, $t(145) = -1.92$, $SE = 1.18$, $p = .057$. The slope for Turkish-origin students in the control condition was not significant, $b = -0.60$, $t(145) = -1.04$, $SE = 0.58$, $p = .301$.[2]

Additional analyses

To rule out the possibility that an entity view of intelligence was associated with the stereotype about group differences in ability, we investigated the correlation between the two measures. Stereotype endorsement and entity theory endorsement were not correlated in the total sample ($r = .11$, $p = .153$), nor did any significant correlations emerge in separate analyses for ethnic groups and experimental conditions (all ps $> .153$).

Discussion

Study 2 provides evidence for the predicted moderation of stereotype activation effects by students' chronically endorsed implicit theory of intelligence. The more Turkish-origin migrants believed their intelligence was fixed, the worse they performed on a verbal test when negative ability-related stereotypes were activated (i.e., the test was described as diagnostic of verbal intelligence). Hence, when negative stereotypes against the in-group are salient, believing that one's intelligence is fixed predicts a more pronounced stereotype threat effect. However, although we expected to find evidence for a stereotype lift effect in the diagnostic condition for Germans who endorsed an entity theory, that effect was not found. This failure to confirm may have been due to the lower number of Germans in the sample compared to Turkish-origin migrants, thus limiting the statistical power to detect the hypothesized effect. As expected, TOI endorsement did not predict test performance for Turkish-origin migrants in the non-diagnostic condition. Unexpectedly, TOI endorsement negatively predicted performance for Germans in the non-diagnostic condition, but this effect seems to be driven by one extreme case and should be interpreted with caution. The influence of TOI on performance cannot be explained by stereotype endorsement, as the two measures are uncorrelated. This absence of a correlation strengthens our view that entity theory endorsement uniquely contributes to the explanation of group differences in performance after stereotype activation.

---

[2] *A regression model without any covariates showed similar results: a simple effect of TOI, b = −3.12, t(165) = −3.27, SE = 0.96, p = .001, two-way interactions between TOI and condition, b = 2.97, t(165) = 2.13, SE = 1.39, p = .034, as well as between TOI and ethnicity, b = 2.56, t(165) = 2.29, SE = 1.12, p = .023, and the three-way interaction between TOI, condition, and ethnicity, b = −3.50, t(165) = −2.19, SE = 1.59, p = .030. The simple slope for German students in the non-diagnostic condition was significant, b = −3.12, t(165) = −3.27, SE = 0.96, p = .001, as well as the slope for Turkish-origin students in the diagnostic condition, b = −1.09, t(165) = −2.13, SE = 0.51, p = .035.*

Although the results of Study 2 provide initial empirical evidence for the moderation hypothesis, the study has a methodological shortcoming: The moderator variable (i.e., TOI) was assessed after test diagnosticity was experimentally manipulated. Although TOI endorsement did not vary between experimental groups, we cannot fully rule out the possibility that it was influenced by the description of the test as diagnostic or non-diagnostic. Therefore, to replicate the results with a more unequivocal test of our hypothesis, in Study 3 students' TOI endorsement was assessed some days before the experimental manipulation of test diagnosticity.

STUDY 3

One hundred and ninety-ninth-grade students from 13 classes of German middle-track high schools (Realschule) in Mannheim, Cologne, and Hannover participated in their classrooms. Four participants were excluded from the analyses because of the lack of compliance. The final sample consisted of 186 participants (84 female, 3 missing). One hundred and twenty-four students had no migration background, while 62 had Turkish migration background. Age ranged from 14 to 18 years ($M = 15.21$, $SD = 0.62$).

Materials and procedure

The same 2 (stereotype activation: diagnostic vs. non-diagnostic test) $\times$ 2 (ethnicity: German vs. Turkish migration background) between-participants design was used. TOI was assessed as a continuous variable 1–3 days prior to the main study. Performance on the verbal ability test served as the dependent variable. In the main study, participants were randomly assigned to one of the two experimental conditions of stereotype activation (diagnostic: 71 Germans and 29 Turkish-origin migrants; non-diagnostic: 53 Germans and 33 Turkish-origin migrants).

