When Accommodations Are Not Enough:

A Multi-Study Examination of Teacher Bias Toward Students with Special Educational Needs Across Student Gender

Fabian Müller[1], Cristina Aelenei[2], Mickaël Jury[1,3]

[1] Laboratoire ACTé, Université Clermont Auvergne, Clermont-Ferrand, France

[2] Université Paris Cité, Laboratoire de Psychologie Sociale, Boulogne-Billancourt, France

[3] Institut Universitaire de France, Paris, France

Author Note

Fabian Müller https://orcid.org/0000-0002-8363-9285

Cristina Aelenei https://orcid.org/0000-0002-4232-6506

Mickaël Jury https://orcid.org/0000-0003-2004-6682

Preregistration, data, analysis code, and additional online materials are openly available at the project's Open Science Framework page (https://osf.io/ckgh7/?view_only=70fddb44c1a846e3b7291119773829ba). We have no conflicts of interest to disclose. We gratefully acknowledge funding from the Clermont Auvergne Métropole. Parts of this research are planned to be presented at upcoming conferences, including the LICIE 2025 conference. A preprint of this manuscript will

also be made available on the PsyArXiv preprint server (https://psyarxiv.com). The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Fabian M. is now affiliated with Université Paris Cité, CNRS, LaPsyDÉ, F-75005 Paris, France.

Correspondence concerning this article should be addressed to Mickaël Jury, Laboratoire ACTé, INSPE Clermont-Auvergne, 36 avenue Jean-Jaurès CS 20001, 63407 Chamalières Cedex, France. Email: mickael.jury@uca.fr

Word count (excluding references, tables, figures, and appendices): 8970

Author Contributions (CRediT Taxonomy):

- Fabian Müller: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Writing—original draft; Writing—review and editing.
- Cristina Aelenei: Conceptualization; Methodology; Supervision; Writing—review and editing.
- Mickaël Jury: Conceptualization; Funding acquisition; Methodology; Resources; Supervision; Writing—review and editing.

Abstract

This research examines how teachers' evaluations of student performance are influenced by special educational needs (SEN) when accommodations are provided, and whether these effects vary by student gender. Across three preregistered experimental studies (N = 1214) with pre-service and in-service teachers in France, we investigated whether students with SEN receiving reduced-exercise accommodations were systematically devalued in grades and competence judgments, and whether this devaluation – a *backlash effect* – was moderated by fairness perceptions. In Studies 1 and 2, students with SEN received lower grades and competence ratings than non-SEN peers, regardless of student gender or relative performance. Study 3 introduced a cross-gender comparison, testing whether female students with SEN faced heightened backlash compared to male non-SEN peers. A consistent backlash effect emerged across studies, unaffected by gender contrast. Notably, fairness perceptions consistently mitigated this bias. These findings highlight persistent SEN-related backlash and suggest that fairness-focused interventions may foster more inclusive evaluations.

Keywords: teacher bias, special educational needs, gender, fairness perceptions, meritocracy

When Accommodations Are Not Enough:

A Multi-Study Examination of Teacher Bias Toward Students with Special Educational Needs Across Student Gender

A teacher sits down to assess two student assignments. One student completed all the tasks without adjustments; the other required accommodations – fewer questions, extra time, or a quieter space. Both demonstrate similar understanding, yet the teacher hesitates: Should they be graded equally? Was the success of the accommodated student due to merit or an unfair advantage? This vignette reflects a central challenge in inclusive education: balancing fairness with individual needs while upholding meritocratic values.

Inclusive education strives to ensure that all students, including those with special educational needs (SEN[1]), have equitable access to learning and academic success (Ainscow et al., 2019; Nilholm & Göransson, 2017). Achieving this requires adaptable learning environments, supportive teacher attitudes, and fair evaluation practices (Amor et al., 2019; Kefallinou et al., 2020). Exam accommodations – such as extended test time or adapted instructional methods – specifically aim to help students reach and demonstrate their potential. Together, these measures promote fairness by addressing individual need, recognizing that equal treatment does not always mean identical treatment (Deutsch, 1975).

However, even when students with SEN are included in general classrooms and benefit from accommodations, they may still face bias or lowered expectations that hinder their progress. Moreover, one critical yet underexplored issue is how teachers evaluate their academic

[1] While definitions of SEN vary globally (Brussino, 2020; European Commission, 2018), and labels like *special needs* may be experienced as stigmatizing (Gernsbacher et al., 2016), we adopt a person-first approach by referring to students with SEN (Dunn & Andrews, 2015).

performance, and whether these evaluations are influenced by *who* the student is – for example, the student's gender. Recent research raises concerns that students with SEN may be devalued in teachers' grading and competence ratings – especially when these students perform well while receiving accommodations (Stanczak, Aelenei, et al., 2024). Such biases risk undermining inclusion by limiting recognition of students with SEN's abilities. The present research addresses this concern by examining teachers' evaluative judgments of students with SEN, while also considering student gender and a psychological factor (i.e., fairness perceptions) that might influence these judgments.

Stereotypes About Students with Special Educational Needs

Teachers, like anyone, can hold implicit or explicit stereotypes about certain groups of students. Students with SEN are often stereotypically viewed as less competent academically (Krämer & Zimmermann, 2023; Krischler & Pit-ten Cate, 2019). While these stereotypes may stem from well-intentioned assumptions (e.g., seeing students with SEN as "needing help" or "struggling"), they can reinforce expectations of lower ability. Research has documented that students with SEN are presumed to have lower academic potential than their non-SEN peers (Hafen et al., 2015; Shifrer, 2013, 2016; Vlachou et al., 2014). Disability-related stereotypes often combine warmth or pity with perceived incompetence (Clément-Guillotin et al., 2018; Louvet & Rohmer, 2016), potentially leading teachers to underestimate students' academic abilities (Cohen et al., 2019; Krischler & Pit-ten Cate, 2020). These expectations can subtly shape teacher behavior, from the difficulty of assigned material to grading strictness. Classic work on teacher expectations (e.g., the Pygmalion effect) shows that biased expectations can become self-fulfilling, affecting both student performance and teacher interpretations of student behavior (Jussim & Harber, 2005). If a teacher assumes a student with SEN will struggle, they may attribute that student's success to external support rather than ability – and grade them more cautiously.

