Received: 5 July 2024 | Revised: 11 October 2024 | Accepted: 14 October 2024

DOI: 10.1111/josi.12649

ORIGINAL ARTICLE

JOURNAL OF Social Issues
A Journal of the Society for the Psychological Study of Social Issues

WILEY

Are low-ability students mentally represented as low-SES, academically incapable, and undeserving of support?

Alexander S. Browman[1] | **David B. Miele**[2]

[1]Department of Psychology, College of the Holy Cross, Worcester, Massachusetts, USA

[2]Department of Counseling, Developmental, and Educational Psychology, Lynch School of Education and Human Development, Boston College, Chestnut Hill, Massachusetts, USA

Correspondence
Alexander S. Browman, Department of Psychology, College of the Holy Cross, 1 College Street, Worcester, MA 01610, USA.
Email: abrowman@holycross.edu

Funding information
College of the Holy Cross; James S. McDonnell Foundation, Collaborative, Grant/Award Number: 220020483

Abstract

In seven studies, this research demonstrates that both the general public and educators may hold culturally-shared, class stereotype-laden mental representations that they reflexively use both to subjectively identify particular students as being high or low in academic ability, and determine who should receive educational support. Using procedures designed to capture people's mental images of others, we first observed that both the general public and aspiring educators mentally represent low-ability students as qualitatively and quantitatively distinct from high-ability students. Furthermore, the representations of low (vs. high) ability students captured from the public and aspiring educators were more likely to be associated with negative class-based academic stereotypes by separate samples of the public and educators, such that a student who "looks" low in ability was also more likely to be labeled as being low-SES, and having poorer academic motivation and work ethic. As a result, the low (vs. high) ability student was more likely to be denied college admissions or scholarship support by members of the American public and to be exposed to unsupportive instructional practices by teachers. Implications for our understanding of teacher biases are discussed.

INTRODUCTION

Americans have long expressed a public belief that the distribution of educational opportunities should be based solely on formal metrics of students' academic skills and abilities. For example, across multiple recent surveys, high school course selection, grades, and standardized test scores were the only factors that a majority of Americans said should play a major role in the university admissions process. By contrast, a majority felt that factors like extracurricular activities, special talents in the arts or athletics, race, gender, and family background should not be considered in making such decisions (Gallup News, 2016; Pew Research Center, 2019; The Associated Press-NORC Center for Public Affairs Research, 2019).

Despite this consensus, it is well established that students are frequently judged based on non-formal metrics. For example, educators' academic *expectations* for their students can lead them to interact with individual pupils in expectation-reinforcing ways, which may result in self-fulfilling prophecy effects. For example, teachers may provide less advancement opportunities or helpful instructional practices to students they perceive as less capable, thereby producing worse academic outcomes for those students (for reviews, see Good et al., 2018; Wang et al., 2018). And critically, just like other members of societies (Fiske et al., 2002), educators' expectations for a student may be influenced by factors that are unrelated to that student's recent academic performance, such as culturally-held beliefs and stereotypes about academic ability for members of different social groups (see Wang et al., 2018).

One highly influential social category is socioeconomic status (SES). Numerous studies have confirmed that both educators and the public tend to have lower academic expectations for individuals from lower (vs. higher) SES backgrounds (Cozzarelli et al., 2001; Doyle & Easterbrook, 2024; Fiske et al., 2002; Wang et al., 2018; Woods et al., 2005). In fact, several of these studies have found that people show these biases in their expectations of lower and higher SES students even when they are not aware of the students' objective SES or academic performance history (see Wang et al., 2018).

The present research explores one potential psychological mechanism that individuals may use in forming their academic expectations in the absence of objective performance or SES information about a student. Specifically, we propose that people operating within the culture of the American educational system may hold shared mental representations of what students with high and low academic abilities look like. This possibility is consistent with research in cognitive and social psychology showing that people form assumptions about an individual's internal characteristics (e.g., trustworthiness, religiosity, intelligence) based on facial cues that may be separate from those associated with race, ethnicity, and gender (Kleisner et al., 2014; Todorov et al., 2005; Zebrowitz et al., 2002). Furthermore, different people within a culture have been found to respond similarly to the same facial cues (Dotsch et al., 2008; Imhoff et al., 2011). This suggests that a given society might share similar mental representations of what a trustworthy, religious, or intelligent person looks like.

Given the pervasiveness of SES-based intellectual stereotypes across many societies (e.g., Cuddy et al., 2009), it follows that many Americans' mental representations of high- and low-ability students may also be imbued with cues that signal SES. Indeed, research has found that raters can reliably determine a target's membership in less visible social groups, such as SES and sexual orientation, from facial cues alone (Becker et al., 2017; Bjornsdottir & Rule, 2017, 2020; Paul et al., 2022; Rule et al., 2008; Schmid Mast & Hall, 2004; Zhang et al., 2021). Educators and members of the general public operating within the culture of the American educational system

JOURNAL OF Social Issues A Journal of the Society for the Psychological Study of Social Issues

may therefore infer both how academically capable students are *and* their SES from aspects of their facial appearance.

Furthermore, research suggests that how people not only evaluate, but also *treat*, a particular person may depend in part on whether that person's appearance matches mental representations that are linked to a stereotyped social group. For example, Brown-Iannuzzi and colleagues (2017) found that Americans' mental representations of welfare recipients (vs. non-welfare recipients) were more likely to perceived as African-American, which caused them to be less supportive of giving that person welfare benefits (see also Brown-Iannuzzi et al., 2018). Therefore, if the mental representations of high- and low-ability students that people hold are imbued with cues that signal SES, this may lead them to treat students who look similar to these representations in ways that reinforce educational disparities – possibly without their conscious awareness or intention (Doyle et al., 2024). Indeed, teachers and members of the public have been shown to systematically make tracking recommendations that favor higher-SES students over lower-SES students, even when they have similar records of past performance (Batruch et al., 2019, 2023). They also grade low-SES students who, contrary to stereotypes, perform well on assessments more harshly than high-SES students who perform equally well (Batruch et al., 2017; Doyle et al., 2023). This is especially likely to occur when the person carrying out the assessment believes that the primary purpose of school is to compare and rank students (Autin et al., 2019).

Of course, as with other biased tendencies (Gawronski & Hahn, 2018), if asked directly, people are likely to dismiss or conceal the possibility that they would evaluate students' intellectual abilities based on what they look like (Doyle et al., 2024). Indeed, teachers in the present work said exactly this (see General Discussion). To address this issue, the present work employed a sophisticated image-generation technique that can capture the mental representations people associate with particular non-physical characteristics in a more implicit, self-presentation-resistant manner. We used this technique to address three questions. First, in general, do people operating within the culture of the American educational system – including both members of the general public and those more directly connected to the education system – share qualitatively distinct mental representations of what lower versus higher ability students look like (Research Question [RQ] 1)? If so, second, are mental representations of low-ability students more closely associated with negative academic attributes and social categories that are associated with negative academic stereotypes (RQ2)? For example, is a student who "looks" low in ability also perceived to have poorer academic motivation and work ethic, and to be lower in SES? Third, do these representations influence the level of support that both the public and educators are willing to provide these students (RQ3)?

Reflections on the present samples and on researcher positionality

We note some elements of the samples examined in the present research, and of the positionality of the researchers who conducted it, that are important for understanding the potential limits on its generality. First, the present studies were conducted with three types of samples: members of the general public recruited on Amazon's Mechanical Turk (MTurk; mturk.com) and CloudResearch's Connect (connect.cloudresearch.com) crowdsourcing networks (Studies 1-5); aspiring educators enrolled at a school of education at an elite, private, predominantly White university in the Northeastern U.S. (Study 2); and in-service elementary school teachers working in a public, suburban school district in the Northeastern U.S. (Study 6). Table 1 presents the representation of each study sample in terms of participants' gender, age, race, income, subjective SES,

TABLE 1 Participant demographics.

	Image generation		Image rating					
	Study 1[1]	**Study 2**[2]	**Study 1**[1]	**Study 2**[1]	**Study 3**[1]	**Study 4**[1]	**Study 5**[1]	**Study 6**[1]
Final *N*	486	134	196	199	297	287	307	41
Male	55.3%	17.2%	61.2%	59.3%	58.6%	52.3%	57.3%	4.9%
Female	44.7%	82.1%	38.8%	39.7%	41.4%	46.3%	41.7%	90.2%
Non-binary or gender fluid	0%	.7%	0%	.5%	0%	.3%	1.0%	0%
Undisclosed	0%	0%	0%	1%	0%	3%	0%	2%
Age [*M* (*SD*)]	40.1 (12.4)	19.1 (1.0)	36.3 (11.3)	35.3 (9.7)	36.8 (10.5)	37.0 (11.1)	40.3 (13.2)	45.7 (11.0)
Race-ethnicity:								
White only	79.2%	64.2%	77.0%	73.9%	72.7%	75.6%	73.9%	90.2%
Black or African-American only	7.0%	4.5%	7.7%	6.0%	10.1%	5.2%	10.4%	2.4%
Latino or Hispanic only	3.1%	5.2%	2.6%	7.0%	5.1%	3.1%	2.9%	2.4%
Asian only	4.5%	14.9%	5.6%	3.0%	8.1%	9.8%	7.5%	0%
Native American, Alaskan, or Hawaiian only	0%	.0%	.5%	.5%	0%	.3%	.7%	0%
Multi-racial	5.8%	10.4%	6.6%	9.0%	4.0%	5.6%	4.2%	0%
Other or undisclosed	.4%	.7%	0%	.5%	0%	.3%	.3%	4.9%
Income [*M* (*SD*)]	4.33 (1.78)	6.30 (2.29)	4.37 (1.85)	4.24 (1.74)	4.23 (1.80)	4.41 (1.85)	4.71 (2.03)	–
(1) Under $15,000	6.2%	2.2%	7.1%	4.0%	7.4%	6.6%	9.4%	–
(2) $15,000–$24,999	11.3%	6.%	10.2%	13.6%	11.8%	12.5%	7.5%	–
(3) $25,000–$34,999	15.0%	6.0%	15.8%	18.1%	14.5%	11.8%	10.4%	–
(4) $35,000–$49,999	17.7%	8.2%	17.3%	22.1%	19.9%	15.0%	14.0%	–
(5) $50,000–$74,999	25.5%	12.7%	23.5%	17.1%	25.9%	28.6%	19.9%	–
(6) $75,000–$99,999	13.4%	9.7%	12.2%	13.6%	9.1%	12.9%	18.9%	–
(7) $100,000-$150,000	7.6%	20.9%	9.7%	8.0%	8.4%	8.4%	13.4%	–
(8) $150,000-$199,999	2.1%	9.7%	3.1%	2.5%	1.7%	2.4%	4.2%	–
(9) Over $200,000	1.2%	23.9%	1.%	.5%	1.3%	1.7%	2.0%	–

(Continues)

TABLE 1 (Continued)

	Image generation		Image rating					
	Study 1[1]	**Study 2[2]**	**Study 1[1]**	**Study 2[1]**	**Study 3[1]**	**Study 4[1]**	**Study 5[1]**	**Study 6[1]**
Undisclosed	0%	.7%	0%	.5%	0%	0%	.3%	–
Subjective SES [*M* (*SD*)]:	4.89 (1.62)	4.38 (1.87)	4.72 (1.53)	4.62 (1.51)	4.75 (1.58)	4.94 (1.58)	4.82 (1.78)	–
Rung 1 (least money, least education, least respected/no jobs)	1.6%	4.5%	1.5%	1.0%	1.7%	1.0%	1.6%	–
Rung 2	4.7%	9.7%	3.6%	4.0%	5.4%	5.2%	9.8%	–
Rung 3	14.0%	19.4%	19.4%	23.6%	15.2%	15.3%	15.3%	–
Rung 4	21.2%	24.6%	20.9%	17.1%	20.5%	15.3%	17.3%	–
Rung 5	21.4%	16.4%	20.4%	26.1%	26.3%	24.0%	14.0%	–
Rung 6	20.2%	11.2%	22.4%	17.1%	17.5%	21.3%	24.1%	–
Rung 7	13.2%	9.0%	10.2%	8.0%	9.8%	14.3%	12.4%	–
Rung 8	2.5%	3.0%	1.0%	3.0%	3.0%	3.5%	5.2%	–
Rung 9	.8%	.7%	0%	0%	.3%	0%	.3%	–
Rung 10 (most money, most education, most respected jobs)	.4%	1.5%	.5%	0%	.3%	0%	0%	–
Undisclosed	0%	0%	0%	0%	0%	0%	0%	–
Own[1]/parental[2] highest education [*M* (*SD*)]:	4.12 (1.27)	4.93 (1.32)	4.06 (1.29)	4.06 (1.25)	4.02 (1.24)	4.23 (1.23)	4.28 (1.33)	–
(1) Did not complete high school	0%	3.0%	0%	0%	.3%	0%	1.3%	0%
(2) Completed high school	13.4%	6.7%	13.8%	14.1%	12.8%	11.5%	12.7%	0%
(3) Began but did not complete college	23.5%	3.7%	26.0%	21.6%	27.3%	21.3%	16.3%	0%
(4) Completed an Associate degree	13.8%	9.0%	9.2%	19.6%	11.8%	10.8%	11.4%	0%
(5) Completed a Bachelor's degree	37%	35.8%	40.3%	33.2%	40.1%	45.3%	43.0%	0%
(6) Completed a post-secondary degree	12.3%	41.8%	10.2%	11.1%	7.7%	11.1%	15.3%	97.6%
Undisclosed	0%	0%	0%	.5%	0%	0%	0%	2.4%

Note: Superscripts indicate whether participant's own (1) or their parents' (2) highest level of educational attainment was assessed in each study.

and educational attainment. Notably, the general public participants had, on average, higher educational attainment, lower household incomes, and higher likelihoods of being male than the U.S. population at large (US Census Bureau, 2023), although our primary results did not reliably differ based on participants' gender or SES (see SOM Tables S3–S7). The gender breakdown of aspiring educators was similar to the overall K-12 teacher population in the U.S. (77% female), as was that of the in-service elementary school teachers in comparison to the overall elementary teacher population in the U.S. (89% female; National Center for Education Statistics, 2023). By contrast, the aspiring educators were wealthier than the average American K-12 educator (National Center for Education Statistics, 2023). Furthermore, the in-service educators taught in a district that scored much higher on standardized assessments than the average U.S. public school district (>3 grade levels above the national average), and that served students who were, on average, from much wealthier families (>2 *SD* above the national average) and were much more likely to be White or Asian (>80%) than the U.S. public school population at large (The Educational Opportunity Project at Stanford University, 2024). In addition, while sensitivity analyses suggested that the sample size was adequate to detect the noted results (see SOM Table S1), the in-service educator sample was very small (see Table 1). Finally, there was also less racial diversity in the present samples than in the U.S. general population (US Census Bureau, 2023). Indeed, as discussed in the Supplementary Online Materials (SOM), non-White participants had to be aggregated into a single group to conduct race-based moderation analyses, as there were not enough participants from individual non-White groups to conduct more granular analyses. Taken together, we expect the present results to hold among White Americans of diverse ages, genders, and SES levels, but we advise caution in extrapolating these results to non-Americans, to Americans from different racial-ethnic groups, or to teachers with experience working in more diverse or less resourced school districts.