The questionnaire content and procedure were similar to those of Study 2. Participants generated a code (i.e., first two letters of mother's name, first two letters of father's name, and month of birth) to ensure anonymity and to enable us to match the first and second questionnaires. As in Study 2, stereotype activation was manipulated by describing a test either as diagnostic or as non-diagnostic of verbal intelligence. The measure of verbal ability and the questionnaire assessing covariates and demographics were similar to those of Study 2.[3]

Results

As in Study 2, there was a main effect of gender on performance (i.e., girls outperformed boys, $F(1, 176) = 10.28$, $p = .002$), but no interaction with test diagnosticity, $F(1, 176) = 0.00$, $p = .993$; thus, gender was not included as a covariate in the subsequent analyses. We note that the students' mean performance in Study 3 was approximately nine points higher than in Study 2, likely due to the inclusion of lower school tracks in Study 2.

We again predicted a three-way interaction between TOI, stereotype activation, and ethnicity on performance. We expected to replicate the finding from Study 2 that entity

---

[3] *Additional variables not included in the analyses: pre-questionnaire: learning goals, effort beliefs; post-questionnaire: learning/ performance goals during test, entity TOI, learning goals, effort beliefs, effort during test, domain identification, perceived test difficulty, ethnic group identification.*

TOI negatively predicts performance for unfavourably stereotyped students in the diagnostic condition. We also hypothesized that entity TOI positively predicts performance for favourably stereotyped students in the diagnostic condition. Performances on the three subtasks were again aggregated in a single verbal ability score (α = .45).

We computed multiple regression analyses with test performance as the dependent variable. Analogous to Study 2, we controlled for previous verbal achievement, task motivation, and domain-specific self-concept (α = .83), in the analyses. Control variables were not significantly correlated with TOI endorsement (all ps > .083), and their means did not vary between experimental conditions or ethnic groups. One exception was a main effect of ethnicity on previous verbal achievement; Germans reported higher grades in German language courses than Turkish-origin migrants, $F(1, 176)$ = 4.17, p = .043, but ethnicity did not interact with experimental condition (all other ps > .265).

The regression model simultaneously included the covariates, their interactions with experimental condition, the simple effects (dummy-coded variables for ethnicity [0 = German, 1 = Turkish-origin] and experimental condition [0 = non-diagnostic, 1 = diagnostic], as well as the continuous group-mean-centred measure of TOI, α = .74), all two-way interactions, and the three-way interaction. The model was significant, $F(13, 149)$ = 3.84, p < .001, R^2 = .25 (see Table 4). Previous verbal achievement, b = 2.11, $t(149)$ = 2.47, SE = 0.86, p = .015, and task motivation, b = 1.92, $t(149)$ = 2.51, SE = 0.76, p = .013, significantly predicted performance. The results also showed a significant simple effect of experimental manipulation, b = 4.63, $t(149)$ = 3.58, SE = 1.29, p < .001. These effects were qualified by the predicted significant three-way interaction between ethnicity, condition, and TOI, b = −4.28, $t(149)$ = −2.40, SE = 1.78, p = .018. Figure 2 depicts the interactive effect of the three predictors on performance. In the non-diagnostic condition, the two-way interaction between TOI and ethnicity was not significant, b = 0.28, $t(149)$ = 0.23, SE = 1.24, p = .823; in the diagnostic condition, it was significant, b = −4.00, $t(149)$ = −3.11, SE = 1.28, p = .002. Simple slopes analyses (Aiken & West, 1991; Preacher *et al.*, 2006) revealed that the slope for Turkish-origin migrants in the diagnostic condition was similar to that of Study 2, b = −2.11, $t(149)$ = −1.96, SE = 1.08, p = .052, indicating again that greater