When students from stigmatized groups exceed stereotypical expectations, they can face a form of bias known as the backlash effect – a penalty or pushback against individuals who defy stereotypes (Rudman et al., 2012). Originally studied in gender contexts (e.g., women who display dominant leadership styles facing social penalties; Rudman & Fairchild, 2004), backlash serves to "punish" counter-stereotypical behavior to preserve cultural stereotypes and the status quo. Applied to education, if a student with SEN (stereotyped as low-achieving) performs at a high level, teachers might unconsciously diminish the student's credit for that success (e.g., lower competence ratings, harsher grading standards, or attributing the success to unfair advantages). Batruch et al. (2017) supported this by demonstrating a backlash pattern in an academic context with low socio-economic status (SES) students. High-achieving, low-SES students were evaluated less favorably than equally high-achieving peers from higher-SES backgrounds. The authors interpreted this as a defensive reaction to unexpected success among disadvantaged students, one that preserves existing beliefs about who deserves to succeed. By analogy, students with SEN who succeed might face a similar backlash, given that they too challenge a stereotype of low competence.

Stanczak, Aelenei, et al. (2024) provide direct empirical evidence consistent with this pattern. In their study with French teachers, identically performing students with and without SEN were not always rated as equally competent. When students with SEN received accommodations perceived as less merit-based (e.g., completing only half of an exercise), they were judged less competent. However, when accommodations seemed more legitimate (e.g., assistive technology), the devaluation was smaller. The authors interpret this as evidence that teachers may use backlash as an "ideological barrier" to full inclusion – essentially, a psychological mechanism that protects the meritocratic status quo in the classroom. Instead of celebrating a student with SEN

who succeeds, a teacher with strong meritocratic leanings might subconsciously think, “Well, of course they did well – they had extra help that others didn’t. It isn’t a real accomplishment”.

Meritocratic Ideology and System Justification in Education

To understand why teachers might engage in biased evaluations, it is important to consider the broader ideological framework of schooling (Florian, 2014). Modern schooling, particularly in industrialized societies, strongly emphasizes meritocracy (Butera et al., 2024) – the belief that academic success should solely reflect individual talent and effort, assuming equal opportunities for all students (Darnon, Wiederkehr, et al., 2018; Mijs, 2016). Ideally, this principle ensures fairness by rewarding students' abilities and hard work, without favoritism. In practice, however, meritocracy assumes a level playing field, overlooking structural disadvantages that necessitate differential support. Consequently, meritocratic discourse can serve an *ideological function*: legitimizing existing inequalities by attributing success or failure solely to individual factors like effort or talent (Darnon, Smeding, et al., 2018; Stanczak, Jury, et al., 2024). If all students are supposedly given the same initial conditions and only personal merit counts, then any outcome inequality may seem justified – high achievers earned their place, while others simply did not measure up (Batruch et al., 2023). This aligns with system justification theory, which argues that people (including educators) are often motivated to defend the status quo as fair, even when it perpetuates injustice (Jost, 2019; Jost et al., 2002).

Within this cultural and ideological framework, inclusive education challenges the meritocratic model by acknowledging that students have different needs and starting points. Achieving equity thus requires differentiated treatment – such as accommodations – rather than uniform standards for all (Sireci et al., 2005). This clashes with a strict meritocratic view of fairness, which assumes that treating everyone the same creates true equality (Benjamin, 2002; De Beco, 2018; Khamzina et al., 2021). As Stanczak, Jury, et al. (2024) argue, there is an ideological

incompatibility between full inclusion of students with SEN and meritocratic principles that resist adapting evaluation practices. Teachers who strongly internalize meritocratic ideals may view accommodations (e.g., extra test time, adjusted standards) as conflicting with the principle of equal competition. In this view, accommodations are not seen as necessary for equity but as distorting merit-based evaluation. A student with SEN's success may then appear less "earned". In other words, if teachers think the student had it "easier" due to an accommodation, they might compensate by grading more strictly or downplaying the student's achievement. Exploratory analyses by Stanczak, Aelenei, et al. (2024) support this interpretation: teachers who perceived accommodations as leveling the field (i.e., fair) rated students more favorably, whereas those who saw them as tilting the field (i.e., unfair) were more likely to discount students' success.

Gender Stereotypes and Intersection with SEN

While research has increasingly explored biases in evaluations of students with SEN, a critical gap remains: how student gender might intersect with SEN status to influence teacher evaluations, as SEN- and gender-related issues seem rather interconnected (Brussino, 2020). Most research treats students with SEN as a homogeneous group, overlooking critical gender-based differences. Boys' overrepresentation in special education – for instance, around 65% in the U.S. (Schaeffer, 2023) – may normalize male SEN diagnoses, whereas female SEN diagnoses could still appear atypical. Specifically, boys are frequently diagnosed with behavioral or neurodevelopmental conditions such as attention-deficit/hyperactivity disorder (ADHD; Hibel et al., 2010; Martin, 2024; OECD, 2005), while girls are more often identified with less visible mental health-related conditions (Thapar et al., 2022). Teachers may therefore expect boys with SEN to struggle academically yet still succeed in stereotypically masculine domains (e.g., mathematics). This expectation aligns with global achievement patterns: boys tend to outperform girls in mathematics, while girls excel in reading (OECD, 2023).

These trends mirror common stereotypes associating mathematical talent with boys and reading abilities with girls – biases that can undermine girls' motivation from an early age (Bian et al., 2017; Jenifer et al., 2024; Leslie et al., 2015) and may also influence how teachers assess competence. Additionally, attribution processes differ by gender: boys' achievements are often linked to assertive traits (e.g., self-confidence), while girls' successes are more frequently attributed to effort (Verniers et al., 2016). A high-achieving girl with SEN may thus violate two sets of expectations – one about gender (especially in stereotypically masculine subjects) and one about disability – potentially invoking unique biases. If her success follows an accommodation implying reduced effort (e.g., fewer assignments), it may be seen as less genuine or "unearned", potentially resulting in a pronounced devaluation. Alternatively, teachers might also be more inclined to give girls the "benefit of the doubt" or show sympathy due to their typically higher classroom engagement and self-regulation (Cornwell et al., 2013), which could help buffer negative bias.

In short, teachers' judgments are not always gender-neutral; they can be influenced by stereotypes such as "girls are better readers" or "boys are naturally better at math", or by behavioral expectations (e.g., girls tend to be more attentive, which teachers may reward in grading). The literature offers conflicting clues, and thus far, no experimental study has systematically examined the intersection of student gender and SEN status in teacher evaluations. Addressing this gap is important for an intersectional understanding of educational equity: policies and trainings need to know if there are "double jeopardy" effects (being female and having SEN compounding bias) or if one stereotype dominates teacher perceptions.