The conceptualization of the present research questions and the selection of methodologies to examine them were led by the first author, a first-generation university graduate with doctoral training in the social psychology of education and SES, and postdoctoral training in educational psychology. Input on these processes was provided by the second author, a continuing-generation university graduate with doctoral training in the learning sciences and social psychology, postdoctoral training in cognitive psychology, and professional experience in the fields of educational psychology and human development. In addition, as straight, White, able-bodied, cisgender males living in the U.S. and working in academia, both authors have long had the psychological privilege of not needing to be chronically vigilant to many of their social identities. As a consequence of these epistemic and experiential backgrounds, the theories, methodologies, and analytic decisions on which the present work is based may contain some important blind spots regarding the experiences of individuals from marginalized communities. For example, in determining our research questions, which samples to recruit, and which variables to measure, we drew a great deal from two fields. The first was social psychological research on the effects of culturally-held race-, ethnicity-, and SES-based stereotypes on person perception and interpersonal judgments. The second was educational psychology and policy research on race-, ethnicity-, and SES-based inequities in academic outcomes, access to educational resources, and support for policies designed to address these inequities. We therefore gave much greater consideration to the roles of the perceived SES and, initially, race-ethnicity (until such effects proved inconsistent and weaker in effect size in early studies) of student targets than to the roles of the targets' other social identities (e.g., gender, age) or to the demographics of the perceivers (i.e., the participants we chose to recruit). We also did not explicitly consider how SES might meaningfully stand out from versus intersect with other identities that are associated with academic stereotypes. For example, might there be

JOURNAL OF Social Issues A Journal of the Society for the Psychological Study of Social Issues

differences in policy support between perceivers who mentally represent low-ability students primarily as low-SES, versus those who represent them as being both low-SES *and* Black or Latinx (Cuddy et al., 2009)? While we conducted some exploratory analyses of these issues with secondary data where possible (see SOM), the priorities we established based on our epistemic and experiential backgrounds meant that the data needed to provide a detailed accounting of these alternative relationships were not collected in the present work.

STUDIES 1–2: CAPTURING THE MENTAL REPRESENTATIONS OF HIGH-AND LOW-ABILITY STUDENTS, AND EXAMINING THE ASSOCIATIONS OF MENTAL REPRESENTATIONS WITH POSITIVE AND NEGATIVE ACADEMIC ATTRIBUTES

We report Studies 1–2 together because their methods and results were similar. Both studies involved 2 phases – an image-generation phase, and an image-rating phase – which involve separate samples of participants. In general, we directly modeled the methods and analyses of the present studies on prior research by Brown-Iannuzzi et al. (2017). We did so because we viewed their studies as robust examples of both how to capture people's mental representations of different categories, and how to test whether such representations influence the level of support that people are willing to provide members of those categories. In studies that used approaches that differed from theirs (e.g., Studies 3 and 5), we provide an explanation of and rationale for those changes when introducing those studies. For all studies in this paper, see https://osf.io/snr97 for materials, data, and analytic syntax, including those not relevant to the present research questions.

Method

Image-generation phases

Participants

Study 1's image-generation phase consisted of 486 American adults recruited from Amazon's Mechanical Turk (MTurk) with approval rates of 99% or higher. Study 2's image-generation phase consisted of 134 American undergraduate students enrolled in introductory courses at a school of education at an elite, private, predominantly White university in the Northeastern U.S. who explicitly identified themselves as planning to become teachers or to go into another career that involves working with students. See Table 1 for complete demographics for all samples used in the present paper. All studies were well-powered to detect our effects of interest – see Table S1 for sensitivity analyses.

Procedure

Image-generation participants completed a reverse-correlation task, a methodology designed to visually capture respondents' mental representations of the kinds of people who exemplify specific social categories. As shown in Figure 1, following established procedures (Brown-Iannuzzi et al., 2017; see also Dotsch et al., 2008; Lei & Bodenhausen, 2017), we began with a 400 × 400 pixel base image used in prior work and used the R package *rcicr* to generate the specific stimuli needed for the present studies. Specifically, this software generates a requested number of stimulus pairs that differ from one another in terms of the level of brightness that the software randomly

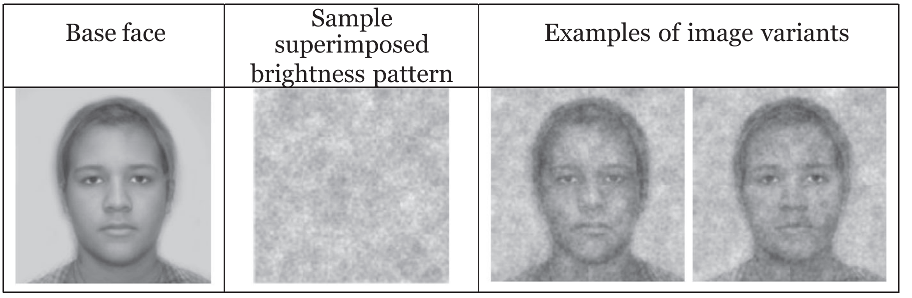


FIGURE 1 Overview of the image-generation task stimuli.

superimposes on each individual pixel that makes up the base image. For example, as shown in Figure 1, superimposing different levels of brightness on the same base image can produce a more downturned mouth (seen in the third image, as compared to the fourth image) or darker eyes in one picture than in others (seen in the fourth image, as compared to the third image). For each stimulus pair that the program generates, one image in each pair has a random pixel brightness pattern superimposed on the base face, and the other image in that pair has the reverse pattern superimposed on the same face (i.e., the brightness level superimposed on each pixel of the second image is the opposite of what was superimposed on the corresponding pixel in the first image).

Participants were presented with these pairs of faces (100 pairs in Study 1, 400 pairs in Study 2), and were randomly assigned to select, from each pair, the face that "looks most like a student with" either "low academic abilities" or "high academic abilities." Both the face pair presentation order and which face from each pair was presented on the left versus the right were randomized. Following the established procedures referenced above, the pixel brightness pattern of each image selected by participants responding to the "low academic abilities" prompt were averaged to create a mean low-ability student representation. The same procedure was applied to the images selected by participants responding to the "high academic abilities" prompt to create a mean high-ability student representation.[1]

Image-rating phases

Participants

For the image-rating phases of both studies, we recruited samples of 196 and 199 American adults on MTurk, with approval rates of 99% or higher.

Procedure

In both studies, participants rated (in random order) both the high- and low-ability images, as well as two filler images that were included to reduce the likelihood that participants would directly contrast the two focal images (but whose ratings were not included in analyses). Some participants (Study 1 $N = 97$; Study 2 $N = 97$) were randomly assigned to indicate the following for each target before advancing to the next target: their perceptions of the student's academic abil-

[1] Following best-practice recommendations (Cone et al., 2021), we determined that it is unlikely that the Type I error inflation that can affect the reverse-correlation procedure influenced the conclusions of the present work. See SOM for details.

JOURNAL OF Social Issues A Journal of the Society for the Psychological Study of Social Issues

ities ("How intelligent does this student look?", "How academically competent does this student look?") and academic attributes that extrapolate beyond ability, including their academic motivation ("How academically motivated does this student look?", "How academically confident does this student look?"), work ethic ("How academically hardworking does this student look?", "How academically lazy does this student look?"), potential ("How likely is this student to finish high school?", "How likely is this student to finish college?"), and behavior ("How receptive would this student be to feedback in school?", "How likely is this student to follow directions in school?", "How likely is this student to exhibit problem behavior in school?", "How likely is this student to cheat on assignments?", "How likely is this student to complete their assignments on time?"). Item order was randomized for each target, and participants responded to all items using 6-point scales (1 = "not at all" to 6 = "extremely").

A second group of participants (Study 1 $N = 99$; Study 2 $N = 102$) were randomly assigned to instead indicate their perception of each target's SES. Specifically, participants were presented with a picture of a 10-rung ladder and were asked to indicate "Where do you think this student stands on this ladder?" (Adler et al., 2000). Participants responded using a 10-point scale ("Rung 10 (families with the most money, most education, and most respected jobs)" to "Rung 1 (families with the least money, least education, and least respected/no jobs)"). This group of participants also indicated the extent to which they perceived each target as "White or European American," "Black or African American," "Hispanic or Latin American," "Asian or Asian American," and "multiracial," all using 6-point scales (1 = "not at all" to 6 = "extremely"). These race-ethnicity perception ratings were included for two reasons. The first is due to the prevalence of race- and ethnicity-based intellectual stereotypes, which result in Black and Hispanic people often being reflexively labeled as less academically capable than White and Asian people (Cuddy et al., 2009). It therefore seemed plausible that Americans' mental representations of high- and low-ability students may be primarily imbued with cues that signal race or ethnicity, rather than SES. Second, the base image used in the image generation task was multiracial in nature, which prior research has shown can result in mean representations that differ in how they are racially perceived and, as a result, treated (Brown-Iannuzzi et al., 2017).[2]

Results

Image-generation results

Addressing RQ1, Table 2 displays the mean images produced by the two image-generation samples. *Within* each study, a visual inspection of the faces generated in the high-ability student condition was quite different from the faces generated in the low-ability student condition, even though all participants selected from the same pairs of images. Furthermore, to provide an objective, quantitative measure of dissimilarity (Brown-Iannuzzi et al., 2017; Ratner et al., 2014), we used the *rcicr* package to calculate the correlations between the brightness levels of each pixel across the four generated images. Specifically, for each of the four generated images, the software compared the brightness of a given pixel in one image to the brightness of the pixel at the same

[2] These second groups of participants also rated the images in terms of other social categories that were not relevant to the primary hypotheses of the present work (e.g., gender). In addition, a third group of participants in Study 2 was assigned to complete personal attribute ratings (e.g., likeability, hostility) that were also not relevant to the present research questions. See SOM and Table S8 for details.

TABLE 2 The resulting images produced by averaging the noise patterns of the images selected by participants in each condition (high-ability student condition vs. low-ability student condition), and then superimposing those averaged noise patterns on the base face, in Studies 1–2.

	Low-ability student image	High-ability student image
Study 1 (general public)		
Study 2 (aspiring educators)		

TABLE 3 Correlations (and confidence intervals) among the pixel brightness levels of the images generated in Studies 1 and 2.

	Study 1	Study 2	
	Low-ability student	**High-ability student**	**Low-ability student**
Study 1			
High-ability student	−.905	–	–
	[−.906, −.903]		
Study 2			
Low-ability student	.461	−.486	–
	[.455, .467]	[−.491, −.481]	
High-ability student	−.565	.573	−.537
	[−.570, −.560]	[.568, .577]	[−.542, −.532]

location in the second image, repeated this comparison for each pixel in both images, and then averaged all of these comparisons to produce an overall pixel brightness correlation between those images. Confirming our qualitative results, there was a significant negative correlation between the brightness of the facial pixels in the low-ability image and the brightness of these pixels in the high-ability image within each study (Table 3).

In addition, looking *across* studies, the faces that were generated in each condition were highly visually similar. This is confirmed by the significant positive correlations in pixel brightness between the two low ability images and between the two high ability images (Table 3). By contrast, the brightness of the pixels in the low-ability student image in each study was significantly

JOURNAL OF Social Issues A Journal of the Society for the Psychological Study of Social Issues WILEY

negatively correlated with the brightness of the pixels in the high-ability student image from the other study.

Overall, then, both the qualitative and quantitative results of the image-generation phases provide convergent conclusions for RQ1: on average, both members of the general public and participants more connected to education hold distinct mental representations of high- and low-ability students. Note that, in secondary analyses, we also found that these results held regardless of participants' beliefs about the fixedness or malleability of intellectual ability – see SOM for details.

Image-rating results

The following results address RQ2. Because many of the same image-rating measures were used in all of the studies discussed in this article, the image-rating results from all six studies are presented in Table 4.

Overall, we found that the images created by having participants select faces that they felt depicted lower-ability students were rated as having significantly weaker academic abilities (i.e., lower intelligence and academic competence) than those created by having participants select faces that they felt depicted high-ability students. And, critically, the participants who made these ratings were different from the participants whose responses were used to generate the images. This supports the hypothesis that people's mental representations of students' academic ability levels are discernible by others.

In addition, the students depicted in the low-ability images were also more likely to be characterized as having negative academic attributes that were separate from the one that the image-generation process was designed to isolate (i.e., ability). Compared to the high-ability images, the low-ability images were perceived to be significantly less academically motivated, confident, and hardworking, less likely to finish high school or college, and more likely to exhibit problematic behavior in school. Finally, the students depicted in the low-ability images were perceived to be from lower SES backgrounds than those depicted in the high-ability images. Virtually all effect sizes described here were very large (see Table 4), and they were not moderated by participants' own race, gender, or SES (see SOM Tables S3–S7).

By contrast, the students depicted in the low-ability images were only perceived to look significantly more Black and less White than those depicted in the high-ability images by participants in Study 2, but not in Study 1. The students depicted in the low-ability images were also perceived to look marginally more Asian than those depicted in the high-ability images. Unexpectedly, the students depicted in the low-ability images were perceived as marginally or significantly less Hispanic and multiracial than those depicted in the high-ability images in both studies. However, the effect sizes for these significant White, Black, Hispanic, and multiracial analyses were much smaller than for all of the other results described here. Further secondary analyses also found that the image-rating results were not reliably moderated by participants' beliefs about the fixedness or malleability of intellectual ability (see SOM for details).

Taken together, Studies 1–2 provide convergent evidence for two of our research questions. Specifically, Americans – including both members of the general public and aspiring educators – may share qualitatively distinct mental representations of what lower versus higher ability students look like (RQ1). In addition, their mental representations of lower- (vs. higher-) ability students may be more closely associated with low-SES and negative academic attributes (RQ2).

TABLE 4 Mean (SDs) perceptions and *t*-tests of high- and low-ability image ratings in all studies.

	Study	High-ability student image [*M* (*SD*)]	Low-ability student image [*M* (*SD*)]	*t*	*df*	*p*	Cohen's *d*
SES	1	4.97 (1.24)	3.29 (1.23)	11.30	98	<.001	1.14
	2	4.91 (1.37)	3.05 (1.44)	11.72	101	<.001	1.16
	3	4.97 (1.43)	3.30 (1.48)	9.84	295	<.001	1.14
	4	5.68 (1.57)	3.08 (1.41)	22.83	286	<.001	1.35
	5	5.54 (1.20)	3.81 (1.33)	23.64	306	<.001	1.35
Intelligence	1	4.67 (1.00)	2.38 (1.13)	16.20	96	<.001	1.64
	2	4.51 (1.17)	2.56 (1.20)	11.73	96	<.001	1.19
	3	4.52 (1.07)	2.74 (1.13)	14.02	295	<.001	1.63
	4	4.76 (.86)	2.71 (1.14)	24.37	286	<.001	1.44
	6	4.73 (.80)	4.09 (1.31)	4.08	32	<.001	.71
Academic competence	1	4.71 (1.01)	2.33 (1.09)	15.90	96	<.001	1.61
	2	4.59 (1.07)	2.43 (1.16)	14.32	96	<.001	1.45
	3	4.59 (1.04)	2.66 (1.16)	15.11	295	<.001	1.75
	4	4.82 (.85)	2.65 (1.20)	24.69	286	<.001	1.46
	5	4.27 (.72)	3.08 (.98)	22.43	306	<.001	1.28
	6	4.79 (.82)	3.91 (1.26)	5.44	32	<.001	.95
Academically motivated	1	4.71 (1.06)	1.99 (1.14)	17.34	96	<.001	1.76
	2	4.65 (1.11)	2.09 (1.18)	16.34	96	<.001	1.66
	3	4.35 (1.08)	2.09 (1.08)	17.98	295	<.001	2.09
	4	4.88 (.85)	2.28 (1.23)	27.18	286	<.001	1.60
	6	4.85 (.76)	3.12 (1.58)	8.06	32	<.001	1.40
Academic confidence	1	4.72 (1.12)	2.05 (1.07)	16.51	96	<.001	1.68
	2	4.57 (1.03)	2.20 (1.21)	14.94	96	<.001	1.52
	3	4.53 (1.11)	2.25 (1.16)	17.25	295	<.001	2.00
Academically hardworking	1	4.63 (1.08)	2.20 (1.15)	15.49	96	<.001	1.57
	2	4.60 (1.13)	2.35 (1.18)	14.42	96	<.001	1.46
	3	4.46 (1.05)	2.41 (1.15)	16.09	295	<.001	1.87
	4	4.82 (.88)	2.42 (1.19)	25.34	286	<.001	1.50
	5	4.27 (.73)	3.01 (1.01)	22.84	306	<.001	1.30
	6	4.82 (.77)	3.52 (1.52)	6.05	32	<.001	1.05
Academically lazy	1	2.04 (1.06)	4.27 (1.36)	−12.11	96	<.001	−1.23
	2	2.19 (1.17)	4.36 (1.35)	−12.87	96	<.001	−1.31
	3	1.99 (.88)	4.10 (1.34)	−16.15	262.91	<.001	−1.86
	4	2.05 (1.11)	4.16 (1.40)	−18.29	286	<.001	−1.08
Likely to finish high school	1	5.18 (1.04)	3.05 (1.38)	13.44	96	<.001	1.36
	2	5.03 (.97)	3.04 (1.35)	12.63	96	<.001	1.28
	3	5.03 (1.03)	3.36 (1.21)	12.81	291.10	<.001	1.48
	6	5.18 (.73)	4.18 (1.26)	4.69	32	<.001	.82