Table 4. Regression results for study 3

|	B	SE (B)	β	t	p
Intercept	19.88	0.94		21.08	<.001
Self-concept	−0.60	0.94	−.08	−0.64	.523
Verbal ability	2.11	0.86	−.29	−2.47	.015
Task motivation	1.92	0.76	.25	2.51	.013
Self-concept × diagnosticity	−1.00	1.30	−.10	−0.77	.443
Verbal ability × diagnosticity	−0.58	1.28	−.05	−0.45	.653
Task motivation × diagnosticity	−0.70	1.18	−.06	−0.59	.556
TOI	0.75	0.70	.13	1.08	.284
Diagnosticity	4.63	1.29	.31	3.58	<.001
Ethnicity	2.04	1.62	.13	1.26	.208
TOI × diagnosticity	1.13	1.01	.14	1.13	.262
TOI × ethnicity	0.28	1.24	.03	0.23	.823
Diagnosticity × ethnicity	−4.23	2.29	−.21	−1.85	.067
TOI × diagnosticity × ethnicity	−4.28	1.78	−.29	−2.40	.018
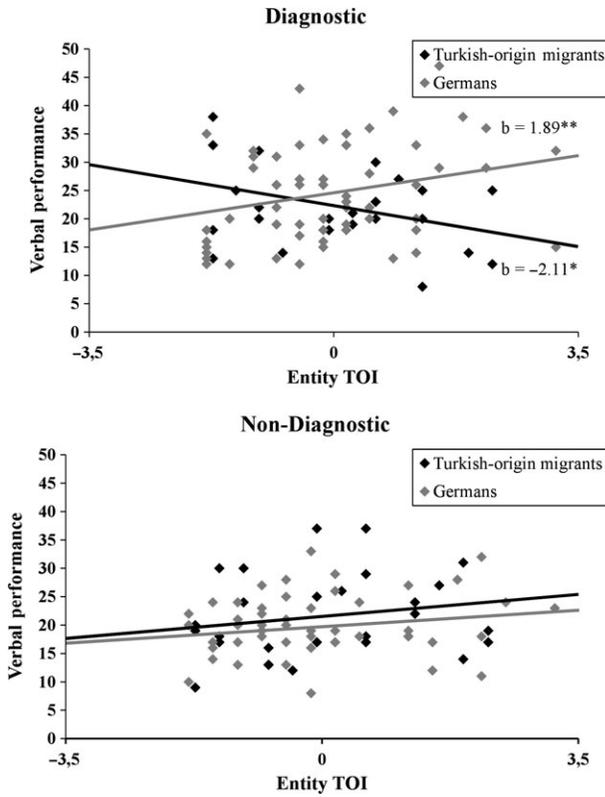
Figure 2. Implicit theory of intelligence moderates the effects of stereotype activation in Study 3 (N = 186). *$p \leq$ .05; **$p \leq$ .01.

endorsement of an entity TOI was associated with lower verbal performance. Furthermore, in this study, the predicted positive relationship for German students with strong entity beliefs was found in the diagnostic condition, b = 1.89, $t(149)$ = 2.61, SE = 0.72, p = .010. The more strongly German students endorsed an entity TOI, the better they performed on the test. Neither of the slopes for Germans or Turkish-origin migrants in the non-diagnostic condition were significant, ps > .283.[4]

As in Study 2, we computed ICCs to analyse the proportions of variance within and between classes. The ICCs of test performance, task motivation, and self-concept were within the acceptable range (between .07 and .10, e.g., Lee, 2000), but the ICCs for verbal ability (.12) and TOI (.14) were slightly above the recommended cut-off of .10. Therefore, we additionally computed hierarchical linear regression with a random-intercept model to account for the nested data structure. Because multilevel regression is recommended only with samples with a high number of level-2 units (at least 20–30; e.g., Bell, Morgan, Kromrey, & Ferron, 2010; Maas & Hox, 2005; Snijders & Bosker, 2012), and in Study 3 we

---

[4] *A regression model without any covariates showed mostly similar results: a simple effect of condition, b = 3.84, $t(160)$ = 2.91, SE = 1.32, p = .004, and the three-way interaction between TOI, condition, and ethnicity, b = −4.29, $t(160)$ = −2.33, SE = 1.84, p = .021. The simple slope for German students in the diagnostic condition was significant, b = 2.14, $t(160)$ = 2.88, SE = 0.74, p = .005, and the slope for Turkish-origin students in the diagnostic condition showed a non-significant trend, b = −1.77, $t(160)$ = −1.58, SE = 1.12, p = .117.*

had only 13 classes, the results should be interpreted with caution. The findings were largely similar to the regression not taking the nested data structure into account in that the three-way interaction between TOI, diagnosticity, and ethnicity was significant; detailed results are displayed in the Appendix B.

Discussion

In Study 3, we replicated the findings from Study 2: When the test was described as diagnostic of verbal ability, the more Turkish-origin students endorsed an entity TOI, the worse they performed. Furthermore, we found the predicted reversed effect for Germans in the diagnostic condition: The more they endorsed an entity TOI, the better they performed. Importantly, in Study 3, the predicted pattern emerged when TOI endorsement was assessed prior to the experimental manipulation of test diagnosticity and the performance test. This finding rules out the possibility that it was influenced by stereotype activation or characteristics of the verbal test (which was not supported in Study 2 either). Even more importantly, this result shows that students' dispositional beliefs about the fixedness of their intelligence, even when measured in a completely different situation, can predict performance when they are exposed to an ability-related stereotype about their in-group. In summary, TOI as a dispositional variable can predict the extent of stereotype threat and stereotype lift effects.