Research Objectives and Overview

Drawing on the above frameworks, this research addresses two critical gaps. First, while prior work suggests that students with SEN may face competence devaluation (Stanczak, Aelenei, et al., 2024), no experimental study has systematically examined whether this bias differs by

gender. It is also unknown whether this bias intensifies when students with SEN outperform their non-SEN peers – thereby increasing perceptions of unexpected success and creating a higher-threat situation (vs. lower-threat, where performance is equal). While Stanczak, Aelenei, et al. (2024) explored the possibility of higher threat, their findings were mixed, highlighting the need for further experimental evaluation. Second, while fairness perceptions have been linked to teacher evaluations of students with SEN, these effects have only been explored post-hoc. No confirmatory research has yet tested whether fairness perceptions moderate backlash against high-performing students with SEN, particularly when gender is considered.

To sum up, Stanczak, Jury, et al. (2024) theorized an incompatibility between inclusive education and meritocratic selection, emphasizing evaluation processes as a central point of tension. They called for empirical research to test these contradictions and their consequences for students with SEN. We respond by extending their work in two key ways: (a) examining whether backlash effects generalize across student gender, and (b) identifying for whom and under what conditions backlash is strongest by probing teachers' fairness perceptions. Through this, we contribute to both theory and practice: providing a more nuanced psychological understanding of teacher biases in inclusive education, and informing interventions that promote fairer recognition of all students' achievements, including those with SEN.

Specifically, the present research addresses the following research questions:

1. Do teachers devalue students with SEN relative to their non-SEN peers, leading to a backlash effect?
2. Does student gender (male vs. female) and performance-based threat (low vs. high) alter the magnitude of bias against students with SEN in teacher evaluations?
3. Do fairness perceptions amplify or mitigate these biases?

We conducted three experimental studies examining the interplay of student gender, performance level (threat), and teacher biases in evaluating students with SEN. Study 1 tested whether student gender and high-threat conditions (outperforming non-SEN peers) amplify backlash effects in pre-service teachers. Study 2 replicated this design with in-service teachers, allowing us to assess whether teaching experience influences these biases. Study 3 refined the experimental design to directly compare backlash effects when a female student with SEN is evaluated against a male student without SEN, further clarifying how gender shapes teacher evaluations.

All analyses were conducted in R (Version 4.2.2). Study materials, data, analysis code, and preregistrations are publicly available on OSF (https://osf.io/ckgh7/?view_only=12b716d0733e4747b3271099ec71c711). The OSF page contains direct links to the preregistrations, which were completed after data collection began but before any analyses were conducted. All studies received prior approval from the Institutional Review Board (N° IRB: 00012024-76).

Study 1

This study examined how students' gender and performance levels shape teachers' grading and competence judgments for students with SEN, and whether fairness perceptions moderate these effects. The preregistered hypotheses were as follows:

Hypothesis 1 (Backlash Effect). Students with SEN receiving accommodations will receive lower grades and competence ratings than their non-SEN peers (H1a). This backlash will be stronger for female than for male students (H1b), and under high-threat conditions, where students with SEN outperform their non-SEN peers, despite equal error rates on math tests (H1c).

Hypothesis 2 (Interaction Effect). The combination of female gender and high threat will amplify the backlash effect, with female students receiving lower grades and competence ratings, particularly in high-threat conditions (i.e., when the student with SEN outperforms the student without SEN).

Hypothesis 3 (Moderation Effect). Perceived fairness of accommodations will moderate the backlash effect. Specifically, the backlash, strongest for female students under high-threat conditions, will intensify when fairness perceptions are low.

Method

Participants and Procedure

Participants. Data were collected online via LimeSurvey from pre-service teachers across France. Multiple INSPE[2] centers distributed the study via their email lists. The goal was to maximize responses within a four-week data collection period in autumn 2024. Although the number of invitations sent is unknown, this approach aimed to obtain a geographically diverse sample. Participation was voluntary and uncompensated. A total of 463 participants completed the experiment (post-imputation, see "Handling Missing Data"; see Table 1 for demographics).

A sensitivity power analysis was conducted using G*Power 3.1.9.4 (Faul et al., 2009) to determine the minimum detectable effect sizes for our hypothesized effects. For the main and interaction effects (student gender, threat level, and their interaction; H1-H2), and with $\alpha = 0.05$, power = 0.80, and a final sample of $N = 463$, the study was powered to detect an effect of partial eta squared, $\eta_p^2 = 0.023$. For the moderation model (H3), the smallest detectable effect size for the key predictor – the three-way interaction between gender, threat level, and perceived fairness – was $\eta p^2 = .017$. These small effects (Cohen, 1988) can hold meaningful implications in education, particularly when accumulated over time or across large populations (Cheung & Slavin, 2016; Götz et al., 2022), justifying the adequacy of our sample size.

Procedure. This study employed a 2 × 2 × 2 mixed-method design, with SEN status (students without SEN vs. students with SEN) as a within-subject factor, and student gender (male vs. female) and threat level (low vs. high) as between-subject factors. Participants were recruited via email and invited to an online study on pedagogical practices. After providing informed consent,

[2] INSPE (*Institut National Supérieur du Professorat et de l'Éducation*): France's national institute for teacher education and training.

they were randomly assigned to one of four experimental conditions (Low Threat, Male; Low Threat, Female; High Threat, Male; High Threat, Female) using LimeSurvey's randomization feature.

Participants viewed two purported math tests from fourth-grade students, presented in successive order. Fourth grade was chosen as it represents a key benchmark in primary education, where teachers begin intensifying competence evaluations and foundational math skills become critical for subsequent academic success (Mullis et al., 2020). One test was attributed to a student without SEN, while the other to a student with ADHD. For the latter, participants were informed that the student had been asked to complete only half of the exercises (or items) due to ADHD accommodations; no such information was provided for the student without SEN.

ADHD was chosen as the SEN context due to its observable behavioral characteristics, such as inattention and impulsivity. These characteristics can disrupt classroom dynamics and often elicit stronger negative attitudes toward students' schooling compared to physical or less behaviorally disruptive neurodevelopmental disabilities, such as dyspraxia (Jury et al., 2021; Stanczak, Aelenei, et al., 2024). These features make ADHD particularly relevant for studying backlash effects.

Student gender was manipulated by assigning traditionally male (Léo, Lucas) or female (Léa, Emma) names to the students. Participants evaluated students of one gender only (i.e., two boys or two girls). Names were counterbalanced and drawn from popular French first names in 2014 (INSEE, 2024) to reflect classroom demographics in 2024. To mirror classroom norms and encourage comparison, the non-SEN student's test was always presented first, as students in ordinary learning conditions represent the majority in classrooms. Threat level was manipulated by varying the difficulty of the errors made by the student with ADHD. In the low-threat condition, both students made errors evenly split between easy and difficult items, indicating similar performance levels. In the high-threat condition, the student with ADHD made errors only on difficult items,

while the non-SEN student's errors remained evenly distributed. This design emphasized the student with SEN's higher competence –error-free performance on easy items and difficulties only with the most challenging ones. Both tests maintained the same overall error rate (i.e., 40%).