(Continues)

TABLE 4 (Continued)

	Study	High-ability student image [*M* (*SD*)]	Low-ability student image [*M* (*SD*)]	*t*	*df*	*p*	Cohen's *d*
Likely to finish college	1	4.55 (.99)	2.21 (1.30)	15.82	96	<.001	1.61
	2	4.42 (1.21)	2.31 (1.27)	12.61	96	<.001	1.28
	3	4.36 (1.10)	2.38 (1.16)	15.16	295	<.001	1.76
	4	4.90 (.87)	2.63 (1.20)	24.31	286	<.001	1.43
	5	4.33 (.80)	3.03 (1.09)	22.69	306	<.001	1.30
Receptive to feedback in school	1	4.66 (1.05)	2.29 (1.03)	15.99	96	<.001	1.62
	2	4.43 (1.19)	2.51 (1.23)	12.70	96	<.001	1.29
	3	4.50 (1.13)	2.51 (1.23)	14.47	295	<.001	1.68
Likely to follow directions in school	1	4.92 (.93)	2.71 (1.31)	14.63	96	<.001	1.49
	2	4.64 (1.13)	2.64 (1.23)	11.79	96	<.001	1.20
	3	4.64 (1.04)	2.72 (1.12)	15.26	295	<.001	1.77
	6	4.88 (.74)	3.85 (1.37)	5.51	32	<.001	.96
Likelihood of problem behavior in school	1	1.97 (.96)	4.06 (1.31)	−12.29	96	<.001	−1.25
	2	2.33 (1.26)	4.27 (1.34)	−10.99	96	<.001	−1.12
	3	2.25 (1.17)	4.20 (1.20)	−14.19	295	<.001	−1.65
	6	2.21 (1.05)	3.15 (1.39)	−3.18	32	.003	−.55
Likely to cheat	1	2.04 (1.14)	3.63 (1.38)	−10.53	96	<.001	−1.07
	2	2.37 (1.32)	3.84 (1.31)	−7.59	96	<.001	−.77
	3	2.22 (1.10)	3.61 (1.30)	−9.98	291.17	<.001	−1.15
Likely to complete assignments on time	1	4.86 (1.02)	2.57 (1.20)	14.89	96	<.001	1.51
	2	4.67 (1.08)	2.53 (1.25)	13.47	96	<.001	1.37
	3	4.59 (1.02)	2.76 (1.22)	13.96	290.27	<.001	1.61
	5	4.38 (.75)	3.07 (1.00)	23.00	306	<.001	1.31
White or European American	1	2.16 (1.12)	1.93 (1.24)	1.77	98	.08	.18
	2	2.46 (1.42)	1.77 (1.12)	5.03	101	<.001	.50
	3	2.31 (1.26)	2.38 (1.40)	−.42	295	.677	−.05
Black or African American	1	3.67 (1.50)	3.95 (1.51)	−1.56	98	.123	−.16
	2	3.70 (1.45)	4.59 (1.32)	−5.95	101	<.001	−.59
	3	3.70 (1.48)	3.96 (1.64)	−1.42	295	.157	−.16
Hispanic or Latin American	1	3.16 (1.38)	2.87 (1.47)	1.93	98	.056	.19
	2	3.24 (1.43)	2.69 (1.44)	3.18	101	.002	.31
	3	3.68 (1.41)	2.98 (1.32)	4.39	295	<.001	.51
Asian or Asian American	1	1.66 (.93)	1.47 (.72)	1.78	98	.077	.18
	2	1.81 (1.21)	1.62 (1.02)	1.83	101	.07	.18
	3	1.77 (1.01)	1.62 (.88)	1.34	295	.181	.16
Multiracial	1	3.86 (1.24)	3.22 (1.43)	3.81	98	<.001	.38
	2	3.87 (1.28)	3.44 (1.57)	2.70	101	.008	.27
	3	4.08 (1.30)	3.56 (1.43)	3.25	295	.001	.38

STUDY 3: BETWEEN-SUBJECTS REPLICATION

It is possible that the results of Studies 1–2 were due in part to participants making direct comparisons between the two focal images that were shown to all participants during the image-rating phases. The goal of Study 3 was thus to replicate these findings with a between-subjects approach to image-rating. Participants in Study 3 therefore rated one image each in all of the rating categories from Studies 1–2 (perceived academic ability, academic attributes separate from ability, SES, and race-ethnicity).

Method

Participants were 297 American adults recruited on MTurk, with approval rates of 99% or higher. Participants were randomly assigned to rate either the high-ability image ($N = 145$) or the low-ability image ($N = 152$) generated in Study 2 (by the aspiring educator sample), using all of the same academic attribute, SES, and race-ethnicity items as in Studies 1–2, again presented in random order.[3]

Results

Addressing RQ2 and replicating Studies 1–2 with a between-subjects approach, the student depicted in the low-ability image was more likely to be perceived as having weaker academic capabilities, as less academically motivated and confident, as having poorer work ethic, as having less academic potential, to be more likely to exhibit problematic behavior in school, and to be from a lower-SES background than the student depicted in the high-ability image (see Table 4). Also replicating Studies 1–2, virtually all effect sizes described here were very large, and they were not reliably moderated by participants' race, gender, SES, or beliefs about the fixedness or malleability of intellectual ability (see SOM Tables S3–S7). By contrast, there were no significant differences in the extent to which the images were perceived to look White, Black, or Asian (see Table 4). The students depicted in the low-ability images were again perceived as significantly less Hispanic and multiracial than those depicted in the high-ability images in both studies; however, the effect sizes of these differences were again much smaller than for all of the other results described here.

STUDIES 4–6: MENTAL REPRESENTATIONS OF LOWER-ABILITY STUDENTS ARE SEEN AS LESS DESERVING OF SUPPORT BY BOTH THE GENERAL PUBLIC AND TEACHERS

Having found convergent evidence for RQs 1 and 2 across multiple studies, the goal of Studies 4–6 was to test a final research question: do the widely-held mental representations of low- and high-ability students influence the level of support that both the public (Studies 4–5) and teachers (Study 6) are willing to provide these students (RQ3)? Generalization to three different types of support were tested: providing college scholarships (Study 4), admitting students to college (Study

[3] Study 3 participants also completed the other social category and personal attribute ratings described in Studies 1-2 for the image they were assigned—see SOM and Table S8 for details.

5), and using motivationally supportive and unsupportive teaching practices (Study 6). Finally, to further confirm that the present results are not primarily the result of Type I error inflation issues that have been associated with the reverse correlation method (Cone et al., 2021), Study 5 used an image-compositing technique that is less susceptible to Type I inflation.

Study 4

Method

Participants in Study 4 were 287 American adults recruited on MTurk, with approval rates of 99% or higher. As in Studies 1–2, Study 4 participants provided responses for both the high- and low-ability images generated in Study 2 (by the aspiring educator sample), as well as for two filler images (the same as in Studies 1–2), which were again included to reduce the likelihood that participants would directly contrast the two focal images. The four images were presented in a random order.

First, participants were told that the images were of graduating high school students who applied for academic scholarships from a private foundation. Then, for each image, they were asked "How supportive or unsupportive would you be of giving this student a scholarship?" (6-point scale; 1 = "completely unsupportive" to 6 = "completely supportive").

Next, participants were simultaneously presented with the images of both the low- and high-ability faces, side by side (left-right position was randomized), with these instructions: "Imagine that you had a say in determining which students would receive a scholarship from this private foundation. Below are the photos of two of the applicants you answered questions about. If you had to give a scholarship to just one of these two students, which student would you choose?" Each participant's choice was recorded.

Finally, participants rated each face on the same SES measure and a subset of the same academic characteristics (and using the same response scales) as in Studies 1–3. These included intelligence, academic competence, academic motivation, academic work ethic, academic laziness, and likelihood of finishing college, presented in random order.

Results

Addressing RQ2 and replicating Studies 1–3, the student depicted in the low-ability image was perceived as being from a significantly lower-SES background, and as having significantly lower intelligence, competence, academic motivation, work ethic, and potential than the student depicted in the high-ability image (see Table 4). More critically, addressing RQ3, participants were significantly less supportive of providing a scholarship to the low-ability student image ($M = 3.43$, $SD = 1.37$) than to the high-ability student image ($M = 4.92$, $SD = .82$), $t(286) = 18.44$, $p < .001$, Cohen's $d = 1.09$. They were also significantly less likely to choose to give a scholarship to the student in the low-ability image ($N = 13$ [4.50%]) than the student in the high-ability image ($N = 274$ [95.50%]), one-sample proportions test comparing to 50%: $\chi^2(1) = 235.54$, $p < .001$, Cohen's $H = 1.16$. Virtually all effect sizes were again very large, and they were not reliably moderated by participants' race, gender, SES, or beliefs about the fixedness or malleability of intellectual ability (see SOM Tables S3–S7).

Study 5

Method

Participants in Study 5 were 307 American adults recruited on CloudResearch's Connect crowdsourcing network. Like Studies 1, 2, and 4, Study 5 used a within-subjects approach to image-rating: participants provided responses for both high- and low-ability images generated in Study 2 (by the aspiring educator sample). However, Study 5 used an image-compositing technique that is less likely to influence Type I error rates: Cone et al.'s (2021) subgrouping approach. Specifically, our prior studies involved averaging the image-generation selections of *all* participants in a condition to provide raters with a single mean low-ability student representation and a single mean high-ability student representation to rate. By contrast, the subgroup approach uses "the judgments of *random subsets* of multiple participants in each condition" (Cone et al., 2021, p. 767, emphasis in original) to create *multiple* low-ability student representations and *multiple* high-ability student representations (e.g., Brown-Iannuzzi et al., 2021; Hutchings et al., 2021). Given our original image-generation sample of 134 (Study 2) and following recommendations from one of the creators of this approach (R. Lei, personal communications, September 18–19, 2022), we created 10 low- and 10 high-ability student representations, each consisting of the averaged judgments of six participants from the associated condition that were randomly selected without replacement.

Similar to Studies 1, 2, and 4, participants rated each of the 20 faces (introduced as high school students) using the same SES measure and a subset of the same academic characteristics (and using the same response scales) as in Studies 1–3. These included academic competence, academic work ethic, likelihood of completing their work on time, and likelihood of finishing college, in that order. For each characteristic, their judgments of the 10 high-ability images were averaged together, as were their judgments of the 10 low-ability images.

Finally, as in Study 4, we tested whether mental representations of low- and high-ability students influence the level of support that people are willing to provide these students. In Study 4, we found that when participants made scholarship judgments and decisions based on facial cues alone – that is, in the absence of any academic performance information (e.g., grades) – they showed a clear bias against the low-ability student image. However, prior research has shown that the effects of non-performance-based information on academic ability judgments may dissipate when explicit performance information is provided (Muenks et al., 2016). Thus, it was an open question whether the findings that emerged in Study 4 would hold when performance information was available.

The final goal of Study 5 was therefore to test whether the preferential support of high (vs. low) ability student representations would hold when both representations were explicitly presented as equally academically qualified. To achieve this, participants completed a modified version of Axt et al.'s (2016; 2018) judgment bias task. Participants were told that they would be playing the role of an admissions officer for an elite U.S. college, and so they would be deciding which of the 20 ostensible high school students they just saw were most qualified and should be accepted, and which were not and should be rejected. They therefore viewed each of the 20 images again, one at a time and in random order. This time, however, they were told that each image would be presented with four pieces of academic performance information. These included the student's GPA (1.0–4.0) in their high school science classes (biology, chemistry, etc.), their GPA (1.0–4.0) in their high school humanities classes (English, foreign languages, etc.), a score for their letters of recommendation (poor, fair, good, or excellent), and an interview score (out of 100). Participants were instructed to weigh each piece of information equally, and to accept approximately

JOURNAL OF Social Issues A Journal of the Society for the Psychological Study of Social Issues

half of the applicants. Critically, half of the high-ability student images and half of the low-ability student images were presented with objectively stronger academic qualifications, and the other half of each category of images were presented with objectively weaker academic qualifications. To do this, following Axt et al. (2016), we standardized each piece of information to have a 1–4 range: the two GPAs already ranged from 1 to 4, and we converted the recommendation letters (poor = 1, fair = 2, good = 3, excellent = 4) and interview scores (dividing by 25). Less qualified applicants had information summing to 13 (out of a possible 16), and more qualified applicants had information summing to 14. All participants saw the same image-qualification pairings.

The use of this design was beneficial because it included objectively correct answers (accept more qualified students, reject less qualified students) and objectively incorrect answers (reject more qualified students, accept less qualified students). Thus, in line with signal detection theory, participants' responses could be used to calculate two key metrics. The first is their level of sensitivity (*d'*), or their ability to distinguish more qualified students from less qualified students. Specifically, a more sensitive participant is one who reliably accepts students with objectively stronger qualifications and reliably rejects students with objectively weaker qualifications. As in prior work (Axt et al., 2016; Axt et al., 2018), because the high- and low-ability student images were presented with equally qualified information on average, we expected participants to be equally able to distinguish objectively more qualified students from objectively less qualified students (i.e., equal sensitivity), regardless of whether the application was presented with a high- or low-ability student mental representation.

By contrast, and more relevant to the present work, the second metric that can be calculated from this paradigm is a participant's level of response bias (also known as their criterion, or *c*), which denotes their personal threshold for accepting or rejecting a student. Specifically, a given participant can have a more liberal threshold, meaning that they are more likely to accept students regardless of their qualifications, or a more conservative threshold, meaning that they are less likely to accept students regardless of their qualifications. More critically, prior research has shown that a participant's threshold for accepting students could vary based on the image that is associated with an application (Axt et al., 2016; Axt et al., 2018). As a result, we tested whether participants would show different levels of response bias (i.e., different criterion scores) based on whether the student is portrayed with a high- versus low-ability image.

Results

Addressing RQ2 and replicating Studies 1–4, the students depicted in the 10 low-ability images were perceived, on average, as being from significantly lower SES backgrounds, and as having lower competence, work ethics, and likelihoods of completing their work on time or finishing college than the students depicted in the 10 high-ability images (see Table 4). This supports our claim that the present results are likely not due to Type I error inflation, as the image-compositing technique used in this study is less likely to influence Type I error rates (Cone et al., 2021).

Following Axt et al. (2016), 15 participants' judgment bias task scores were excluded from analyses because they accepted more than 80% of students, suggesting that they did not follow the instructions (i.e., to accept approximately half of the applicants). There was not a reliable difference in sensitivity (*d'*) to applications with with high-ability student images ($M = 1.08$, $SD = .78$) versus to those with low-ability student images ($M = 1.00$, $SD = .69$), $t(291) = 1.45$, $p = .149$, Cohen's $d = .08$, meaning that participants were capable of distinguishing between more and less qualified applicants, regardless of the image associated with their application. This serves as a fidelity check for the paradigm, as it suggests that differences between students' academic qualifications were sufficiently explicit and salient.

By contrast, participants' criterion scores (c) differed significantly based on the image associated with students' applications, $t(291) = -9.46$, $p < .001$, Cohen's $d = -.55$. That is, applications with high-ability student images were held to a lower acceptance criterion ($M = -.25$, $SD = .40$) than applications with low-ability student images ($M = .05$, $SD = .41$). This means that even when they included the same objective academic qualifications, applications with high-ability student images were more likely to be accepted than applications with low-ability student images. Specifically, more qualified applicants were 13.4% more likely to be accepted when their applications were paired with a high-ability representation than when they were paired with a low-ability representation, and less qualified applicants were 9.6% more likely to be accepted when their applications were paired with a high-ability representation than when they were paired with a low-ability representation. Virtually all effect sizes observed in Study 5 were again very large and emerged regardless of participants' race, gender, or SES (see SOM Tables S4–S7).