GENERAL DISCUSSION

We investigated implicit theories of intelligence (Study 1) and their interaction with ability-related stereotypes as predictors of verbal performance of German and Turkish-origin students (Studies 2 and 3). When negative stereotypes were activated by describing a task as diagnostic of verbal intelligence, entity theory endorsement was related to higher performance for favourably stereotyped group members (Study 3) and to lower performance for negatively stereotyped group members (Studies 2 and 3). Thus, the effect of TOI on performance differs as a function of group membership (i.e., belonging to the negatively stereotyped group of Turkish-origin migrants or the favourably stereotyped group of Germans). For Germans, higher entity theory endorsement seems to be advantageous for performance; for Turkish-origin migrants, it appears to be detrimental.

Interaction between stereotype activation and implicit theory of intelligence

It has been hypothesized that stereotype threat and implicit theory of intelligence interact to predict the performance of negatively stereotyped group members (e.g., Aronson *et al.*, 2002; Dweck, Mangels, & Good, 2004; Good *et al.*, 2003). The present research contributes to this earlier work by empirically investigating the interactive effect. For the first time, an experimental activation of ability-related stereotypes was combined with the assessment of students' dispositional TOI and a standardized performance measure.

For negatively stereotyped group members, we predicted that the description of a test as diagnostic of intelligence would indicate a performance goal and at the same time activate the threat that the ability of one's group is low. Thus, reminders of past failure on the group level may create a risk of future failure on the individual level. The results of the present studies support this line of reasoning. Failing in the task at hand would likely lead group members with an entity view to the conclusion that they are not smart. Such a

threatening scenario increases negative cognitions and worries, reducing working-memory capacity (Schmader *et al.*, 2008) and decreasing performance. In other words, a dispositional belief in the fixedness of intelligence enhances vulnerability to stereotype threat effects.

For favourably stereotyped students, however, describing a verbal test as diagnostic results in higher performance for those with an entity view (Study 3). This finding is in line with earlier work (e.g., Mendoza-Denton *et al.*, 2008) and with our theoretical assumption that when confronted with a favourable stereotype, entity theorists will show a performance boost, assuring them that their membership in a group that is expected to do well predicts their own future success. This pattern was significant only in Study 3; however, descriptively the slope in Study 2 pointed in the same direction. One explanation for the non-significant slope in Study 2 could be the relatively smaller number of German students in that sample. As gaining precise information about the ethnic composition of classes before recruitment was difficult, in Study 2 the sample consisted of more Turkish-origin migrants than the sample of Study 3, but at the same time of less Germans. The lesser statistical power would make it more difficult to detect the moderated stereotype lift effect. Future research should test how reliable this effect is and investigate its underlying processes.

The present results also showed that when the test was described as non-diagnostic of intelligence (i.e., no stereotype activation), TOI endorsement did not predict test performance. This description of the test as under development implied that students' test performance is primarily relevant for test constructors and that performance does not provide any information for the students about their ability levels. Therefore, their belief about the fixedness of intelligence was not expected to be related to performance.

Theoretical implications

We measured students' TOI as a dispositional trait instead of a state-specific manipulation, combined with an experimental activation of ability-related stereotypes and a standardized performance measure. Thus, the present research investigated the previously proposed moderator effect of TOI on stereotype threat and stereotype lift effects (e.g., Aronson *et al.*, 2002; Good *et al.*, 2003; Mendoza-Denton *et al.*, 2008). Dispositional TOI can predict which individuals are prone to stereotype threat and stereotype lift effects, thereby contributing to the explanation of the high variability in effect sizes found in meta-analyses of stereotype activation effects (e.g., Nguyen & Ryan, 2008; Walton & Cohen, 2003). The present research has identified a single variable (i.e., TOI endorsement) that moderates both stereotype threat and stereotype lift effects, pointing to a potential unifying process. Considerable research has investigated the processes of stereotype threat effects (for a review, see Schmader & Beilock, 2012), but the processes of stereotype lift effects are less well understood (e.g., Chatard *et al.*, 2008; Walton & Cohen, 2003). Implicit theories of intelligence might shape the degree to which performance expectations about groups conveyed by ability-related stereotypes are considered relevant to one's own performance. Because entity theorists want to determine their ability level, they might perceive their in-group's assumed ability level as a reasonable gauge of their own expected performance. Previous research has shown that performance expectations mediate stereotype threat effects for minority group members (e.g., Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003); future studies should investigate whether implicit theories of intelligence can further specify which individuals are most susceptible to these effects.