After viewing each test, participants graded each student's performance. They then completed a social judgment task assessing competence, effort, and warmth (Louvet & Rohmer, 2016) for both students, judged the fairness of multiple SEN accommodations (Stanczak, Aelenei, et al., 2024), including reduced-exercise allowances.[3] Finally, basic demographics (e.g., age, gender) were collected. All procedures followed ethical guidelines, ensuring voluntary, anonymous participation and a full debriefing.

[3] Participants also completed the Belief in Meritocracy Scale (Wiederkehr et al., 2015), preregistered in the study design. These analyses are reported in the Supplementary Material (SM) to maintain narrative focus.

Table 1

Demographic Characteristics (Study 1, Study 2, and Study 3)

Variable	Category	Study 1: *N* (%)	Study 2: *N* (%)	Study 3: *N* (%)
Gender	Woman	306 (69.4 %)	322 (85.9 %)	308 (87.3%)
	Man	125 (28.3 %)	50 (13.3 %)	44 (12.5%)
	Self-Identified	3 (0.7 %)	2 (0.5 %)	1 (0.3%)
	Prefer Not to Say	7 (1.6 %)	1 (0.3 %)	0 (0.0%)
Year of Study	Year 1 - Undergraduate	4 (0.9 %)	—	—
	Year 2 - Undergraduate	0 (0.0 %)	—	—
	Year 3 - Undergraduate	1 (0.2 %)	—	—
	Year 1 - Postgraduate	222 (50.9 %)	—	—
	Year 2 - Postgraduate	194 (44.6 %)	—	—
	Diplôme d'Université	15 (3.4 %)	—	—
Marginalized Group	No	—	298 (79.9 %)	288 (81.8%)
	Yes (incl. Prefer Not to Say)	—	75 (20.1 %)	64 (18.2%)
Trainee Status	In Training	—	18 (4.8 %)	17 (4.8%)
	Not in Training	—	358 (95.2 %)	336 (95.2%)
Number of Educational	1	—	298 (80.1 %)	281 (80.1%)
Stages Taught	2	—	72 (19.4 %)	67 (19.1%)
	3	—	2 (0.5 %)	3 (0.9%)

Note. For Study 1, *N* = 463; for Study 2, *N* = 387; for Study 3, *N* = 364. Percentages are calculated based on the valid analytic sample (excluding missing values).
Age: Study 1 (*M* = 23.9, *SD* = 5.40, Median = 22, Range = 18–65); Study 2 (*M* = 43.96, *SD* = 9.80, Median = 46, Range = 22–65); Study 3 (*M* = 44.10, *SD* = 9.47, Median = 45, Range = 23–65).
Teaching Experience (Years): Study 2 (*M* = 17.39, *SD* = 9.80, Median = 18, Range = 1–43); Study 3 (*M* = 17.42, *SD* = 9.72, Median = 19, Range = 1–41).
Marginalized Group and Trainee Status were only collected in Studies 2 and 3. Marginalized group refers to participants identifying as part of a group facing discrimination in France (e.g., unequal opportunities or unfair treatment).
Number of Educational Stages Taught: Study 2 and Study 3 include *Maternelle* (Preschool), *Elementaire* (Primary), *College* (Middle), and *Lycée* (High School). The majority of participants taught a single stage (most commonly *Elementaire*). Fewer participants taught across multiple stages (e.g., *Maternelle* and *Elementaire*).
The *Diplôme d'Université* is a French university-specific diploma offering targeted professional or academic training, in this case focused on teacher education.

Measures

Full item wordings are available in the codebook on the project's OSF page.

Grading (Math Tests). Participants graded two math tests using a 10-point scale (1 = competence not acquired, 10 = fully acquired), consistent with assessment practices in French primary education.

Each test, presented separately, included five sections (addition, subtraction, multiplication, word problems, mixed problems), with four items per section (20 items total;

adapted from national evaluations aligned with the French curriculum, Ministère de l'Éducation Nationale et de la Jeunesse, 2023). Each section included two easy and two difficult items, with difficult items highlighted in bold. Participants rated each test immediately after viewing it. In the version attributed to the student without SEN, all 20 items were completed, and eight errors were randomly distributed (four on easy items, four on difficult ones). In the version attributed to the student with SEN, accommodations were simulated using a half-exercise condition: only the first two items (one easy, one difficult) of each section were completed (10 items total). This version contained four errors: in the low-threat condition, errors were evenly split across difficulty levels; in the high-threat condition, all four errors appeared on difficult items. Errors were marked with red crosses to guide attention to overall performance rather than error detection. Error rates were equivalent across both tests (8/20 for the student without SEN vs. 4/10 for the student with SEN). The only differences between the two tests within a given condition were the number of completed items and the distribution of errors, ensuring a controlled comparison.

Perceived Competence. Perceived competence was assessed using the competence subscale of the Social Judgment Scale (Louvet & Rohmer, 2016). The scale comprised 15 items – five each measuring competence, effort, and warmth – rated on a 5-point Likert scale (1 = not at all, 5 = completely). Each item was presented with two side-by-side rating fields – one for the student without SEN (left) and one for the student with SEN (right) – with names displayed according to the assigned gender condition. This format enabled direct student comparisons. The competence subscale included items describing competence-related attributes (e.g., "competent", "efficient", "productive"), with gendered adjectives adapted for grammatical accuracy in French. Items were randomized. Responses were averaged to create composite competence scores for the student without SEN ($\alpha = .83$) and with SEN ($\alpha = .85$), demonstrating good internal consistency. Descriptive

statistics and reliability results for the warmth and effort subscales are presented in Table S1 (Supplementary Material, SM).

Fairness of Accommodations. Participants evaluated the perceived fairness of five accommodations for students with SEN – extra time, oral exams, use of a computer, separate-room assessments, and half-exercise requirements (Stanczak, Aelenei, et al., 2024) – on a 5-point Likert scale (1 = very unfair, 5 = very fair), in the context of overcoming ADHD-related barriers during assessments. To prevent bias and avoid drawing undue attention to the half-exercise accommodation, all items were presented equally to participants in random order. Preregistered moderation analyses focused exclusively on the item concerning the half-exercise accommodation (M = 3.42, SD = 1.22, Range = 1–5), which was mean-centered prior to analysis.