Study 6

Method

Participants were 41 American in-service elementary school teachers, employed in a K-12 school district near a major metropolitan area in the Northeastern United States, who completed Study 6 as part of a larger multi-wave study of teachers and students. Like Studies 1, 2, 4, and 5, Study 6 used a within-subjects approach to image-rating: teachers provided responses for both the high- and low-ability images generated in Study 2 (by the aspiring educator sample), which were presented in random order.

Given the perceived ages of the composite faces, teachers were first asked to imagine that they were teaching a high school class and that each image was of a student in their class who was struggling on a math assignment. For each image, teachers indicated how likely they would be to engage in ten teaching behaviors with that student (6-point scale; 1 = "very unlikely" to 6 = "very likely"; Miele et al., 2019). Five were practices that have been established as motivationally supportive and which promote more positive academic outcomes for students (e.g., "Encourage the student to keep working hard on the assignment"; high-ability image: $\alpha = .53$; low-ability image: $\alpha = .58$; Park et al., 2016; Stipek et al., 2001). The other five practices are considered to be motivationally unsupportive and can contribute to more negative student outcomes (e.g., "Give the student an easier assignment to work on"; high-ability image: $\alpha = .43$; low-ability image: $\alpha = .44$).[4]

Teachers then rated each image on a subset of the same academic characteristics and using the same response scales as in Studies 1–3. This included intelligence, academic competence, academic motivation, academic work ethic, and the likelihood of following directions in school, exhibiting problem behavior in school, and finishing high school. For both the teaching practices

[4] We report Cronbach's α here because it is a standard measure of internal consistency reliability. However, note that alpha is likely to be inaccurate for the present instrument, for reasons that we explain in the SOM. We had originally intended to report categorical omega ($\omega_{\text{u-cat}}$) as the primary index of internal consistency for these scales. However, for various reasons (detailed in the SOM), we came to view these estimates as unreliable. We report these omega estimates in Table S9, along with somewhat more reliable estimates from a larger study that used the same measure.

task and the academic ratings task, the face presentation order and the items-within-face order were both randomized.

Results

Addressing RQ2 and replicating Studies 1–5, the student depicted in the low-ability images was significantly more likely to be perceived as having lower intelligence, academic competence, motivation, work ethic, and academic potential, and as less likely to follow directions and more likely to exhibit problematic behavior in school than those depicted in the high-ability images (see Table 4).

To examine teachers' endorsement of using different instructional practices with the different students (RQ3), we submitted their likelihood of using supportive and unsupportive practices to a mixed ANCOVA. This analysis included instructional practice type (supportive vs. unsupportive) and image type (high-ability vs. low-ability) as within-subjects factors, and the associated interaction of these terms. We used this analytic approach because it enabled us to determine whether any difference that we observed in teachers' likelihood of using supportive practices with the high- versus low-ability student was similar to or different from the difference in their likelihood of using unsupportive practices with the high- versus low-ability student (e.g., Miele et al., 2019).

A significant student ability level × instructional practice type interaction emerged, F (1, 40) = 4.94, p = .032, η^2 = .11. Critically, simple slopes analyses revealed that while teachers were equally likely to endorse using supportive practices with both students (high-ability image: M = 5.38, SD = .51; low-ability image: M = 5.36, SD = .51), t(40) = .50, p = .618, Cohen's d = .08, they were significantly more likely to endorse using unsupportive practices with the student depicted in the low-ability image (M = 1.69, SD = .52) than with the student depicted in the high-ability image (M = 1.58, SD = .47), t(40) = −2.71, p = .010, Cohen's d = .42. Finally, the results of Study 6 were not reliably moderated by teachers' beliefs about the fixedness or malleability of intellectual ability (see SOM Table S3).

Taken together, Studies 4–6 provide convergent evidence for two of our research questions. Specifically, they suggest that Americans' mental representations of lower-ability students may not only be more associated with negative academic attributes (RQ2), but may also influence the level of support that both the public and teachers are willing to provide these students (RQ3).

GENERAL DISCUSSION

A large majority of Americans feel that the distribution of educational opportunities should be based primarily on formal assessments of students' academic skills and abilities (e.g., Gallup News, 2016; Pew Research Center, 2019). The success of such a system hinges not only on these assessment methods being fair and accurate, but on people being able to prevent non-assessment factors from subtly influencing how students are classified. However, prior work on biases among both teachers and the general public has shown that this is often not the case, as individuals frequently make negative ability inferences about particular students based on their perceived membership in social groups that are associated with negative academic stereotypes (e.g., race, ethnicity, gender, SES; Doyle & Easterbrook, 2024; Moss-Racusin et al., 2012; Okonofua et al., 2016).

The present work provides a valuable extension to these findings. Specifically, our studies suggest that beyond simply making ability inferences about students who are currently in front of them based on the social categories they appear to belong to, individuals may carry with them mental representations of what low- and high-ability students look like. Furthermore, the representations that people hold may be culturally shared, as the representations that emerged in the present work were remarkably similar across somewhat different groups of individuals (aspiring educators vs. members of the general public). And most critically, the present findings suggest that these representations may be imbued with facial cues that naïve observers may use to make automatic judgments of a target's other educationally-relevant characteristics (e.g., motivation), as well as with social categories that are associated with academic stereotypes (e.g., SES).

Establishing that these shared mental representations exist, and demonstrating that they influence people's educational-relevant judgments of individual students, represents a step toward advancing an understudied area of the teacher expectancy literature. Specifically, such findings may contribute to our understanding of the specific visual cues that teachers and other individuals may unintentionally draw on when forming inferences about the ability level of individual students, especially in the absence of information about the students' prior academic performances. Indeed, prior studies on teacher expectancies have shown that teachers report higher academic expectations for higher (vs. lower) SES students, even in the absence of explicit SES information about those students (for reviews, see Good et al., 2018; Wang et al., 2018). It is possible that, in these cases, teachers may be inferring students' abilities based on a set of facial cues that they assume differ between lower- and higher-SES students. Though our current work does not systematically test and catalog these cues, it does offer some promising starting points that align with findings from beyond the teacher expectancy literature. For example, Bjornsdottir and Rule (2017, 2020) have found that people can accurately identify a person's SES based on displays of positive and negative emotions. Specifically, their participants perceived the neutrally-posed faces of higher-SES individuals as exhibiting more positive and less negative affect than the neutrally-posed faces of lower-SES individuals, and they used these affect cues to accurately identify targets' SES. These results are in line with both secondary quantitative affect rating data collected in two of our studies (see Table S8), as well as with a simple visual inspection of the mental representations generated by the participants in our own studies (see Table 2). Indeed, the images of the high-ability student depict more positive and less negative affect than the images of the low-ability student. By assessing people's implicit mental representations of high- versus low- ability students, our research represents an important step in identifying the specific visual cues that some teachers may use to automatically form academic and SES judgments about individual students. And critically, such judgments have the potential to produce expectancy effects that undermine the achievement of vulnerable (e.g., low-SES) students. Thus, to the extent that such cues do exist and are used by teachers to form automatic evaluations of their students' SES and (because of culturally-shared stereotypes linking the two) academic ability, it may be possible to develop interventions that call teachers' attention to their unwanted use of these cues, thus reducing one form of teacher bias effects.

Participants also consistently perceived the students depicted in the low-ability images as lower in SES than the students depicted in the high-ability images, while their racial perceptions of these students were less consistent across studies. Specifically, participants perceived the low-ability student to look more Black and less White and Asian than the high-ability student in some studies, which is consistent with widely-held racial stereotypes. However, these results did not replicate in other studies. Additionally, in supplementary exploratory analyses (see SOM Tables S8), we found that the low-ability student was consistently perceived to be older and more masculine than the high-ability student. This is noteworthy, as perceptions of young students' age and masculinity are negatively associated with perceptions of their academic competence and classroom behavior,

JOURNAL OF Social Issues A Journal of the Society for the Psychological Study of Social Issues

respectively (Glock & Kleen, 2017; Mallman & Lee, 2016). By contrast, participants in Studies 1–3 perceived the low-ability student to look significantly less Hispanic and multiracial than the high-ability student, which runs contrary to prevailing racial-ethnic stereotypes about intellectual abilities. However, the effect sizes were much larger for perceived differences in SES than they were for any of the perceived racial-ethnic difference findings. These findings are clearly complex and deserve further exploration, as quantifying the relative influence of various social categories in people's mental representations of low- and high-ability students was not a goal of the present studies. The present studies also cannot establish whether observers first perceive the student in terms of a particular SES, and then use this information to make inferences about the student's other social attributes (e.g., age, gender, race, ethnicity), as well as their academic capabilities, or whether this sequence occurs in a different order. Future research is required to directly address these questions.

That being said, the current findings do suggest that people's mental representations of high- and low-ability students may have intersectional associations with a range of different social categories (Valle & Covarrubias, 2024). That is, in American society, people may form visual representations of low-ability students by automatically combining or integrating their representations of multiple groups that are stereotyped as less academically capable or well-behaved (e.g., lower-SES students, male students, older students, Black students). Alternatively, they may incorporate visual cues (e.g., partially closed eyes) that are associated with certain characteristics (e.g., laziness) into their mental representations of all the categories that are stereotyped as exhibiting these characteristics (e.g., low ability, low SES, Black, etc.). In either case, the intersectional nature of people's ability-based representations suggests that people's low-ability student representation may be particularly likely to be activated when they have categorized a present student as belonging to multiple negatively-stereotyped social categories. And, the activation of this representation may lead them to provide that student with relatively low levels of academic support.

These findings also have potential implications for educational policy and advocacy. Prior research has shown that the extent to which certain social groups are depicted as lazy, unmotivated troublemakers drives opposition to policies designed to support those individuals (Brown-Iannuzzi et al., 2017; Rose & Baumgartner, 2013). Relatedly, if the images that come to mind when Americans think of low-ability students (or of students from groups that are stereotyped as being low-ability) are of lazy unmotivated troublemakers, versus motivated hard-working youth facing major barriers, this may reduce backing for policies designed to provide for these students. It therefore seems plausible that the extent to which the mental representations corresponding to particular ability levels and particular (minoritized) social categories overlap may contribute to the disproportionate allocation of educational opportunities to particular (non-minoritized) students. Supporting this possibility, the composite images of the low (vs. high) ability student were consistently rated as lazier, less hardworking, more likely to exhibit problematic behavior in school, and (in secondary ratings; see Table S8) as more masculine, hostile, and threatening. While the present research was not designed to thoroughly examine how SES might meaningfully intersect with other identities that are associated with academic stereotypes, it is noteworthy that these findings align with widely-held intersectional stereotypes about Black, male, and low-SES individuals, specifically (Okonofua & Eberhardt, 2015; Rose & Baumgartner, 2013). Future research should provide a more extensive investigation of this possible intersectionality in people's mental representations of high- and low-ability students.

Finally, it is noteworthy that teachers were much more positive in their evaluations and treatment of the lower-ability student (see Table 4). In addition, in their comments about the study, some teachers explicitly expressed that they found it problematic to base their judgments purely on facial images. These findings suggest that there may be some elements of teacher training or classroom experience that can help reduce the tendency to categorize and interact with students based on mental representations. However, due to self-presentation concerns, Study 6 may also underestimate the actual effects of teachers' mental representations on the kinds of judgments and interactions that occur in authentic classroom contexts, where teachers may be responding to students in a more automatic manner. In addition, while sensitivity analyses suggested that the sample size was adequate to detect the noted results (see SOM Table S1), the in-service educator sample was very small. Future research with larger and more representative samples of educators should examine how teacher training and experience might reduce or reinforce reliance on these culturally-shared mental representations of students.

In addition to these contributions, it is important to acknowledge the limitations of the present work. First, we did not preregister the methods or analytic approaches employed in the present studies. Although we have tried to be as transparent as possible about the major decisions we made regarding the methods and analyses used in this work, we acknowledge that the lack of preregistration may have unintentionally introduced some amount of experimenter error at the analytic and reporting stages of the project.

Second, while this research suggests that people's mental representations of high- and low-ability students may be imbued with SES-related cues, this work did not capture people's specific representations of high- and low-SES students. However, the potential overlap between ability-based and SES-based mental representations is supported by prior research that has directly assessed participants' mental representations of low- versus high-SES individuals. Specifically, Lei and Bodenhausen (2017) used the same image-generation methodology to capture participants' mental representations of "poor" and "rich" individuals. Complementing the present findings, they found that people held qualitatively distinct mental representations of "poor" and "rich" individuals. And, when separate groups of participants were asked to rate these images, the results showed that images depicting "poor" individuals were rated as being significantly lazier, less intelligent, and less motivated than images depicting "rich" individuals.

Finally, the image-generation method employed does not allow us to determine the *extent* to which the generated images reflect people's true mental images of high- and low-ability students. Indeed, experts in this technique have noted that "while reverse correlation aims to visualize the content of mental representations, it can… only provide an approximation of the true mental representations" (Brinkman et al., 2017, p. 334). This might raise questions about the potential influence of demand effects. That is, participants may be responding to the image generation task not by drawing on representations of high- and low-ability students that they are carrying around in their minds, but by spontaneously creating mental images that reflect their understanding of the task stimuli and instructions (i.e., that reflect the specific demands of the task). However, we believe that the way in which the task was administered actually makes this possibility unlikely. Specifically, participants were *not instructed to generate* an image of a low- or high-ability student. Rather, they were "shown several pairs of blurry faces" and were instructed, "for each pair of faces that you will see, we want you to tell us which one looks most like a student with" low or high academic abilities, and to "decide quickly … [and] base your responses on your immediate 'gut' reactions to the photos." Thus, it seems unlikely that, before making their selections, participants spontaneously generated an image of a low-ability student for the sole purpose of completing the task.

ACKNOWLEDGMENTS

This work was supported by funding from the College of the Holy Cross (awarded to ASB) and from the James S. McDonnell Foundation (Collaborative Grant #220020483, awarded to DBM). The authors thank Ashley Dumais, Gabriela Jimenez-Thompson, Camille McCobb, Caroline Muniz, and Ahana Nagarkatti for their assistance with data collection.

CONFLICT TO INTEREST STATEMENT

The authors hereby declare that they have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

For all studies in this paper, see https://osf.io/snr97/ for materials, data, and analytic syntax, including those not relevant to the present research questions.

ORCID

Alexander S. Browman https://orcid.org/0000-0002-2957-3262

REFERENCES

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, white women. *Health Psychology*, *19*(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586

Autin, F., Batruch, A., & Butera, F. (2019). The function of selection of assessment leads evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology*, *111*(4), 717–735. https://doi.org/10.1037/edu0000307

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2016). An unintentional, robust, and replicable pro-black bias in social judgment. *Social Cognition*, *34*(1), 1.