In sum, by measuring TOI as a trait variable, this research shed light on how students high or low in entity theory endorsement perceive performance situations and cope with the presence of group-related ability stereotypes. The fact that dispositional TOI seems to be relevant in the context of ability stereotypes about social groups creates further research possibilities. For example, the interplay between ability stereotypes and students' trait TOI could be assessed in longitudinal study designs to investigate whether repeated stereotype activation and TOI endorsement mutually influence each other.

Measuring implicit theory of intelligence as a dispositional variable rather than influencing it by experimental manipulation or intervention raises the question of how far TOI is shaped by demographic or context variables. The present research showed that the endorsement of ability-related stereotypes was unrelated to TOI. Further, in contrast to Aronson *et al.* (2002), we did not find increased incremental theory endorsement as an identity-management strategy for negatively stereotyped group members. It may be the case that not only ethnicity and ethnic stereotypes shape TOI, but also social class might be a related variable (e.g., Croizet & Claire, 1998). Study 2, which investigated schools from the merged lower and middle track, showed higher entity endorsement among Turkish-origin students than among Germans. Rheinschmidt and Mendoza-Denton (2014) showed that an entity view interacted with rejection sensitivity based on social class to predict college performance. As Turkish-origin migrants are overrepresented in the lower track as well as the lower social class in Germany (e.g., Matzner, 2012), future research should investigate the interrelation of these variables in predicting outcomes in the German educational context.

Limitations

First, the measure created to reflect heterogeneous verbal abilities needed in everyday school situations had a comparatively low internal consistency. Second, there were limitations due to the samples collected in school settings: The unequal number of German and Turkish-origin participants in the samples resulted in varying statistical test power. It could not be investigated whether the effects found for Turkish-origin migrants would extend to other negatively stereotyped immigrant groups, because these groups were too small for quantitative analyses. These limitations are results of our approach to investigate stereotype activation effects in real-world settings to ensure high ecological validity. Further studies should replicate our results in both natural and more controlled settings to ensure their robustness.

Implications for educational settings

In the German educational system, the differences between German and Turkish-origin students have been found on academic achievement and standardized test performance. Sociological and educational research has identified the factors such as socioeconomic status or language spoken at home as potential reasons for these differences (e.g., Bos, 2012a,b; Diefenbach, 2011; Klieme *et al.*, 2010). However, a substantial amount of variance still remains unexplained when controlling for these demographic variables (e.g., Neumann & Schneider, 2011). The present research introduces evidence for social–psychological factors contributing to the explanation of ethnic performance differences. Because in educational settings tests are assumed to be diagnostic of ability, these settings are likely to activate performance goals and

ability-related stereotypes about different groups. It is therefore important to develop more stereotype-neutral testing situations. Recent research has shown that Turkish-origin migrants are negatively stereotyped in the domain of academic competence in German society and especially among student teachers (Froehlich *et al.*, 2016). Therefore, it is very likely that Turkish-origin migrants are confronted with negative stereotypes in educational settings, and thus, the findings from stereotype threat research (e.g., Inzlicht & Schmader, 2012) need to be applied to the educational system. Specifically, in Germany, Turkish-origin students who endorse an entity theory of intelligence seem to be especially vulnerable to stereotype threat effects.

Acknowledgements

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Appel, M. (2012). Anti-immigrant propaganda by radical right parties and the intellectual performance of adolescents. *Political Psychology*, *33*, 483–493. doi:10.1111/j.1467-9221.2012.00902.x

Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in Psychology*, *6*, 900. doi:10.3389/fpsyg.2015.00900

Aronson, J. (Ed.) (2002). *Improving academic achievement: Impact of psychological factors on education* (6th ed.). San Diego, CA: Academic Press.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*, 113–125. doi:10.1006/jesp.2001.1491

Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology*, *41*, 76–81. doi:10.1027/1864-9335/a000011

Bell, B. A., Morgan, G. B., Kromrey, J. D., & Ferron, J. M. (2010). The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors. *Proceedings of the Joint Statistical Meetings – Section on Survey Research Methods*, *2010*, 4057–4067.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*, 246–263. doi:10.1111/j.1467-8624.2007.00995.x

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, *12*, 225–229. doi:10.1111/1467-9280.00340

Bos, W. (2012a). *IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [IGLU 2011: Reading competencies of elementary school children in Germany and in international comparison]. Münster, Germany [u.a.]: Waxmann.