Results

Analysis Strategy

Handling Missing Data. Missing data were found to be missing completely at random (nonparametric test of homoscedasticity, p = .458; Jamshidian & Jalal, 2010; Little, 1988). We addressed missingness using multiple imputation by chained equations (mice; v3.16.0; van Buuren & Groothuis-Oudshoorn, 2011), generating 20 imputed datasets. Participants with >50% missing data on key measures (i.e., grading, perceived competence, and fairness) were excluded before imputation to ensure data quality (Enders, 2022). Outliers were screened using the median absolute deviation (MAD) criterion, excluding cases with completion times more than 2.5 MADs below the median (Leys et al., 2013); no such cases were identified. Likert-scale items, gradings, age, and year of study were treated as continuous and imputed using predictive mean matching (Norman, 2010). Convergence diagnostics confirmed stable imputations.

Confirmatory Factor Analyses. To validate the factor structures of the Social Judgment Scales (SEN and non-SEN), confirmatory factor analyses (CFA)[4] were conducted on the 20 imputed datasets using lavaan (v0.6-19; Rosseel, 2012), estimated via maximum likelihood with robust standard errors. Both versions followed a three-factor structure (competence, effort, and warmth), and model fit was good to excellent: non-SEN version (robust root mean square error of approximation [RMSEA] = .060, 90% confidence interval [CI] [.047, .074]; robust comparative fit index [CFI] = .977; robust Tucker-Lewis index [TLI] = 1.000; standardized root mean square residual [SRMR] = .050); SEN version (robust RMSEA = .070, 90% CI [.057, .083]; robust CFI = .968; robust TLI = 1.000; SRMR = .059). Standardized factor loadings for the competence subscale were moderate to strong: non-SEN (.551–.788) and SEN (.549–.819).

Statistical Analyses. Table 2 presents pooled means and paired-sample *t*-tests for grading and perceived competence, comparing students with and without SEN. Hypotheses were tested using multiple regression analyses on the pooled datasets (handled via mice), with pooling via the mitools package (v2.4; Lumley, 2019). To examine the main (H1) and interaction effects (H2), we tested whether difference scores in grading and perceived competence (non-SEN minus SEN ratings) were greater than zero, indicating a backlash effect. Regression models included student gender (male vs. female) and threat level (low vs. high) as predictors, plus their interaction to assess whether backlash amplified under high threat, particularly for female students. For the moderation analysis (H3), a regression model tested whether perceived fairness of accommodations moderated the impact of student gender and threat level on grading and

[4] While EFA was preregistered to explore factor structures of the Social Judgment Scales (SEN and non-SEN), pooling EFA results across multiple imputations is statistically inappropriate. Therefore, we conducted CFA instead. An exploratory EFA on one imputed dataset supported these theoretical expectations.

competence difference scores.[5] Exploratory analyses tested perceived fairness as an independent predictor of grading and competence ratings, separately for students with and without SEN, to explore whether the effect was stronger for students with SEN. Model assumptions (normality, homoscedasticity, multicollinearity) were conducted in one representative imputed dataset, revealing minor violations (e.g., moderate skewness). To assess robustness, we conducted two sensitivity analyses: (1) robust regression across all imputed datasets, and (2) a complete-case analysis on the non-imputed sample (N = 424). Results were consistent; thus, standard linear models are reported.

Primary Analyses

Grading (Math Tests). Results confirmed a backlash effect: students with SEN received lower grading ratings than their non-SEN peers (b = 0.60, 95% CI [0.44, 0.77], SE = 0.08, t(456) = 7.20, p < .001, η_p^2 = 0.102). However, this effect was unaffected by student gender, threat level, or their interaction (all ps > .320).

Perceived Competence. Similarly, students with SEN were judged as less competent than their non-SEN peers (**b** = 0.22, 95% CI [0.16, 0.27], SE = 0.03, t(452) = 7.69, p < .001, η_p^2 = 0.115). Again, neither student gender, threat level, nor their interaction affected this backlash effect (all ps > .328).

Table 3 summarizes the primary regression results. Controlling for participants' gender did not alter these findings (Table S2 in SM).

Moderation Analyses

[5] Preregistered secondary analyses – including stratified regressions for SEN and non-SEN students and models of perceived effort – are presented in SM.

Fairness of Accommodations. Higher fairness ratings were associated with reduced grading differences (b = -0.30, 95% CI [-0.44, -0.17], SE = 0.07, $t(439) = -4.43$, $p < .001$, $\eta_p^2 = 0.042$) and reduced perceived competence differences (b = -0.07, 95% CI [-0.12, -0.03], SE = 0.02, $t(425) = -3.09$, $p = .002$, $\eta_p^2 = 0.021$) between students with and without SEN. Teachers who viewed accommodations as fair evaluated students with SEN more similarly to their non-SEN peers in grades and competence. However, fairness did not interact with student gender, threat level, or their combination to moderate the backlash effect (all ps > .120; Table S3 in SM for full results).

Exploratory Analyses

To examine whether fairness differentially influenced evaluations, fairness was tested as an independent predictor separately for students with and without SEN. Fairness significantly influenced ratings for both groups, but effects were stronger for students with SEN. For non-SEN students, fairness predicted perceived competence (b = 0.09, 95% CI [0.04, 0.15], SE = 0.03, $t(438) = 3.47$, $p = .001$, $\eta_p^2 = 0.026$) but not grading ($p = .705$). For students with SEN, fairness was a strong predictor of both grading (b = 0.32, 95% CI [0.19, 0.45], SE = 0.06, $t(420) = 5.00$, $p < .001$, $\eta_p^2 = 0.054$) and perceived competence (b = 0.17, 95% CI [0.11, 0.22], SE = 0.03, $t(425) = 5.77$, $p < .001$, $\eta_p^2 = 0.071$). See Table S4 and corresponding section in SM for exploratory interaction results not central to our hypotheses.

Discussion

The goal of Study 1 was to test whether students with SEN receive lower grading and competence evaluations (a "backlash effect") compared to their non-SEN peers (H1a), and whether this bias is magnified by student gender (H1b) and/or performance-based threat (H1c). Additionally, we explored whether gender and threat interact (H2) and whether teachers' fairness perceptions moderate the backlash (H3).

Students with SEN were rated lower in both grading and competence than students without SEN, confirming a backlash effect (H1a). However, neither student gender, performance-based threat, nor their interaction further amplified this backlash (contrary to H1b, H1c, and H2). Thus, while SEN status alone – when accompanied by accommodations – triggered devaluation, neither being female nor surpassing non-SEN peers (high threat) exacerbated that bias among pre-service teachers.

Fairness perceptions emerged as a significant predictor: teachers who regarded the half-exercise accommodation as fair rated students with SEN more favorably, mitigating the size of the backlash. However, fairness did not interact with student gender or threat, yielding no evidence that fairness matters more under female-student or high-threat conditions (disconfirming H3).