Axt, J. R., Nguyen, H., & Nosek, B. A. (2018). The judgment bias task: A flexible method for assessing individual differences in social judgment biases. *Journal of Experimental Social Psychology*, *76*, 337–355. https://doi.org/10.1016/j.jesp.2018.02.011

Batruch, A., Autin, F., Bataillard, F., & Butera, F. (2019). School selection and the social class divide: How tracking contributes to the reproduction of inequalities. *Personality and Social Psychology Bulletin*, *45*(3), 477–490. https://doi.org/10.1177/0146167218791804

Batruch, A., Autin, F., & Butera, F. (2017). Re-establishing the social-class order: Restorative reactions against high-achieving, low-SES pupils. *Journal of Social Issues*, *73*(1), 42–60. https://doi.org/10.1111/josi.12203

Batruch, A., Geven, S., Kessenich, E., & Van De Werfhorst, H. G. (2023). Are tracking recommendations biased? A review of teachers' role in the creation of inequalities in tracking decisions. *Teaching and Teacher Education*, *123*, 103985. https://doi.org/10.1016/j.tate.2022.103985

Becker, J. C., Kraus, M. W., & Rheinschmidt-Same, M. (2017). Cultural expressions of social class and their implications for group-related beliefs and behaviors. *Journal of Social Issues*, *73*(1), 158–174. https://doi.org/10.1111/josi.12209

Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of Personality and Social Psychology*, *113*(4), 530–546. https://doi.org/10.1037/pspa0000091.supp

Bjornsdottir, R. T., & Rule, N. O. (2020). Negative emotion and perceived social class. *Emotion*, *20*(6), 1031–1041. https://doi.org/10.1037/emo0000613

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Brown-Iannuzzi, J. L., Cooley, E., Marshburn, C. K., McKee, S. E., & Lei, R. F. (2021). Investigating the interplay between race, work ethic stereotypes, and attitudes toward welfare recipients and policies. *Social Psychological and Personality Science*, *12*(7), 1155–1164. https://doi.org/10.1177/1948550620983051

Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological Science*, *28*(1), 92–103. https://doi.org/10.1177/0956797616674999

Brown-Iannuzzi, J. L., McKee, S., & Gervais, W. M. (2018). Atheist horns and religious halos: Mental representations of atheists and theists. *Journal of Experimental Psychology: General*, *147*(2), 292–297. https://doi.org/10.1037/xge0000376

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2021). Type I error is inflated in the two-phase reverse correlation procedure. *Social Psychological and Personality Science*, *12*(5), 760–768. https://doi.org/10.1177/1948550620938616

Cozzarelli, C., Wilkinson, A. V., & Tagler, M. J. (2001). Attitudes toward the poor and attributions for poverty. *Journal of Social Issues*, *57*(2), 207–227. https://doi.org/10.1111/0022-4537.00209

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), 1–33. https://doi.org/10.1348/014466608X314935

Dotsch, R., Wigboldus, D. H. J., Langner, O., & Knippenberg, A. van. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Doyle, L., & Easterbrook, M. J. (2024). Biased career choices? It depends what you believe: Trainee teachers' aversions to working in low-income schools are moderated by beliefs about inequality, meritocracy, and growth mindsets. *Journal of Social Issues*. https://doi.org/10.1111/josi.12648

Doyle, L., Easterbrook, M. J., & Harris, P. R. (2023). Roles of socioeconomic status, ethnicity and teacher beliefs in academic grading. *British Journal of Educational Psychology*, *93*(1), 91–112. https://doi.org/10.1111/bjep.12541

Doyle, L., Easterbrook, M. J., & Harris, P. R. (2024). It's a problem, but not mine: Exploring bias-related message acceptance among teachers. *Social Psychology of Education*, *27*(3), 909–933. https://doi.org/10.1007/s11218-023-09832-9

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

Gallup News. (2016). *Most in U.S. Oppose colleges considering race in admissions*. https://news.gallup.com/poll/193508/oppose-colleges-considering-race-admissions.aspx

Gawronski, B., & Hahn, A. (2018). *Implicit measures: Procedures, use, and interpretation*. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in Social Psychology* (pp. 29–55). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780429452925-2

Glock, S., & Kleen, H. (2017). Gender and student misbehavior: Evidence from implicit and explicit measures. *Teaching and Teacher Education*, *67*, 93–103. https://doi.org/10.1016/j.tate.2017.05.015

Good, T. L., Sterzinger, N., & Lavigne, A. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, *24*(3-5), 99–123. https://doi.org/10.1080/13803611.2018.1548817

Hutchings, R. J., Simpson, A. J., Sherman, J. W., & Todd, A. R. (2021). Perspective taking reduces intergroup bias in visual representations of faces. *Cognition*, *214*, 104808. https://doi.org/10.1016/j.cognition.2021.104808

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science*, *22*(12), 1583–1590. https://doi.org/10.1177/0956797611419675

Kleisner, K., Chvátalová, V., & Flegr, J. (2014). Perceived intelligence is associated with measured intelligence in men but not women. *PLoS ONE*, *9*(3), e81237. https://doi.org/10.1371/journal.pone.0081237

Lei, R. F., & Bodenhausen, G. V. (2017). Racial assumptions color the mental representation of social class. *Frontiers in Psychology*, *8*, 1–7. https://doi.org/10.3389/fpsyg.2017.00519

Mallman, M., & Lee, H. (2016). Stigmatised learners: Mature-age students negotiating university culture. *British Journal of Sociology of Education*, *37*(5), 684–701. https://doi.org/10.1080/01425692.2014.973017

Miele, D. B., Perez, S. A., Butler, R., Browman, A. S., O'Dwyer, L. M., & McNeish, D. (2019). Elementary school teachers' growth mindsets predict their differential treatment of high versus low ability students. PsyArXiv. https://doi.org/10.31234/osf.io/qcd83

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Muenks, K., Miele, D. B., & Wigfield, A. (2016). How students' perceptions of the source of effort influence their ability evaluations of other students. *Journal of Educational Psychology*, *108*(3), 438–454. https://doi.org/10.1037/edu0000068

National Center for Education Statistics. (2023). *Characteristics of Public School Teachers*. https://nces.ed.gov/programs/coe/indicator/clr/public-school-teachers

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624. https://doi.org/10.1177/0956797615570365

Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social–psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, *11*(3), 381–398. https://doi.org/10.1177/1745691616635592

Park, D., Gunderson, E. A., Tsukayama, E., Levine, S. C., & Beilock, S. L. (2016). Young children's motivational frameworks and math achievement: Relation to teacher-reported instructional practices, but not teacher theory of intelligence. *Journal of Educational Psychology*, *108*(3), 300–313. https://doi.org/10.1037/edu0000064

Paul, M., Gaither, S. E., & Darity, W. (2022). About face: Seeing class and race. *Journal of Economic Issues*, *56*(1), 1–17. https://doi.org/10.1080/00213624.2022.2008750

Pew Research Center. (2019). *Most Americans say colleges should not consider race or ethnicity in admissions*. Pew Research Center.

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., Knippenberg, A. van, & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, *106*(6), 897–911. https://doi.org/10.1037/a0036498

Rose, M., & Baumgartner, F. R. (2013). Framing the poor: Media coverage and U.S. Poverty Policy, 1960–2008. *Policy Studies Journal*, *41*(1), 22–53. https://doi.org/10.1111/psj.12001

Rule, N. O., Ambady, N., Adams Jr, R. B., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology*, *95*(5), 1019.

Schmid Mast, M., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior*, *28*(3), 145–165. https://doi.org/10.1023/B:JONB.0000039647.94190.21

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*(2), 213–226. https://doi.org/10.1016/S0742-051X(00)00052-4

The Associated Press-NORC Center for Public Affairs Research. (2019). *Perceptions of college admissions practices*.

The Educational Opportunity Project at Stanford University. (2024). The 2009–2019 Educational Opportunity Explorer. In *The Educational Opportunity Project at Stanford University*. https://edopportunity.org/

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

US Census Bureau. (2023). *U.S. Census Bureau QuickFacts: United States*. https://www.census.gov/quickfacts/fact/table/US/PST045222

Valle, I., & Covarrubias, R. (2024). A critical race, interdisciplinary approach to examining socioeconomic disparities in higher education. *Journal of Social Issues*.

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

Woods, T. A., Kurtz-Costes, B., & Rowley, S. J. (2005). The development of stereotypes about the rich and poor: Age, race, and family income differences in beliefs. *Journal of Youth and Adolescence*, *34*(5), 437–445. https://doi.org/10.1007/s10964-005-7261-0

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*, *28*(2), 238–249. https://doi.org/10.1177/0146167202282009

Zhang, X., Yan, R., Sun, S., & Zuo, B. (2021). Facial expression stereotypes of rich and poor adults and children. *Cognitive Processing*, *22*(4), 649–657. https://doi.org/10.1007/s10339-021-01040-7

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Browman, A. S., & Miele, D. B. (2024). Are low-ability students mentally represented as low-SES, academically incapable, and undeserving of support? *Journal of Social Issues*, 1–26. https://doi.org/10.1111/josi.12649

AUTHOR BIOGRAPHIES

Alexander S. Browman, Ph.D., is an Assistant Professor of Psychology at the College of the Holy Cross. A social and educational psychologist by training, his research explores the beliefs that people hold about education and educational disparities, and how understanding these beliefs can help make education more equitable. In particular, he examines how students, teachers, and the public come to internalize inequities in educational contexts and policies, and the associated consequences academic motivation and performance, especially for students from less advantaged, lower-opportunity backgrounds. He then works to develop interventions designed to reform these internalized beliefs and the structural policies and practices that maintain them.

David B. Miele, Ph.D., is an Associate Professor in the Department of Counseling, Developmental, and Educational Psychology at Boston College. He investigates students' beliefs about their ability, effort, and motivation, and examines how these beliefs influence their goal pursuit and self-regulation in academic contexts. At the broadest level, he is interested in how students can develop into effective, independent learners. And, though much of his research has examined the motivation of college students, he is also interested in the learning and development of elementary school students and adolescents. In addition, he conducts research with parents and teachers in order to better understand how their beliefs influence the ways in which they support students' learning and motivation.

Supplementary Information for "Are Low-Ability Students Are Mentally Represented as Low-SES, Academically Incapable, and Undeserving of Support?"

Length of Studies and Compensation

Participation in the image-generation phase of Study 1 took ~10 minutes to complete and paid $1.50. Participation in the image-generation phase of Study 2 took ~30 minutes to complete and compensation was partial course credit. Participation in the image-rating phases of Studies 1-2, and in Studies 3-4 took ~5 minutes to complete and paid $0.75. Participation in the image-generation phase of Study 5 took ~15 minutes to complete and paid $3.75. Study 6 was completed as part of a larger multi-wave project, for which teachers were paid up to $70 for their participation.

Data Exclusions

The final samples in the image generation phases of Studies 1 and 2 excluded 28 and 75 responses from Studies 1 and 2, respectively, for one or more of the following reasons: the participant did not complete the entire image-generation task; the participant had an IP address or location data that either was not American or was flagged as suspicious (identified using https://itaysisso.shinyapps.io/Bots; see Dennis et al., 2019); the participant began or completed the study twice, so their repeated completions were removed; or the participant did not plan to become an educator (in Study 2).

The final samples in the image rating phases of Studies 1 and 2 excluded 4 and 4 responses from Studies 1 and 2, respectively, for one or more of the following reasons: the participant had an IP address or location data that either was not American or was flagged as suspicious; or they provided text box responses that suggested that they were a bot.

The final sample in Study 3 excluded the responses of 3 participants who had an IP address or location data that either was not American or was flagged as suspicious.

The final sample in Study 4 excluded 36 responses for one or more of the following reasons: the participant began but did not complete any or much of the study; the participant had an IP address or location data that either was not American or was flagged as suspicious; or the participant failed a formal attention check ("This item is here to screen out random responding; do not give a response to this item").

This final sample in Study 5 excluded 41 responses for one or more of the following reasons: the participant began but did not complete any or much of the study; the participant completed the study twice, so their second response was removed; the participant's Connect ID number was not captured, so we could not determine whether they had completed the study more than once; the participant had an IP address or location data that either was not American or was flagged as suspicious; or the participant failed a formal attention check ("I endorse the importance of screening out random responding, so I will select 'strongly disagree.'").

In Study 6, 7 teachers provided deliberately invalid (i.e., straightlined) responses to the academic rating task because, as noted in the General Discussion, they explicitly expressed to us that they found it problematic to base their judgments purely on facial images. Thus, these participants' academic rating data were excluded from our analyses.

Sensitivity Analyses

Table S1 displays the results from sensitivity analyses for each study in the present work. These indicate the smallest effect size of interest that each study could reliably detect with three levels of statistical power—.80, .99, and .999. The smallest significant effect that emerged in each study is also provided for comparison. As shown, Studies 1-5 provided statistical power of > .999 to detect all observed effects. Study 6 provided statistical power of > .999 to detect the majority of effects of the sizes observed, while it provided a power level between .99 and .999 to detect one observed effect, between .80 and .99 to detect another, and slightly < .80 to detect another. Overall, then, studies were well-powered to detect our effects of interest.

Table S1

Sensitivity analyses for all studies, including the smallest effect size of interest that each study could reliably detect with statistical power = .80, .99. and .999, and the smallest effect actually observed in each study for comparison.

	.80	.99	.999	Smallest effect observed
Study 1				
Image-generation phase: correlations between pixels (*r*)	0.127	0.192	0.226	0.905
Image-rating phase: within-subjects differences between images (*d*)	0.201	0.308	0.363	0.16
Study 2				
Image-generation phase: correlations between pixels (*r*)	0.239	0.357	0.413	0.537
Image-rating phase: within-subjects differences between images (*d*)	0.200	0.305	0.360	0.18
Study 3: Between-subjects differences between images (*d*)	0.326	0.499	0.588	0.05
Study 4: Within-subjects differences between images (*d*)	0.166	0.254	0.299	1.08
Study 5: Within-subjects differences between images (*d*)	0.160	0.245	0.289	1.28
Study 6: Within-subjects differences between images (*d*)	0.448	0.686	0.809	0.55

Addressing the Potential for Inflated Type I Error Rates.

The procedures used in the present work were based on established techniques (e.g., Brown-Iannuzzi et al., 2017; Dotsch et al., 2008). However, some scholars have recently cautioned that the image-averaging method used in the image-generation phases of Studies 1 and 2 have the potential to inflate Type I error rates (Cone et al., 2021). Screening Studies 1, 2, 3, 4, and 6 as recommended by Cone et al. (2021) revealed that (a) it is unlikely that Type I error inflation was an issue in Studies 1, 2, and 6, as the image-rating samples were smaller than the corresponding image-generation samples (see Table 1), and (b) even though Studies 3 and 4 did involve more image-raters than image-generators, for the least significant *p*-value that emerged (likelihood of cheating in Study 3: p = 6e-20) to become non-significant would require the same level of Type I error inflation as conducting ~8.3×10^{17} comparison tests (based on the Bonferroni method of adjusting p-values). It therefore seems unlikely that Type I error inflation influenced the conclusions of the present work. However, to further confirm that the present results are not primarily the result of Type I error inflation, Study 5 used a different image-compositing technique that is less likely to influence Type I error rates (creating and rating multiple subgroup images; Cone et al., 2021) and found similar results.

Potential Moderation by Beliefs about the Fixedness or Malleability of Intellectual Abilities

In addition to the primary research questions addressed in the main text, we also examined a secondary question in the present studies. While some view intellectual ability as malleable (a growth mindset), others view it as innate and fixed (a fixed mindset), and therefore tend to believe that some people inherently have less ability than others (Dweck, 2000). To the extent that fixed mindset individuals assume that innate characteristics are associated with one's physical appearance, they may believe that levels of intellectual ability can be accurately inferred just by looking at someone (Keller, 2005; Suzuki et al., 2017; Thomas & Sarnecka, 2015). Thus, we examined whether people with a growth (versus fixed) mindset are less likely to differentially represent low- and high-ability students, or to evaluate students' academic attributes and make decisions about whether or not to provide them with academic support based on their visual appearance.

To test whether differences in people's mental representations of high- and low-ability students might be moderated by their beliefs about the fixedness or malleability of intelligence, participants in Studies 1, 2, 3, 4, and 6 completed a validated 8-item measure (e.g., "You have a certain amount of intelligence, and you can't really do much to change it"; Dweck, 2000), using a 1 ("strongly disagree") to 6 ("strongly agree") response scale; *M* s = 3.15 (Study 1 image-generation), 2.95 (Study 2 image-generation), 3.14 (Study 1 image-rating), 3.18 (Study 2 image-rating), 3.17 (Study 3), 3.21 (Study 4), and 2.26 (Study 6); *SD*s = 1.28, 0.92, 1.47, 1.45, 1.36, 1.28, and 0.91; ω_us = 0.98, 0.94, 0.99, 0.98, 0.98, 0.98, and 0.98.