Bos, W. (2012b). *TIMSS 2011: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [TIMSS 2011: Mathematics and science competencies of elementary school children in Germany and in international comparison]. Münster, Germany [u.a.]: Waxmann.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research. Methodology in the social sciences*. New York: Guilford Press.

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on Raven's advanced progressive matrices. *Journal of Applied Psychology*, *91*, 979–985. doi:10.1037/0021-9010.91.4.979

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology*, *76*, 246–257. doi:10.1037/0022-3514.76.2.246

Bundesamt für Migration und Flüchtlinge [Federal Office for Migration and Refugees] (2010). Migrationsbericht 2010. *Migrationsbericht des Bundesamtes für Migration und Flüchtlinge im Auftrag der Bundesregierung. Bundesministerium des Innern* [Migration Report 2010. Migration Report of the Federal Office for Migration and Refugees on Behalf of the Federal Government. Federal Ministry of the Interior].

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, *33*, 267–285. doi:10.1002/ejsp.145

Chatard, A., Selimbegović, L., Konan, P., & Mugny, G. (2008). Performance boosts in the classroom: Stereotype endorsement and prejudice moderate stereotype lift. *Journal of Experimental Social Psychology*, *44*, 1421–1424. doi:10.1016/j.jesp.2008.05.004

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*, 588–594. doi:10.1177/0146167298246003

Croizet, J. C., & Millet, M. (2012). Social class and test performance: From stereotype threat to symbolic violence and vice versa. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 188–201). New York, NY: Oxford University Press.

Deaux, K., Bikmen, N., Gilkes, A., Ventuneac, A., Joseph, Y., Payne, Y. A., & Steele, C. M. (2007). Becoming American: Stereotype threat effects in Afro-Caribbean immigrant groups. *Social Psychology Quarterly*, *70*, 384–404. doi:10.1177/019027250707000408

Diefenbach, H. (2011). Der Bildungserfolg von Schülern mit Migrationshintergrund im Vergleich zu Schülern ohne Migrationshintergrund. [Educational success of students with migration background compared to students without migration background]. In R. Becker (Ed.), *Lehrbuch der Bildungssoziologie* [*Sociology of Education*] (pp. 449–473). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York: Random House.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273. doi:10.1037/0033-295X.95.2.256

Dweck, C. S., Mangels, J. A., & Good, C. (2004). Motivational effects on attention, cognition, and performance. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition. Integrative perspectives on intellectual functioning and development* (pp. 41–55). Mahwah, NJ: Lawrence Erlbaum.

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*, 74–86. doi:10.1027/1864-9335/a000260

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology*, *24*, 645–662. doi:10.1016/j.appdev.2003.09.002

Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology*, *85*, 541–553. doi:10.1037/0022-3514.85.3.541

Hattie, J., & Anderman, M. (Eds.) (2013). *International guide to student achievement*. New York: Routledge.

Herbert, U. (2001). *Geschichte der Ausländerpolitik in Deutschland: Saisonarbeiter, Zwangsarbeiter, Gastarbeiter, Flüchtlinge* [History of immigration policy in Germany: Seasonal workers, forced laborers, guest workers, and refugees]. München, Germany: Beck.

Inzlicht, M., & Schmader, T. (Eds.) (2012). *Stereotype threat: Theory, process, and application*. New York: Oxford University Press.

Kahraman, B., & Knoblich, G. (2000). "Stechen statt Sprechen": Valenz und Aktivierbarkeit von Stereotypen über Türken. ["Stereotypes about Turks": Valence and activation]. *Zeitschrift für Sozialpsychologie*, *31*, 31–43. doi:10.1024/0044-3514.31.1.31

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., & Prenzel, M., . . . Stanat, P. (Eds.). (2010). *PISA 2009: Bilanz nach einem Jahrzehnt* [PISA 2009: Review after one decade]. Münster, Germany: Waxmann Verlag.

Kline, R. B. (2011). *Principles and practice of structural equation modeling. Methodology in the social sciences*. New York: Guilford Press.

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., . . . Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente* [PISA 2000: Documentation of Scales]. Berlin, Germany: Heenemann GmbH.

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, *35*, 125–141. doi:10.1207/S15326985EP3502_6

Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model. *International Journal of Intercultural Relations*, *30*, 751–768. doi:10.1016/j.ijintrel.2006.06.005

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology*, *74*, 1421–1436. doi:10.1037/0022-3514.74.6.1421

Little, T. D. (2013). *Longitudinal structural equation modeling. Methodology in the social sciences*. New York: The Guilford Press.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92. doi:10.1027/1614-2241.1.3.86

Marsh, H. W. (1990). *Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph*. Macarthur, NSW, Australia: University of Western Sydney.