Exploratory analyses further indicated that fairness perceptions influenced teacher judgments for both students with and without SEN, but the effect was stronger for students with SEN. This suggests that fairness concerns are more salient – and perhaps more consequential – in inclusive education contexts, where accommodations challenge normative standards of performance.

Overall, Study 1 shows that SEN status alone invites devaluation in teacher evaluations, and that fairness perceptions may help mitigate this bias. Yet neither student gender nor high threat intensified the backlash in this pre-service teacher sample. Study 2 extends these findings by examining whether in-service teachers, who possess more classroom experience and prolonged exposure to institutional norms and inclusion practices, show similar or divergent patterns of bias and moderation.

Study 2

The goal of Study 2 was to replicate the findings from Study 1 using a sample of in-service teachers to examine whether the effects observed among pre-service teachers generalize to

experienced educators. The methodology, hypotheses, and experimental design were identical to Study 1, with the only difference being the participant sample.

Method

Participants and Procedure

Participants. Data were collected from in-service teachers across France via LimeSurvey in autumn 2024. To ensure geographic diversity and account for anticipated low response rates, 18,000 preschool and primary school principals were randomly selected and invited to participate if they also taught or to forward the invitation to their staff. Participants received the same study information as in Study 1. Participation was voluntary and uncompensated. As the number of teachers who received the invitation is unknown, a precise response rate could not be determined.

A preregistered power analysis using G*Power 3.1.9.4 determined that 264 participants were needed to detect a small effect ($\eta_p^2 = 0.040$; based on Stanczak, Aelenei, et al., 2024) with 80% power at $\alpha = .05$. The final sample ($N = 387$, post-imputation) exceeded this threshold, and a sensitivity power analysis confirmed sufficient power for small effects ($\eta_p^2 = 0.028$ for main/interaction effects; $\eta_p^2 = 0.020$ for moderation models). Table 1 presents demographic details.

Procedure. The 2 × 2 × 2 mixed-method design, materials, manipulations, and measures were identical to Study 1, with SEN status (students without vs. with SEN) as a within-subject factor and student gender (male vs. female) and threat level (low vs. high) as between-subject factors. Participants graded two math tests, evaluated each student's competence, effort, and warmth, and judged the fairness of accommodations.

Ethical approval, randomization procedures, and all study protocols mirrored Study 1 to ensure consistency in replication.

Measures

All measures were identical to Study 1. Descriptive statistics for grading and perceived competence (both αs = .86) are presented in Table 2. Perceived fairness' item concerning the half-exercise accommodation (*M* = 3.93, *SD* = 1.07, Range = 1–5) was mean-centered prior to analysis.

Results

Analysis Strategy

Analyses followed the same procedure as in Study 1. Multiple imputation was performed using mice, and results were pooled using mitools. CFA revalidated the Social Judgement Scales, yielding similar model fit and factor loadings as in Study 1 (see R code on OSF). Primary analyses tested main effects (H1), interaction effects (H2), and fairness as a moderator (H3) via multiple regression. Exploratory analyses examined fairness perceptions as an independent predictor of grading and competence ratings, separately for students with and without SEN. Model assumptions were checked as in Study 1, and sensitivity analyses were conducted for robustness.

Primary Analyses

Grading (Math Tests). Results replicated the backlash effect: students with SEN received lower grading ratings than their non-SEN peers (*b* = 0.63, 95% CI [0.47, 0.79], *SE* = 0.08, *t*(380) = 7.75, *p* < .001, η_p^2 = 0.136). As in Study 1, this effect was unaffected by student gender, threat level, or their interaction (all *p*s > .225).

Perceived Competence. Similarly, students with SEN were judged as less competent (*b* = 0.19, 95% CI [0.14, 0.24], *SE* = 0.03, *t*(379) = 7.10, *p* < .001, η_p^2 = 0.117), with no significant effects of student gender, threat level, or their interaction (all *p*s > .255).

Findings closely mirrored Study 1 (Table 3). Controlling for participants' gender or self-identification as part of a marginalized group did not alter findings (Table S2 in SM).

Moderation Analyses

Fairness of Accommodations. Higher fairness perceptions predicted again smaller grading differences (b = -0.16, 95% CI [-0.32, -0.01], SE = 0.08, $t(353) = -2.10$, $p = .036$, $\eta_p^2 = 0.012$), but not perceived competence differences ($p = .073$), although the descriptive pattern was consistent. As in Study 1, fairness did not interact with student gender, threat level, or their combination (all ps > .073; Table S3 in SM).

Exploratory Analyses

As in Study 1, fairness predicted higher ratings for students with SEN and, to a lesser extent, for non-SEN students. Among students with SEN, greater fairness perceptions were associated with higher grading ratings (b = 0.24, 95% CI [0.08, 0.40], SE = 0.08, $t(358) = 3.02$, $p = .003$, $\eta_p^2 = 0.024$) and greater perceived competence (b = 0.16, 95% CI [0.09, 0.24], SE = 0.04, $t(338) = 4.31$, $p < .001$, $\eta_p^2 = 0.049$). For non-SEN students, fairness was more weakly associated with perceived competence (b = 0.12, 95% CI [0.05, 0.19], SE = 0.04, $t(343) = 3.27$, $p = .001$, $\eta_p^2 = 0.029$) but not with grading ($p = .183$; Table S4 in SM).

Discussion

The primary goal of Study 2 was to replicate Study 1's findings among in-service teachers, extending our inquiry to a group with more classroom experience. As anticipated, students with SEN were again devalued in both grading and competence ratings – a robust backlash effect (H1a) that remained unaffected by student gender, performance-based threat, or their interaction (contrary to H1b, H1c, and H2). Thus, in-service teachers appear no more or less susceptible to these biases than pre-service teachers.

In line with Study 1, fairness perceptions once again emerged as a key moderator – but only for grading: teachers who viewed the "half-exercise" accommodation as fair rated students with SEN more favorably. However, fairness did not interact with student gender or threat, contradicting H3.

Exploratory analyses further showed that fairness perceptions influenced grading ratings for students with SEN but not for their non-SEN peers. While fairness perceptions did not emerge as a key factor in the moderation analysis of competence ratings, exploratory analyses distinguishing SEN and non-SEN competence ratings revealed that fairness had a stronger influence on competence ratings for students with SEN than for their non-SEN peers, mirroring Study 1. This reinforces the idea that fairness concerns shape how teachers evaluate accommodated students, suggesting they may rely on fairness heuristics in both grading and competence judgments.