The results of these analyses are described in the following section, but we provide a summary and discussion here. We found that participants with stronger fixed mindsets about intelligence (i.e., participants whose mindsets were in the top tertile of responses) produced similar high- and low-ability faces to those produced by participants with weaker fixed mindsets about intelligence (i.e., participants whose mindsets were in the bottom tertile of responses; see SOM Table S2). There was also little evidence across the five studies that these image rating disparities were mitigated among the public or teachers with a growth mindset: only 5 of the 59 interaction terms tested reached significance, and only 1 of those 4 effects replicated in a second study (see SOM Table S3). In other words, the findings reported in this work—that the public and aspiring educators' representations of lower (versus higher) ability students are associated with poorer academic attributes and behavioral support—were not reliably moderated by participants' beliefs about the nature of intellectual ability.

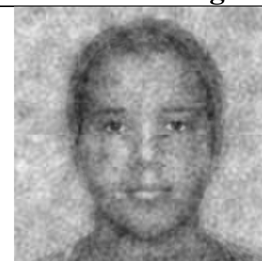
These results are noteworthy, as prior research has suggested that the magnitude of teacher biases and their consequences for students might be moderated by teachers' beliefs about the fixedness or malleability of intelligence (i.e., teacher growth mindsets; e.g., Canning et al., 2019). Specifically, a teacher who believes that intellectual abilities are malleable may be more likely to treat students equally than a teacher who believes that some people inherently and permanently have less ability than others. However, across our 5 studies, we found little evidence that people's growth mindsets influenced either how they mentally represented low- and high-ability students or how they evaluated and supported these students. The present work therefore also contributes to the literature on potential moderators of expectancy effects, as they suggest that such mental representations may be culturally shared by many teachers and members of the public, regardless of their personal beliefs about the nature of ability. As a result, interventions designed to modify beliefs about the nature of intellectual ability may not be helpful when it comes specifically to reducing educators' reliance on culturally-shared mental representations about students' academic ability levels, at least not among those well represented in the present samples—White, female educators working in high-achieving districts that predominantly serve higher-SES White and Asian students. Future research should therefore test whether the present findings hold among different and larger samples of educators and members of the public.

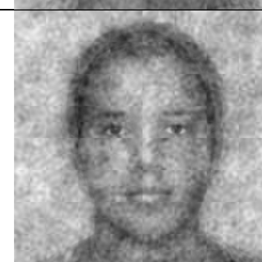
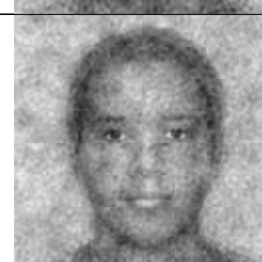
Moderation of Image-Generation Results by Participants' Race, Gender, SES, and Fixed Mindsets

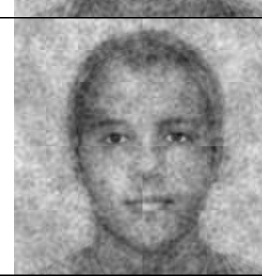
We tested whether the image-generation results varied as a result of participants' gender identities (male-identified versus female-identified, as only a very small number of participants indicated a gender identity other than male or female), racial identities (White versus non-White, as our samples did not include enough participants from other individual racial groups to conduct more granular analyses), SES levels (participants with scores in the 1st versus 3rd tertiles for income and subjective SES), beliefs about the nature of intellectual abilities (participants with scores in the 1st versus 3rd tertiles for fixed mindsets), or non-aspiring educators (in Study 2). As Table S2 demonstrates, the images produced were qualitatively very similar regardless of these participant characteristics. In addition, Figure S1 demonstrates that across studies and participant subgroups, the pixel brightness of same-condition images (e.g., the high-ability student generated by male participants, and the high-ability student generated by female participants) were positively correlated, while the pixel brightness of across-condition images (e.g., the high-ability student generated by male participants, and the low-ability student generated by male participants) were negatively correlated. These quantitative results align with those discussed in the main text.

Table S2

Images generated by subsamples in Studies 1-2.

	Low-ability student image	**High-ability student image**
Male-identified participants (Study 1)		
Male-identified participants (Study 2)		
Female-identified participants (Study 1)		
Female-identified participants (Study 2)		
White-only participants (Study 1)		
White-only participants (Study 2)		

Non-White-only participants (Study 1)		
Non-White-only participants (Study 2)		
Stronger fixed mindsets (Study 1)		
Stronger fixed mindsets (Study 2)		
Weaker fixed mindsets (Study 1)		
Weaker fixed mindsets (Study 2)		

Lower income participants (Study 1)		
Lower income participants (Study 2)		
Higher income participants (Study 1)		
Higher income participants (Study 2)		
Lower subjective SES participants (Study 1)		
Lower subjective SES participants (Study 2)		

Higher subjective SES participants (Study 1)		
Higher subjective SES participants (Study 2)		
Non-aspiring educators (Study 2)		

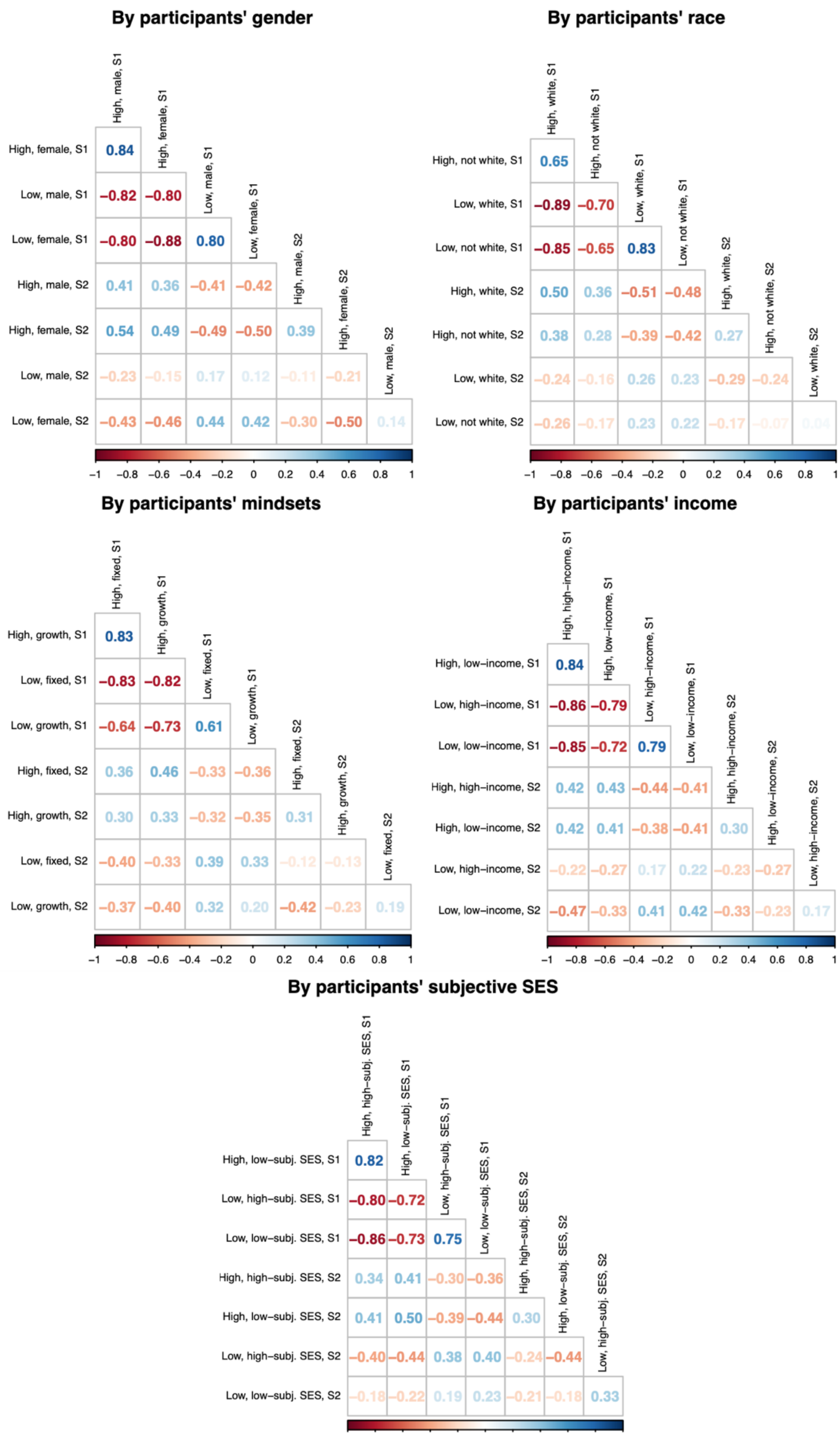


Figure S1.

Correlations among the pixel brightness levels of imaged generated by participant subgroups (high/low = image ability level).

Moderation of Image-Rating Results by Participants' Race, Gender, SES, and Fixed Mindsets

We also tested whether the image-rating results varied as a result of participants' gender identities (male-identified versus female-identified, as only a very small number of participants indicated a gender identity other than male or female), racial identities (White versus non-White, as our samples did not include enough participants from other individual racial groups to conduct more granular analyses), income (continuous: 1 = "Under $15,000"; 9 = "Over $200,000"), or subjective SES (continuous; self-reported position on a 10-runged ladder, ranging from "Rung 1 (least money, least education, least respected/no jobs)" to "Rung 10 (most money, most education, most respected jobs)").

The "support measures" referenced are (a) self-reported and (b) behavioral support for providing scholarship support in Study 4, granting college admissions in Study 5, and support for using unsupportive instructional practices in Study 6. Note that we were not permitted to collect data on teachers' income or subjective SES in Study 6, and there was not sufficient racial or gender diversity in this sample (only 2 male-identified teachers and only 2 non-White teacher) to conduct those analyses either.

As Table S3 demonstrates (and as discussed in the main text), there was little evidence across the five studies that the image rating disparities noted between the high- and low-ability image were reliably moderated by participants' beliefs about the nature of intellectual ability; only 5 of the 59 image × mindset interaction terms tested reached significance.

Table S3
Image × fixed mindset interactions predicting academic ratings and perceived SES in all studies.

	Study	*F*	*df*	*p*	Partial η^2
SES	1	0.44	1, 97	.507	0.005
	2	1.11	1, 100	.295	0.011
	3	0.40	1, 293	.527	0.001
	4	0.86	1, 285	.353	0.003

Intelligence	1	0.48	1, 95	.491	0.005
	2	7.88	1, 95	.006	0.077
	3	3.96	1, 293	.047	0.013
	4	0.21	1, 285	.644	0.001
	6	0.29	1, 31	.595	0.009
Academic competence	1	1.19	1, 95	.279	0.012
	2	1.79	1, 95	.184	0.019
	3	2.50	1, 293	.115	0.008
	4	1.21	1, 285	.273	0.004
	6	1.22	1, 31	.277	0.038
Academically motivated	1	0.16	1, 95	.693	0.002
	2	4.36	1, 95	.040	0.044
	3	0.92	1, 293	.339	0.003
	4	0.08	1, 285	.779	0.000
	6	2.25	1, 31	.143	0.068
Academic confidence	1	0.79	1, 95	.375	0.008
	2	2.27	1, 95	.135	0.023
	3	1.83	1, 293	.177	0.006
Academically hardworking	1	0.04	1, 95	.847	0.000
	2	1.35	1, 95	.249	0.014
	3	1.55	1, 293	.214	0.005
	4	0.23	1, 285	.632	0.001
	6	1.67	1, 31	.206	0.051
Academically lazy	1	0.79	1, 95	.377	0.008
	2	0.52	1, 95	.471	0.005
	3	10.90	1, 293	.001	0.036
	4	1.88	1, 285	.172	0.007
Likely to finish high school	1	0.87	1, 95	.352	0.009
	2	1.00	1, 95	.320	0.010
	3	0.71	1, 293	.399	0.002
	6	1.21	1, 31	.280	0.038
Likely to finish college	1	2.41	1, 95	.124	0.025
	2	2.32	1, 95	.131	0.024
	3	0.15	1, 293	.698	0.001
	4	1.11	1, 285	.293	0.004
Receptive to feedback in school	1	0.37	1, 95	.545	0.004
	2	2.43	1, 95	.122	0.025
	3	0.29	1, 293	.589	0.001
Likely to follow directions in school	1	0.36	1, 95	.548	0.004
	2	2.88	1, 95	.093	0.029
	3	1.03	1, 293	.310	0.004
	6	2.98	1, 31	.094	0.088
Likelihood of problem behavior in school	1	0.26	1, 95	.611	0.003
	2	0.22	1, 95	.637	0.002
	3	0.67	1, 293	.413	0.002
	6	2.51	1, 31	.123	0.075
Likely to cheat	1	0.06	1, 95	.804	0.001
	2	1.07	1, 95	.304	0.011
	3	0.66	1, 293	.416	0.002
Likely to complete assignments on time	1	0.11	1, 95	.738	0.001
	2	0.59	1, 95	.444	0.006
	3	2.28	1, 293	.132	0.008
Support measures	4 (a)	0.79	1, 285	.374	0.003
	4 (b)	z = -2.31	1, 285	.021	OR = 0.545
	6	0.14	1, 39	.713	0.004

Table S4 shows that there was no evidence across studies that these image rating disparities reliably varied between White and non-White participants; 0 of the 57 image × race interaction terms tested reached significance.

Table S4

Target image × participant race interactions predicting academic ratings and perceived SES in all studies.

	Study	*F*	*df*	*p*	Partial η^2
SES	1	0.01	1, 97	.928	0.000
	2	0.17	1, 99	.679	0.002
	3	2.43	1, 293	.120	0.008
	4	2.98	1, 284	.085	0.010
	5	0.00	1, 304	.949	0.000
Intelligence	1	0.08	1, 95	.775	0.001
	2	3.53	1, 95	.063	0.036
	3	0.10	1, 293	.756	0.000
	4	0.78	1, 284	.378	0.003
Academically motivated	1	0.00	1, 95	.949	0.000
	2	1.84	1, 95	.179	0.019
	3	0.00	1, 293	.997	0.000
	4	0.75	1, 284	.388	0.003
	5	1.42	1, 304	.235	0.005
Academic competence	1	3.72	1, 95	.057	0.038
	2	2.17	1, 95	.144	0.022
	3	0.10	1, 293	.749	0.000
	4	1.42	1, 284	.234	0.005
Academic confidence	1	0.01	1, 95	.911	0.000
	2	1.02	1, 95	.314	0.011
	3	0.26	1, 293	.611	0.001
Academically hardworking	1	0.15	1, 95	.700	0.002
	2	0.87	1, 95	.354	0.009
	3	0.00	1, 293	.959	0.000
	4	1.86	1, 284	.173	0.007
	5	1.41	1, 304	.236	0.005
Academically lazy	1	0.88	1, 95	.351	0.009
	2	1.97	1, 95	.163	0.020
	3	0.09	1, 293	.769	0.000
	4	0.02	1, 284	.890	0.000
Likely to finish high school	1	0.33	1, 95	.565	0.004
	2	0.23	1, 95	.635	0.002
	3	1.39	1, 293	.240	0.005
Likely to finish college	1	1.18	1, 95	.281	0.012
	2	2.28	1, 95	.134	0.023
	3	0.17	1, 293	.683	0.001
	4	1.55	1, 284	.214	0.005
	5	1.20	1, 304	.273	0.004
Receptive to feedback in school	1	1.29	1, 95	.258	0.013
	2	1.46	1, 95	.230	0.015
	3	0.74	1, 293	.390	0.003
Likely to follow directions in school	1	0.33	1, 95	.567	0.003

	2	2.25	1, 95	.137	0.023
	3	0.44	1, 293	.507	0.002
Likelihood of problem behavior in school	1	0.53	1, 95	.469	0.006
	2	1.28	1, 95	.260	0.013
	3	0.02	1, 293	.877	0.000
Likely to cheat	1	0.00	1, 95	.991	0.000
	2	0.17	1, 95	.681	0.002
	3	1.31	1, 293	.253	0.004
Likely to complete assignments on time	1	0.34	1, 95	.562	0.004
	2	2.62	1, 95	.109	0.027
	3	0.37	1, 293	.545	0.001
	5	1.35	1, 304	.246	0.004
Support measures	4 (a)	0.79	1, 284	.374	0.003
	4 (b)	z = -0.57	1, 284	.568	OR = 0.703
	5	0.60	1, 290	.438	0.002

Table S5 shows that there was little evidence across studies that these image rating disparities reliably varied between male-identified and female-identified participants; only 3 of the 57 image × gender interaction terms tested reached significance.