Martiny, S. E., & Götz, T. (2011). Stereotype Threat in Lern- und Leistungssituationen: Theoretische Ansätze, empirische Befunde und praktische Implikationen. [Stereotype threat in learning and achievement situations. Theoretical approaches, empirical results and practical implications]. In M. Dresel & L. Lämmle (Eds.), *Talentförderung – Expertiseentwicklung – Leistungsexzellenz: Vol. 9. Motivation, Selbstregulation und Leistungsexzellenz [Talent promotion, expert development, performance excellence: Vol. 9. Motivation, self-regulation and performance excellence]* (pp. 153–177). Münster, Germany: LIT.

Martiny, S. E., Mok, S. Y., Deaux, K., & Froehlich, L. (2015). Effects of activating negative stereotypes about Turkish-origin students on performance and identity management in German high schools. *International Review of Social Psychology*, *3*, 205–225.

Matzner, M. (2012). Migration und Bildung: Daten und Fakten. [Migration and education: Data and facts]. In M. Matzner (Ed.), *Handbuch Migration und Bildung [Handbook migration and education]* (1st ed., pp. 90–101). Weinheim, Germany: Beltz.

Mendoza-Denton, R., Kahn, K., & Chan, W. (2008). Can fixed views of ability boost performance in the context of favorable stereotypes? *Journal of Experimental Social Psychology*, *44*, 1187–1193. doi:10.1016/j.jesp.2008.03.005

Molden, D. C., & Dweck, C. S. (2006). Finding "meaning" in psychology: A lay theories approach to self-regulation, social perception, and social development. *American Psychologist*, *61*, 192–203. doi:10.1037/0003-066X.61.3.192

Neumann, U., & Schneider, J. (Eds.). (2011). *Schule mit Migrationshintergrund* [School with migration background]. Münster, Germany: Waxmann.

Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314–1334. doi:10.1037/a0012702

OECD (2010). *Closing the gap for immigrant students: Policies, practice, and performance*. Paris, France: Author.

OECD (2015). Can the performance gap between immigrant and non-immigrant students be closed? *PISA in Focus*, *53*, 1–4. doi:10.1787/22260919

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448. doi:10.3102/10769986031004437

Rheinschmidt, M. L., & Mendoza-Denton, R. (2014). Social class and academic achievement in college: The interplay of rejection sensitivity and entity beliefs. *Journal of Personality and Social Psychology*, *107*, 101–121. doi:10.1037/a0036553

Ruehl, S., & Babka von Gostomski, C. (2012). Menschen mit Migrationshintergrund in Deutschland: Daten und Fakten [People with migration background in Germany: Data and facts]. In M. Matzner (Ed.), *Handbuch Migration und Bildung* (1st ed., pp. 22–37). Weinheim, Germany: Beltz.

Schmader, T., & Beilock, S. L. (2012). An integration of processes that underlie stereotype threat. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 34–50). New York, NY: Oxford University Press.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356. doi:10.1037/0033-295X.115.2.336

Schoppe, K. J. (1975) *Verbaler Kreativitäts-Test*. [Verbal Creativity Test]. Göttingen: Hogrefe Verlag.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: Sage.

Statistisches Bundesamt [Federal Statistical Office] (2012). *Bevölkerung und Erwerbstätigkeit 2011. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2011* [Population and Employment 2011. Population with Migration Background – Results of the Microcensus 2011]. Fachserie 1 Reihe 2.2. Wiesbaden, Germany: Author.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811. doi:10.1037/0022-3514.69.5.797

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, *64*, 324–335. doi:10.1007/s11199-010-9924-x

van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research. Cross-cultural psychology series: V. 1*. Thousand Oaks, CA: Sage.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467. doi:10.1016/S0022-1031(03)00019-2

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, *40*, 424–431. doi:10.1016/j.jesp.2003.10.001