Taken together, these findings underscore that SEN status – and how fair accommodations are perceived – constitutes the central axis of bias in teachers' judgments. Study 3 further refines our design to directly contrast evaluations of a female student with SEN versus a male student without SEN, clarifying whether and when intersecting stereotypes produce the most pronounced backlash. This final study aims to pinpoint the specific conditions in which teacher bias toward SEN students is amplified or mitigated by gender-related assumptions.

Study 3

Study 3 refined the experimental design and extended the findings from Studies 1 and 2 by eliminating threat manipulations (which yielded no significant effects) and introducing a gender contrast manipulation. While prior studies manipulated student gender by evaluating student pairs of the same gender, Study 3 focused on gender contrast (same-gender vs. cross-gender student pairs) to assess whether backlash effects are amplified when a female student with SEN is directly compared to a male peer without SEN. Male students without SEN are often perceived as the normative standard in academic evaluations. Comparing a female student with SEN to this standard was expected to reinforce both gender stereotypes (linking girls' success with effort rather than competence; Verniers et al., 2016) and SEN-related biases (stereotyping students with

SEN as less competent; Stanczak, Jury, et al., 2024). This dual norm violation was hypothesized to amplify grading and perceived competence disparities.

Four gender contrast conditions were examined:

1. Same-Gender Male (SGM): Boy without SEN vs. boy with SEN
2. Same-Gender Female (SGF): Girl without SEN vs. girl with SEN
3. Cross-Gender Male-Female (CGMF): Boy without SEN vs. girl with SEN
4. Cross-Gender Female-Male (CGFM): Girl without SEN vs. boy with SEN

The preregistered hypotheses were:

Hypothesis 1 (Backlash Effect). The CGMF condition will produce the largest backlash effect, as it represents the most pronounced intersection of gender- and SEN-related biases. This will be reflected in greater differences in grading and perceived competence between students without SEN and students with SEN, relative to all other conditions.

Hypothesis 2 (Moderation Effect). This backlash effect in the CGMF condition will be stronger when fairness perceptions are low.

Method

Participants and Procedure

Participants. Data were collected from in-service teachers across France via LimeSurvey in autumn 2024, following the same recruitment strategy as in Study 2. This time, 21,000 preschool and primary school principals were randomly selected and invited to participate or to forward the invitation to their staff.

Given the anticipated small effects in gender-related biases in educational settings, the preregistered power analysis (G*Power 3.1.9.4) for the planned contrast in H1 targeted effect sizes of η_p^2 = 0.020–0.024 and determined that 316–395 participants were needed for 80% power at α = .05. The final sample (*N* = 364, post-imputation) fell within this range, and a sensitivity power analysis confirmed sufficient power for small effects (η_p^2 = 0.021 for main effects; η_p^2 = 0.029 for moderation models). Table 1 presents demographic details.

Procedure. A 2 × 2 × 2 mixed-method experimental design was used, with SEN status (students without vs. with SEN) as a within-subject factor and gender contrast (same-gender vs. cross-gender) and student gender (male vs. female) as between-subject factors. Study materials and measures were identical to Studies 1 and 2, except that the threat manipulation was removed, and only the low-threat condition was retained (i.e., same error rate on easy and difficult items). Gender contrast was introduced to examine whether backlash effects are amplified when a female student with SEN is compared to a male student without SEN. Student names were counterbalanced across conditions. To heighten gender salience, we incorporated more gendered pronouns in the instructions and reframed the accommodation description to emphasize student agency (e.g., "the student completed only the first two exercises" rather than "the teacher instructed the student to complete only the first two exercises").

Participants graded two math tests, evaluated competence, effort, and warmth, rated accommodation fairness, and provided demographic information. Ethical approval and randomization procedures mirrored prior studies.

Measures

Measures were identical to Studies 1 and 2. Descriptive statistics for grading and perceived competence ($\alpha_{\text{without SEN}}$ = .84; $\alpha_{\text{with SEN}}$ = .86) are shown in Table 2. Fairness ratings for the half-exercise accommodation (*M* = 3.95, *SD* = 1.10, Range = 1–5) were mean-centered.

Results

Analysis Strategy

The statistical procedures mirrored Studies 1 and 2: missing data were imputed (mice), results pooled (mitools), and CFA reconfirmed factor structures. To test the backlash effect (H1), planned orthogonal contrasts were specified to compare experimental conditions (Brauer & McClelland, 2005). The primary contrast (Contrast 1) compared the CGMF condition (boy without SEN vs. girl with SEN), which was hypothesized to elicit the strongest backlash, to the three other conditions (CGFM, SGM, SGF), using contrast coding: CGMF = +3; all others = –1. Two additional orthogonal contrasts partitioned the remaining variance: Contrast 2 contrasted the CGFM condition (girl without SEN vs. boy with SEN; coded +2) against the same-gender conditions (SGM and SGF = –1; CGMF = 0); Contrast 3 contrasted the two same-gender pairings (SGM = +1; SGF = –1; CGMF and CGFM = 0). To test the moderation hypothesis (H2), perceived fairness of accommodations was examined as a moderator of the primary backlash contrast. Interaction terms were created by multiplying each contrast-coded predictor with fairness perceptions, and separate models were estimated for grading and competence difference scores. Finally, exploratory analyses assessed fairness perceptions as independent predictors of grading and competence ratings, separately for students with and without SEN, using the same contrast specifications. Model assumptions were checked as in previous studies, with sensitivity analyses confirming robustness.

Primary Analyses

Grading (Math Tests). Results confirmed the backlash effect: students with SEN received lower grading ratings than their non-SEN peers (b = 0.65, 95% CI [0.48, 0.82], SE = 0.09, $t(358)$ = 7.54, $p < .001$, $\eta_p^2 = 0.136$). However, Contrast 1 (CGMF vs. other conditions) was not significant (p = .988), indicating that the grading gap did not vary by condition.

Perceived Competence. Similarly, students with SEN were rated as less competent (*b* = 0.26, 95% CI [0.20, 0.32], *SE* = 0.03, *t*(352) = 8.62, *p* < .001, η_p^2 = 0.173). Again, Contrast 1 (CGMF vs. other conditions) was not significant (*p* = .107), suggesting that competence judgments were similarly unaffected by gender contrast manipulations.

See Table 3 for an overview of primary regression results. Controlling for participants' gender or marginalized group status did not alter findings (Table S2 in SM).