Table S5

Target image × participant gender interactions predicting academic ratings and perceived SES in all studies.

	Study	*F*	*df*	*p*	Partial η^2
SES	1	5.93	1, 97	.017	0.058
	2	0.24	1, 99	.627	0.002
	3	1.86	1, 293	.174	0.006
	4	2.55	1, 281	.111	0.009
	5	1.87	1, 302	.172	0.006
Intelligence	1	2.28	1, 95	.134	0.023
	2	1.58	1, 94	.212	0.017
	3	0.59	1, 293	.444	0.002
	4	2.24	1, 281	.135	0.008
Academic competence	1	4.13	1, 95	.045	0.042
	2	0.07	1, 94	.791	0.001
	3	0.91	1, 293	.340	0.003
	4	3.23	1, 281	.073	0.011
	5	2.31	1, 302	.130	0.008
Academically motivated	1	3.25	1, 95	.074	0.033
	2	2.08	1, 94	.152	0.022
	3	0.35	1, 293	.553	0.001
	4	0.24	1, 281	.623	0.001
Academic confidence	1	0.94	1, 95	.334	0.010
	2	0.91	1, 94	.341	0.010
	3	0.08	1, 293	.777	0.000
Academic hardworking	1	3.17	1, 95	.078	0.032
	2	2.43	1, 94	.122	0.025
	3	0.87	1, 293	.351	0.003
	4	0.32	1, 281	.570	0.001
	5	1.06	1, 302	.305	0.003
Academically lazy	1	2.84	1, 95	.095	0.029

	2	0.51	1, 94	.479	0.005
	3	1.41	1, 293	.237	0.005
	4	0.00	1, 281	.976	0.000
Likely to finish high school	1	0.27	1, 95	.607	0.003
	2	0.23	1, 94	.632	0.002
	3	0.63	1, 293	.429	0.002
Likely to finish college	1	0.51	1, 95	.479	0.005
	2	0.60	1, 94	.441	0.006
	3	0.65	1, 293	.420	0.002
	4	1.20	1, 281	.275	0.004
	5	2.83	1, 302	.094	0.009
Receptive to feedback in school	1	0.45	1, 95	.504	0.005
	2	0.18	1, 94	.675	0.002
	3	0.23	1, 293	.634	0.001
Likely to follow directions in school	1	0.05	1, 95	.832	0.000
	2	0.14	1, 94	.709	0.001
	3	0.13	1, 293	.716	0.000
Likelihood of problem behavior in school	1	0.00	1, 95	.981	0.000
	2	0.06	1, 94	.809	0.001
	3	0.01	1, 293	.913	0.000
Likely to cheat	1	0.08	1, 95	.774	0.001
	2	0.09	1, 94	.761	0.001
	3	2.02	1, 293	.157	0.007
Likely to complete assignments on time	1	2.78	1, 95	.099	0.028
	2	0.00	1, 94	.947	0.000
	3	0.21	1, 293	.646	0.001
	5	1.99	1, 302	.159	0.007
Support measures	4 (a)	7.60	1, 281	.006	0.026
	4 (b)	$z = -1.06$	1, 281	.289	OR = 0.539
	5	1.58	1, 287	.210	0.005

Table S6 shows that there was little evidence across studies that these image rating disparities reliably varied based on participants' income; only 2 of the 57 image × income interaction terms tested reached significance.

Table S6

Target image × income interactions predicting academic ratings and perceived SES in all studies.

	Study	F	df	p	Partial η^2
SES	1	0.09	1, 97	.766	0.001
	2	1.12	1, 99	.292	0.011
	3	0.66	1, 293	.416	0.002
	4	0.18	1, 285	.671	0.001
	5	1.05	1, 304	.307	0.003
Intelligence	1	1.23	1, 95	.271	0.013
	2	0.14	1, 95	.708	0.001
	3	2.07	1, 293	.152	0.007
	4	1.62	1, 285	.205	0.006
Academic competence	1	0.01	1, 95	.930	0.000
	2	0.22	1, 95	.642	0.002
	3	0.10	1, 293	.753	0.000

	4	5.52	1, 285	.020	0.019
	5	0.05	1, 304	.821	0.000
Academically motivated	1	0.09	1, 95	.761	0.001
	2	0.00	1, 95	.976	0.000
	3	0.61	1, 293	.436	0.002
	4	0.05	1, 285	.816	0.000
Academic confidence	1	1.10	1, 95	.297	0.011
	2	0.02	1, 95	.887	0.000
	3	1.93	1, 293	.165	0.007
Academic hardworking	1	0.24	1, 95	.625	0.003
	2	0.62	1, 95	.432	0.007
	3	0.06	1, 293	.809	0.000
	4	0.94	1, 285	.334	0.003
	5	0.00	1, 304	.999	0.000
Academically lazy	1	0.86	1, 95	.355	0.009
	2	0.03	1, 95	.852	0.000
	3	1.29	1, 293	.258	0.004
	4	1.22	1, 285	.270	0.004
Likely to finish high school	1	0.73	1, 95	.394	0.008
	2	0.34	1, 95	.561	0.004
	3	0.67	1, 293	.414	0.002
Likely to finish college	1	0.63	1, 95	.430	0.007
	2	0.01	1, 95	.927	0.000
	3	3.70	1, 293	.055	0.012
	4	0.51	1, 285	.476	0.002
	5	0.01	1, 304	.903	0.000
Receptive to feedback in school	1	1.15	1, 95	.286	0.012
	2	0.82	1, 95	.369	0.009
	3	0.12	1, 293	.731	0.000
Likely to follow directions in school	1	0.00	1, 95	.978	0.000
	2	0.53	1, 95	.469	0.006
	3	0.32	1, 293	.573	0.001
Likelihood of problem behavior in school	1	0.67	1, 95	.413	0.007
	2	0.44	1, 95	.508	0.005
	3	0.91	1, 293	.340	0.003
Likely to cheat	1	0.02	1, 95	.882	0.000
	2	1.15	1, 95	.285	0.012
	3	0.94	1, 293	.332	0.003
Likely to complete assignments on time	1	0.10	1, 95	.756	0.001
	2	0.04	1, 95	.843	0.000
	3	0.90	1, 293	.343	0.003
	5	0.00	1, 304	.967	0.000
Support measures	4 (a)	1.65	1, 285	.200	0.006
	4 (b)	z = -1.27	1, 285	.204	OR = 0.819
	5	4.47	1, 289	.035	0.015

Finally, Table S7 that there was little evidence across studies that these image rating disparities reliably varied as a result of participants' subjective SES; only 2 of the 57 image × subjective SES interaction terms tested reached significance. In other words, the findings reported in this work were not reliably moderated by participants' beliefs about the nature of intellectual ability, whether they were White or non-White, whether they identified as male or female, or based on their income or subjective SES. These findings provide further support for the possibility that such mental representations may be culturally transmitted and shared.

Table S7

Target image × subjective SES interactions predicting academic ratings and perceived SES in all studies.

	Study	F	df	p	Partial η^2
SES	1	0.00	1, 97	.948	0.000
	2	0.04	1, 100	.843	0.000
	3	0.05	1, 293	.822	0.000
	4	0.74	1, 285	.389	0.003
	5	0.60	1, 305	.441	0.002
Intelligence	1	0.37	1, 95	.547	0.004
	2	0.00	1, 95	.988	0.000
	3	0.00	1, 293	.994	0.000
	4	1.32	1, 285	.252	0.005
Academic competence	1	0.09	1, 95	.760	0.001
	2	0.50	1, 95	.483	0.005
	3	0.48	1, 293	.488	0.002
	4	3.35	1, 285	.068	0.012
	5	4.28	1, 305	.040	0.014
Academically motivated	1	0.35	1, 95	.554	0.004
	2	0.10	1, 95	.748	0.001
	3	1.64	1, 293	.202	0.006
	4	0.96	1, 285	.327	0.003
Academic confidence	1	0.03	1, 95	.854	0.000
	2	0.72	1, 95	.399	0.008
	3	1.51	1, 293	.219	0.005
Academic hardworking	1	0.51	1, 95	.475	0.005
	2	0.82	1, 95	.366	0.009
	3	1.11	1, 293	.294	0.004
	4	0.64	1, 285	.424	0.002
	5	4.49	1, 305	.035	0.015
Academically lazy	1	0.03	1, 95	.856	0.000
	2	0.19	1, 95	.661	0.002
	3	0.56	1, 293	.455	0.002
	4	2.05	1, 285	.154	0.007
Likely to finish high school	1	3.11	1, 95	.081	0.032
	2	0.26	1, 95	.615	0.003
	3	0.02	1, 293	.893	0.000
Likely to finish college	1	0.63	1, 95	.431	0.007
	2	0.10	1, 95	.758	0.001
	3	0.14	1, 293	.704	0.000
	4	0.39	1, 285	.535	0.001
	5	3.46	1, 305	.064	0.011
Receptive to feedback in school	1	1.35	1, 95	.248	0.014
	2	0.10	1, 95	.758	0.001
	3	1.12	1, 293	.291	0.004
Likely to follow directions in school	1	0.17	1, 95	.678	0.002
	2	0.34	1, 95	.563	0.004

	3	0.96	1, 293	.329	0.003
Likelihood of problem behavior in school	1	0.11	1, 95	.746	0.001
	2	0.01	1, 95	.905	0.000
	3	0.00	1, 293	.965	0.000
Likely to cheat	1	1.10	1, 95	.297	0.011
	2	0.76	1, 95	.386	0.008
	3	0.35	1, 293	.556	0.001
Likely to complete assignments on time	1	0.02	1, 95	.885	0.000
	2	0.06	1, 95	.805	0.001
	3	0.02	1, 293	.894	0.000
	5	3.18	1, 305	.076	0.010
Support measures	4 (a)	1.13	1, 285	.289	0.004
	4 (b)	$z = -0.59$	1, 285	.557	OR = 0.901
	5	0.67	1, 290	.414	0.002

Additional Image Ratings

As noted in the main text, in Studies 1-2, the group of participants assigned to rate the images in terms of SES also rated the images in terms of other social categories.

Specifically, they indicated the extent to which the students in the images looked "masculine" and "feminine" (1 = "not at all" to 6 = "extremely"), as well as how old the student looked (1 = "Much younger than the average high school student" to 7 = "Much older than the average high school student"). In addition, a third group of participants in Study 2 was assigned to rate the images in terms of a number of personal attributes: "respectful," "trustworthy," "physically attractive," "happy," "likable," "a good person," "kind," "gentle," "warm," "hostile," "threatening," "aggressive," "angry," "approachable," and "dominant" (1 = "not at all" to 6 = "extremely"). Study 3 participants also completed these same additional social category and personal attribute ratings.

As shown in Table S8, compared to the higher-ability image, the lower-ability image was consistently perceived to be more hostile, threatening, aggressive, angry, dominant, and masculine, as well as less respectful, trustworthy, happy, likeable, like a good person, kind, gentle, warm, approachable, and feminine. Thus, while the present research was not designed to thoroughly examine how SES might meaningfully intersect with other identities that are associated with academic stereotypes, these findings suggest that people's mental representations of low-ability students may be a product of widely-held intersectional stereotypes about low-SES, Black, *and* male individuals, specifically (Okonofua & Eberhardt, 2015; Rose & Baumgartner, 2013). However, all of these additional social category differences had much smaller effect sizes than nearly all of the primary results highlighted in the present work. Thus, the strongest and most consistent social categorization difference that emerged in the present work was the perceived difference in SES between the images. Future research would be needed to provide substantive support for the notion that people's mental representations of high- and low-ability students are in fact intersectional in nature.

Table S8
Mean (SDs) perceptions and t-tests of social category and personal characteristics ratings in Studies 1-3.

	Study	High-ability student image [*M* (*SD*)]	Low-ability student image [*M* (*SD*)]	*t*	*df*	*p*	Cohen's *d*
Respectful	2	4.60 (1.09)	2.14 (1.16)	15.47	97	< .001	1.56
	3	4.61 (1.03)	2.34 (1.02)	19.01	295	< .001	2.21
Trustworthy	2	4.44 (1.18)	2.20 (1.17)	13.90	97	< .001	1.40
	3	4.52 (1.05)	2.33 (1.01)	18.34	295	< .001	2.13
Physical attractive	2	3.78 (1.28)	2.17 (1.06)	10.18	97	< .001	1.03
	3	3.85 (1.25)	2.31 (1.04)	11.52	280.16	< .001	1.34
Happy	2	4.58 (1.07)	1.50 (0.99)	20.19	97	< .001	2.04
	3	4.15 (1.20)	1.37 (0.70)	24.23	228.56	< .001	2.85
Likeable	2	4.59 (1.18)	2.01 (1.08)	16.98	97	< .001	1.72
	3	4.65 (1.12)	2.24 (1.01)	19.44	295	< .001	2.26
Like a good person	2	4.68 (1.00)	2.32 (1.15)	15.75	97	< .001	1.59
	3	4.70 (1.08)	2.61 (0.99)	17.44	295	< .001	2.02
Kind	2	4.51 (1.15)	1.94 (1.07)	15.55	97	< .001	1.57
	3	4.53 (1.12)	2.05 (1.05)	19.74	295	< .001	2.29
Gentle	2	4.45 (1.09)	1.89 (1.06)	18.02	97	< .001	1.82
	3	4.41 (1.14)	1.97 (1.10)	18.79	295	< .001	2.18
Warm	2	4.60 (1.03)	1.72 (1.03)	17.65	97	< .001	1.78
	3	4.43 (1.10)	1.83 (0.98)	21.65	295	< .001	2.51
Hostile	2	1.63 (0.96)	4.35 (1.42)	-16.17	97	< .001	-1.63
	3	1.48 (0.97)	4.07 (1.36)	-18.95	272.33	< .001	-2.18
Threatening	2	1.68 (0.89)	4.20 (1.44)	-14.94	97	< .001	-1.51
	3	1.60 (0.89)	3.98 (1.31)	-18.43	266.22	< .001	-2.12
Aggressive	2	1.76 (1.02)	4.06 (1.50)	-12.75	97	< .001	-1.29
	3	1.71 (1.03)	4.09 (1.28)	-17.70	287.37	< .001	-2.04
Angry	2	1.54 (0.95)	4.10 (1.48)	-14.00	97	< .001	-1.41
	3	1.38 (0.73)	4.26 (1.34)	-23.08	234.65	< .001	-2.64
Approachable	2	4.81 (1.02)	1.73 (0.93)	21.25	97	< .001	2.15

	3	4.70 (1.13)	1.87 (0.97)	23.17	295	< .001	2.69
Dominant	2	2.64 (1.20)	3.81 (1.46)	-5.44	97	< .001	-0.55
	3	2.80 (1.20)	3.80 (1.29)	-6.87	295	< .001	-0.80
Masculine	1	2.71 (1.39)	4.46 (1.28)	-12.00	98	< .001	-1.21
	2	3.32 (1.37)	4.34 (1.33)	-6.79	101	< .001	-0.67
	3	3.59 (1.32)	4.50 (1.21)	-6.24	295	< .001	-0.72
Feminine	1	3.79 (1.54)	2.00 (1.17)	11.80	97	< .001	1.19
	2	3.26 (1.56)	2.12 (1.31)	6.52	101	< .001	0.65
	3	2.77 (1.41)	1.88 (1.18)	5.94	281.31	< .001	0.69
Perceived age	1	3.82 (0.68)	5.37 (0.98)	-15.14	98	< .001	-1.52
	2	3.88 (0.88)	5.13 (1.01)	-9.35	101	< .001	-0.92
	3	4.12 (0.84)	4.66 (1.02)	-4.92	288.68	< .001	-0.57

Reliability Metrics for the Teaching Behavior Measures in Study 6

As discussed in the main text, the teachers in Study 6 indicated how likely they would be to engage in a five motivationally supportive and five motivationally unsupportive teaching behaviors with both the student pictured in the high-ability image and the student pictured in the low-ability image. Thus, each participant responded to a total of 20 items.