Appendix A: Confirmatory factor analysis in study 1

To further test measurement invariance for Germans and Turkish-origin migrants, we conducted CFA with multiple-group comparison. The results should be interpreted with caution due to the relatively small sample sizes ($N = 61$ for Germans and $N = 47$ for Turkish-origin migrants). We used a step-up procedure to test for

different types of measurement invariance (e.g., Brown, 2006). In the first step, we tested for configural invariance by investigating the measurement model in both subsamples separately. Factor loadings were not restricted, and the variance of the latent variable was fixed to 1. For Turkish-origin migrants, the measurement model had a good fit (χ^2 = 18.960, df = 24, p = .754; CFI = 1.000; RMSEA = .000 [90% CI = .000/.087]; SRMR = .079). However, for Germans, the measurement model fit the data less well (χ^2 = 29.393, df = 24, p = .206; CFI = .095; RMSEA = .061 [90% CI = .000/.126]; SRMR = .080), and the solution was not reliable due to a non-positive definite psi matrix involving the effort factor. Likely, the problem occurred because of the three effort items: One was framed positively (i.e., high values indicate high effort beliefs) and two were framed negatively (i.e., high values indicate low effort beliefs). Similarly, the three effort items did not form a consistent factor in the exploratory factor analysis for Germans as well. Therefore, we performed another CFA without the effort items and specified only the entity and goal orientation factors. The measurement models with two factors had an acceptable fit in both subsamples (for Turkish-origin migrants: χ^2 = 7.068, df = 8, p = .529; CFI = 1.000; RMSEA = .000 [90% CI = .000/.159]; SRMR = .040; for Germans: χ^2 = 10.552, df = 8, p = .228; CFI = .970; RMSEA = .072 [90% CI = .000/.179]; SRMR = .047). In a second step, we estimated an unrestricted baseline model of the multiple-group comparison by allowing the model parameters to be freely estimated in both subgroups. The model also had an acceptable fit (χ^2 = 17.620, df = 16, p = .347; CFI = .989; RMSEA = .043 [90% CI = .000/.137]; SRMR = .044). In a third step, we estimated a model in which the factor loadings were fixed between the two subgroups. This model again showed an acceptable fit (χ^2 = 22.996, df = 22, p = .402; CFI = .994; RMSEA = .029 [90% CI = .000/.119]; SRMR = .087). Finally, we compared the fit of the baseline model and the more restricted model with a χ^2 difference test, which was not significant, χ^2_{diff} = 5.38, df_{diff} = 6, p = .497. This indicates that the more restricted model does not fit the data worse and is therefore preferable to the baseline model. Thus, metric invariance for the two factors entity and goal orientation can be assumed between Germans and Turkish-origin migrants. The results of the CFA are in line with the results of the exploratory factor analysis and strengthen the finding that entity theory endorsement can be used as a moderator of stereotype activation effects in the subsequent studies.

Appendix B: Hierarchical linear regression in study 3

We computed a random-intercept model to account for the fact that students were nested in classes. All predictors and covariates were entered at level 1 (the student level), and the variance of the intercept was allowed to vary randomly. Regression results are displayed in Table B1. Simple slopes analyses (Aiken & West, 1991; Preacher *et al.*, 2006) showed that the slope for Turkish-origin migrants in the diagnostic condition was negative and marginally significant, b = −1.95, SE = 1.08, t (149) = −1.80, p = .073. In contrast to the non-multilevel model, the slope for Germans in the diagnostic condition was not significant, b = 0.86, SE = 0.95, t (149) = 0.90, p = .367, but the slope for Turkish-origin students in the non-diagnostic condition was positive and marginally significant, b = 1.74, SE = 0.91, $t(149)$ = 1.92, p = .057. The slope for Germans in the non-diagnostic condition was not significant, b = 0.55, SE = 0.32, $t(149)$ = 1.68, p = .194.

Table B1. Hierarchical linear regression results for study 3

Fixed effects	Coefficient	SE	t	p
Intercept	24.26	5.21	4.66	<.001
Self-concept	−0.12	0.69	−0.18	.860
Verbal ability	−2.44	1.04	−2.34	.021
Task motivation	0.98	0.43	2.67	.025
Self-concept × diagnosticity	−0.84	0.95	−0.89	.375
Verbal ability × diagnosticity	−0.63	1.55	−0.41	.685
Task motivation × diagnosticity	−0.34	0.66	−0.51	.610
TOI	0.55	0.71	0.77	.441
Diagnosticity	11.12	7.49	1.49	.140
Ethnicity	1.95	1.63	1.20	.234
TOI × diagnosticity	1.20	1.00	1.19	.236
TOI × ethnicity	0.32	1.23	0.26	.798
Diagnosticity × ethnicity	−4.04	2.31	−1.75	.082
TOI × diagnosticity × ethnicity	−4.00	1.80	−2.23	.027
Variance component	Coefficient	df	χ^2	p
Intercept variance	2.31	12	19.67	.073

Note. $N = 186$ at level 1 and $N = 13$ at level 2.