In determining how best to quantify internal consistency for these scales, we relied on recent research showing that Cronbach's α makes rigid assumptions that can lead to considerable downward bias, particularly when scales have a small number of items or are multidimensional (Flora, 2020; McNeish, 2018). Indeed, a previous study, in which we administered the same self-report measure to a large sample of teachers, prompted us to investigate omega because of the low Cronbach's α observed (Miele et al., 2019).[1] Furthermore, research has shown that Pearson correlations (which are used to compute both α and *standard* omega coefficients [ω_u]) can misestimate the relation between two Likert-type (ordinal) items that exhibit skewed response distributions, and can therefore bias reliability and factor analytic estimates (Baglin, 2014; Foldnes & Grønneberg, 2021; Gadermann et al., 2019; Yang & Green, 2011). Because many of

[1] In our previous study, the teachers were asked to think about a student who they perceived as having high or low ability in the math or verbal domain and to imagine that the student was struggling to complete a "math" or "verbal or language arts" assignment in their classroom, rather than (as in the present Study 6) looking at an image of student who was described as struggling on a math assignment.

the items in our scales exhibited a high level of skew (in our previous teacher study, 12 of the 20 items had a skew > |1| and 4 of these had a skew > |2|; in the current study, 17 of the 20 items had a skew > |1|, and 4 of these had a skew > |2|), we therefore made an initial decision to report categorical omega ($w_{\text{u-cat}}$) as the primary index of internal consistency for these scales (see Table S9). These were computed in R based on polychoric correlation matrices—as polychoric correlations generally (though not always) produce less biased estimates than Pearson correlations (Flora, 2020; McNeish, 2018)—following the steps described by Flora (2020).

It is important to note that, because (a) the sample was small (Lorenzo-Seva & Ferrando, 2021), (b) the polychoric correlation between two items from one scale was close to 1, (c) the variance-covariance matrices from the CFAs for two other scales were positive semidefinite (rather than positive definite), (d) not all CFA models had adequate fit or fit statistics that appeared reliable (Flora, 2020), and (e) three of the models had a low standardized loading (< .4) for one item, the $\omega_{\text{u-cat}}$ reliability estimates should be treated with caution. For this reason, we reported α (a more standard measure of reliability that does not rely on factor analysis) rather than $\omega_{\text{u-cat}}$ in the main manuscript.

In our previous study (Miele et al., 2019), which included a much larger sample of teachers (N = 245), we conducted exploratory factor analyses (EFAs) and confirmatory factor analyses (CFAs) of our teaching behaviors scale. EFAs suggested that the four of the five supportive items load onto one factor, and that all five unsupportive items load onto a second factor. However, CFAs of the same two factor-model suggest relatively poor fit overall, with some low loadings for the supportive factor.

To allow for comparison of internal consistency coefficients across studies, we re-estimated $\omega_{\text{u-cat}}$ for the scales administered in our previous study following procedures described by Flora (2020) (see Table S9). The single-factor CFA models used to recompute $\omega_{\text{u-cat}}$

for the two unsupportive practices scales (one for a low-ability student and another for a high-ability student) did not have adequate fit, though the standardized loadings were >.4 for all items. Both CFA models used to recompute $\omega_{\text{u-cat}}$ for the two supportive practices scales appeared to have at least marginal fit, but three of the standardized loadings for one of the scales were <.4.

It is possible that the pattern of high residual correlations (>|.1|) that we observed in the two CFA models of the unsupportive practice scales in our previous study corresponded to a meaningful source of variance that was unrelated to the factor of interest, thus inflating our estimates of $\omega_{\text{u-cat}}$. Indeed, Flora (2020) has argued that $\omega_{\text{u-cat}}$ "estimates depend on correct specification of the model underlying a given test (e.g., ω_{u} is not an appropriate reliability estimate if the population model is multidimensional, as evidenced by poor fit of a one-factor model)" (p. 497). Thus, in our previous study, we re-computed the CFAs with four error-covariance parameters added (see Flora, 2020, p. 490), and then used the output of the CFAs to recompute omega (see $\omega_{\text{u-cat-res}}$ in Table S9). The respecified CFA models had very good fit.

We then attempted to follow the same procedure for the two unsupportive practices scales in the current study, by adding the same error-covariance parameters that we identified in our previous study. We then computed $\omega_{\text{u-cat-res}}$ based on the output of these models (see Table S9). For one of the scales, the model converged successfully, though the covariance matrix appeared to be positive semidefinite (rather than positive definite). For the other scale, the covariance matrix again appeared to be positive semidefinite. We therefore had to remove one of the error-covariance parameters before finding an admissible solution, and though the fit for this model was an improvement, it was still poor. We re-computed omega based on the output of these models (see Table S9). However, for the reasons noted above, these estimates should also be treated with caution.

Overall, we emphasize that the internal consistency estimates for the teaching behaviors scales in both studies (Study 6 and our previous study) should be treated with caution. However, we view the estimates from the previous study as somewhat more reliable than the estimates for Study 6.

Table S9

Internal consistency reliability coefficients for the four teacher practice scales used in Study 6, and for same scales used in a larger study of elementary school teachers (Miele et al., 2019).

Study	Practice type	Image	α	ω	$\omega_{\text{u-cat}}$	$\omega_{\text{u-cat-res}}$
Study 6 from the present work ($N = 41$)	Supportive practices	High-ability student	0.53	0.52	0.59	-
		Low-ability student	0.58	0.60	0.69	-
	Unsupportive practices	High-ability student	0.43	0.55	0.59	0.48
		Low-ability student	0.44	0.49	0.61	0.44
A larger study of teachers from other work (Miele et al., 2019)	Supportive practices	High-ability student	0.54	0.54	0.57	-
		Low-ability student	0.53	0.54	0.56	-
	Unsupportive practices	High-ability student	0.56	0.59	0.62	0.58
		Low-ability student	0.55	0.57	0.59	0.58

Data Analytic Software

All statistical analyses described in this work were conducted using R (Version 4.4.1; R Core Team, 2021) and the R-packages *afex* (Version 1.4.1; Singmann et al., 2021), *broom* (Version 1.0.6; Robinson et al., 2021), *car* (Fox et al., 2020; Version 3.1.2; Fox & Weisberg, 2019), *carData* (Version 3.0.5; Fox et al., 2020), *corrplot2024* (Wei & Simko, 2024), *data.table* (Version 1.16.0; Dowle & Srinivasan, 2021), *dplyr* (Version 1.1.4; Wickham et al., 2022), *effsize* (Version 0.8.1; Torchiano, 2020), *emmeans* (Version 1.10.4; Lenth, 2021), *ggplot2* (Version 3.5.1; Wickham, 2016), *heplots* (Version 1.7.0; Friendly, 2007, 2010), *jpeg* (Version 0.1.10; Urbanek, 2021), *kableExtra* (Version 1.4.0; Zhu, 2021), *knitr* (Version 1.48; Xie, 2015), *lavaan* (Version 0.6.18; Rosseel, 2012), *lme4* (Version 1.1.35.5; Bates et al., 2015), *Matrix* (Version 1.7.0; Bates & Maechler, 2021), *papaja* (Version 0.1.2; Aust & Barth, 2022), *plyr* (Wickham, 2011; Version 1.8.9; Wickham et al., 2022), *psych* (Version 2.4.6.26; Revelle, 2021), *pwr* (Version 1.3.0; Champely, 2020), *qualtRics* (Version 3.2.1; Ginn et al., 2022), *questionr* (Version

0.7.8; Barnier et al., 2022), *rcicr* (Version 1.0.1; Dotsch, 2017), *rempsyc* (Thériault, 2022), *semTools* (Version 0.5.6; Jorgensen et al., 2021), *tidyr* (Version 1.3.1; Wickham, 2021), *tinylabels* (Version 0.2.4; Barth, 2022), and *VGAM* (Yee, 2010, 2013, 2020; Yee et al., 2015; Yee & Hadi, 2014; Version 1.1.11; Yee & Wild, 1996).

Contributor Roles Taxonomy (CRediT) Statement

Conceptualization: ASB

Methodology: ASB, DBM

Software: ASB

Validation: ASB

Formal Analysis: ASB

Investigation: ASB

Resources: ASB, DBM

Data curation: ASB

Writing – Original Draft: ASB

Writing – Review & Editing: ASB, DBM

Visualization: ASB

Supervision: ASB

Project administration: ASB

Funding acquisition: ASB, DBM

(Version Date: October 11, 2024 at 17:02:03 EDT)

References

Aust, F., & Barth, M. (2022). papaja: Prepare reproducible APA journal articles with R Markdown. Retrieved from https://github.com/crsh/papaja

Baglin, J. (2014). Improving Your Exploratory Factor Analysis for Ordinal Data: A Demonstration Using FACTOR. *Practical Assessment, Research & Evaluation*, 19 (5), 1–14. https://doi.org/10.1146/annurev.psych.53.100901.135239

Barnier, J., Briatte, F., & Larmarange, J. (2022). Questionr: Functions to make surveys processing easier. Retrieved from https://CRAN.R-project.org/package=questionr

Barth, M. (2022). tinylabels: Lightweight variable labels. Retrieved from https://cran.r-project.org/package=tinylabels

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., & Maechler, M. (2021). Matrix: Sparse and dense matrix classes and methods. Retrieved from https://CRAN.R-project.org/package=Matrix

Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological Science*, 28(1), 92–103. https://doi.org/10.1177/0956797616674999

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, 5, eaau4734. https://doi.org/10.1126/sciadv.aau4734

Champely, S. (2020). Pwr: Basic functions for power analysis. Retrieved from https://CRAN.R-project.org/package=pwr

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2021). Type I Error Is Inflated in the Two-Phase Reverse Correlation Procedure. *Social Psychological and Personality Science*, 12(5), 760–768. https://doi.org/10.1177/1948550620938616

Dotsch, R. (2017). Rcicr: Reverse correlation image classification toolbox.

Dotsch, R., Wigboldus, D. H. J., Langner, O., & Knippenberg, A. van. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, 19 (10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Dowle, M., & Srinivasan, A. (2021). Data.table: Extension of ‘data.frame‘. Retrieved from https://CRAN.R-project.org/package=data.table

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Lillington, NC: Taylor & Francis.

Flora, D. B. (2020). Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates.

Advances in Methods and Practices in Psychological Science, 3 (4), 484–501. https://doi.org/10.1177/2515245920951747

Foldnes, N., & Grønneberg, S. (2021). The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychological Methods*. https://doi.org/10.1037/met0000385

Fox, J., & Weisberg, S. (2019). An R companion to applied regression (Third). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fox, J., Weisberg, S., & Price, B. (2020). carData: Companion to applied regression data sets. Retrieved from https://CRAN.R-project.org/package=carData

Friendly, M. (2007). HE plots for multivariate general linear models. *Journal of Computational and Graphical Statistics*, 16(4), 421–444.

Friendly, M. (2010). HE plots for repeated measures designs. *Journal of Statistical Software*, 37(4), 1–40. https://doi.org/10.18637/jss.v037.i04

Gadermann, A., Guhn, M., & Zumbo, B. (2019). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation*, 17 (1). https://doi.org/10.7275/n560-j767Ginn, J., O'Brien, J., & Silge, J. (2022). qualtRics: Download 'qualtrics' survey data. Retrieved from https://CRAN.R-project.org/package=qualtRics

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). semTools: Useful tools for structural equation modeling. Retrieved from https://CRAN.R-project.org/package=semTools

Keller, J. (2005). In genes we trust: The biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology*, 88(4), 686–702. https://doi.org/10.1037/0022-3514.88.4.686

Lenth, R. V. (2021). Emmeans: Estimated marginal means, aka least-squares means. Retrieved from https://CRAN.R-project.org/package=emmeans

Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not Positive Definite Correlation Matrices in Exploratory Item Factor Analysis: Causes, Consequences and a Proposed Solution. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 138–147. https://doi.org/10.1080/10705511.2020.1735393

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. https://doi.org/10.1037/met0000144

Miele, D. B., Perez, S. A., Butler, R., Browman, A. S., O'Dwyer, L. M., & McNeish, D. (2019). Elementary School Teachers' Growth Mindsets Predict Their Differential Treatment of High Versus Low Ability Students. *PsyArXiv*. https://doi.org/10.31234/osf.io/qcd83

Okonofua, J. A., & Eberhardt, J. L. (2015). Two Strikes: Race and the Disciplining of Young Students. *Psychological Science*, 26(5), 617–624. https://doi.org/10.1177/0956797615570365

R Core Team. (2021). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R- project.org/

Revelle, W. (2021). Psych: Procedures for psychological, psychometric, and personality research. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Robinson, D., Hayes, A., & Couch, S. (2021). Broom: Convert statistical objects into tidy tibbles.

Rose, M., & Baumgartner, F. R. (2013). Framing the Poor: Media Coverage and U.S. Poverty Policy, 1960-2008. *Policy Studies Journal*, 41 (1), 22–53. https://doi.org/10.1111/psj.12001

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48 (2), 1–36. Retrieved from https://www.jstatsoft.org/v48/i02/

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). Afex: Analysis of factorial experiments. Retrieved from https://CRAN.R- project.org/package=afex

Suzuki, A., Tsukamoto, S., & Takahashi, Y. (2017). Faces tell everything in a just and biologically determined world: Lay theories behind face reading. *Social Psychological and Personality Science*, 10(1), 62–72. https://doi.org/10.1177/1948550617734616

Thériault, R. (2022). rempsyc: Convenience functions for psychology. Retrieved from https://rempsyc.remi-theriault.com

Thomas, A. J., & Sarnecka, B. W. (2015). Exploring the relation between people's theories of intelligence and beliefs about brain development. *Frontiers in Psychology*, 6(921). https://doi.org/10.3389/fpsyg.2015.00921

Torchiano, M. (2020). Effsize: Efficient effect size computation. https://doi.org/10.5281/zenodo.1480624

Urbanek, S. (2021). Jpeg: Read and write JPEG images. Retrieved from https://CRAN.R-project.org/package=jpeg

Wei, T., & Simko, V. (2024). R package ‘corrplot’: Visualization of a correlation matrix. Retrieved from https://github.com/taiyun/corrplot

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40 (1), 1–29. Retrieved from http://www.jstatsoft.org/v40/i01/

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2021). Tidyr: Tidy messy data. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., & Müller, K. (2022). Dplyr: A grammar of data manipulation. Retrieved from https://CRAN.R-project.org/package=dplyr

Xie, Y. (2015). Dynamic documents with R and knitr (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.org/knitr/

Yang, Y., & Green, S. B. (2011). Coefficient Alpha: A Reliability Coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, 29 (4), 377–392. https://doi.org/10.1177/0734282911406668

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32(10), 1–34. Retrieved from https://www.jstatsoft.org/v32/i10/

Yee, T. W. (2013). Two-parameter reduced-rank vector generalized linear models. *Computational Statistics and Data Analysis*. Retrieved from https://ees.elsevier.com/csda

Yee, T. W. (2020). The VGAM package for negative binomial regression. *Australian and New Zealand Journal of Statistics*, 61.

Yee, T. W., & Hadi, A. F. (2014). Row-column interaction models, with an R implementation. *Computational Statistics*, 29(6), 1427–1445.

Yee, T. W., Stoklosa, J., & Huggins, R. M. (2015). The VGAM package for capture- recapture data using the conditional likelihood. *Journal of Statistical Software*, 65(5), 1–33. Retrieved from https://www.jstatsoft.org/v65/i05/

Yee, T. W., & Wild, C. J. (1996). Vector generalized additive models. *Journal of Royal Statistical Society, Series B*, 58(3), 481–493.

Zhu, H. (2021). kableExtra: Construct complex table with 'kable' and pipe syntax. Retrieved from https://CRAN.R-project.org/package=kableExtra