Moderation Analyses

Fairness of Accommodations. Fairness perceptions were not significantly associated with overall grading differences (*p* = .133) or competence differences (*p* = .110). No significant interactions with Contrast 1 emerged for either grading (*p* = .907) or competence (*p* = .319; Table S3 in SM).[6]

Exploratory Analyses

Fairness perceptions significantly predicted grading and competence ratings for students with SEN and, to a lesser extent, competence ratings for students without SEN. Among students with SEN, higher fairness perceptions were associated with higher grading ratings (*b* = 0.18, 95% CI [0.02, 0.33], *SE* = 0.08, *t*(342) = 2.28, *p* = .023, η_p^2 = 0.015) and increased perceived competence (*b* = 0.13, 95% CI [0.06, 0.20], *SE* = 0.04, *t*(346) = 3.50, *p* = .001, η_p^2 = 0.034). For students without SEN, fairness perceptions were significantly related to perceived competence (*b* = 0.08, 95% CI [0.02, 0.15], *SE* = 0.03, *t*(340) = 2.48, *p* = .014, η_p^2 = 0.018), but not grading (*p* = .315; Table S4 in SM).

[6] A significant interaction emerged for a residual contrast – Contrast 2 (CGFM vs. same-gender conditions) – and fairness on grading (*b* = 0.14, 95% CI [0.01, 0.27], *SE* = 0.07, *t*(354) = 2.12, *p* = .035, ηp^2 = 0.012). When fairness perceptions were low, boys with SEN received lower relative grades than their non-SEN peers; when fairness perceptions were high, this gap was reduced. This contrast was included to partition residual variance; no hypothesis was stated, and the result is reported for completeness only.

Discussion

Study 3 refined our experimental design by removing threat manipulations – shown to have no effect in Studies 1 and 2 – and introducing gender contrast (same-gender vs. cross-gender). We hypothesized that a cross-gender male–female (CGMF) pairing (boy without SEN vs. girl with SEN) would elicit the strongest backlash (H1), and that fairness perceptions would amplify or mitigate this effect (H2).

Consistent with Studies 1 and 2, students with SEN were again devalued in grading and competence, confirming a robust backlash effect that persisted across all conditions. Contrary to H1, however, the CGMF condition did not elicit a stronger backlash than the other pairings (boy–boy, girl–girl, girl–boy). That is, cross-gender comparisons did not intensify the penalty for students with SEN. Regarding H2, fairness perceptions did not moderate this effect as expected.

Exploratory analyses indicated that fairness perceptions more strongly influenced evaluations of students with SEN than those without SEN, echoing the pattern observed in Studies 1 and 2. When teachers viewed the “half-exercise” accommodation as fair, they assigned higher grades and competence ratings to students with SEN.

Overall, these findings reinforce that SEN status alone triggers systematic devaluation across various gender pairings. Fairness perceptions remain a critical factor in shaping teachers’ evaluations, potentially mitigating bias when accommodations are perceived as legitimate. However, the expected heightened backlash in cross-gender comparisons did not emerge, suggesting that gender does not intensify bias against students with SEN in teacher judgments.

Table 2

Descriptive Statistics and t-Tests for Grading and Competence by SEN Status (Study 1, Study 2, and Study 3)

	Without SEN		With SEN		Test Statistics		
Variables	*M (SD)*	Range	*M (SD)*	Range	*t (df)*	*p*	*d* [95% CI]
Study 1							
Grading	6.21 (1.12)	2-10	5.62 (1.68)	1-9	7.15 (462)	<.001***	0.33 [0.24, 0.43]
Perceived Competence	3.67 (0.70)	1-5	3.46 (0.76)	1-5	7.79 (462)	<.001***	0.36 [0.27, 0.46]
Study 2							
Grading	6.22 (1.21)	2-10	5.58 (1.66)	1-10	7.89 (386)	<.001***	0.40 [0.30, 0.51]
Perceived Competence	3.45 (0.75)	1-5	3.25 (0.79)	1-5	7.26 (386)	<.001***	0.37 [0.27, 0.47]
Study 3							
Grading	6.34 (1.17)	2-9	5.69 (1.59)	1-9	7.58 (363)	<.001***	0.40 [0.29, 0.50]
Perceived Competence	3.58 (0.70)	1-5	3.32 (0.77)	1-5	8.58 (362)	<.001***	0.45 [0.34, 0.56]

Note. Means (*M*) and standard deviations (*SD*) were pooled across 20 imputed datasets using Rubin's Rules (Rubin, 1987). Paired *t*-tests examined differences by SEN status, with degrees of freedom (*df*) estimated using the Barnard & Rubin (1999) small-sample adjustment. Repeated-measures Cohen's *d* was computed as the mean difference divided by its standard deviation, with confidence intervals (CI) reflecting both within- and between-imputation variance. *SD*s were approximated as the average sample *SD* across imputations (Enders, 2022). Statistical differences are highlighted as follows: ***$p < .001$.

Table 3

Summary of Primary Regression Analyses for Grading and Competence (Study 1, Study 2, and Study 3)

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Study 1												
Intercept	0.60	0.08	[0.44, 0.77]	7.20 (456)	<.001***	0.102	0.22	0.03	[0.16, 0.27]	7.69 (452)	<.001***	0.115
Threat Level	0.06	0.17	[-0.27, 0.38]	0.34 (456)	0.736	0.000	-0.02	0.06	[-0.13, 0.09]	-0.28 (454)	0.783	0.000
Student Gender	-0.17	0.17	[-0.49, 0.16]	-1.00 (456)	0.320	0.002	0.05	0.06	[-0.05, 0.16]	0.98 (454)	0.328	0.002
Threat Level × Student Gender	-0.23	0.33	[-0.89, 0.42]	-0.70 (456)	0.485	0.001	0.07	0.11	[-0.15, 0.29]	0.63 (455)	0.529	0.001
Study 2												
Intercept	0.63	0.08	[0.47, 0.79]	7.75 (380)	<.001***	0.136	0.19	0.03	[0.14, 0.24]	7.10 (379)	<.001***	0.117
Threat Level	-0.04	0.16	[-0.36, 0.28]	-0.26 (380)	0.797	0.000	-0.02	0.05	[-0.13, 0.08]	-0.44 (373)	0.659	0.001
Student Gender	0.15	0.16	[-0.17, 0.47]	0.91 (380)	0.365	0.002	0.05	0.05	[-0.06, 0.15]	0.91 (372)	0.363	0.002
Threat Level × Student Gender	0.40	0.33	[-0.24, 1.04]	1.22 (380)	0.225	0.004	0.12	0.11	[-0.09, 0.33]	1.14 (375)	0.255	0.003
Study 3												
Intercept	0.65	0.09	[0.48, 0.82]	7.54 (358)	<.001***	0.136	0.26	0.03	[0.20, 0.32]	8.62 (352)	<.001***	0.173
Contrast 1 (CGMF vs. others)	0.00	0.05	[-0.10, 0.09]	-0.01 (358)	0.988	0.000	-0.03	0.02	[-0.06, 0.01]	-1.61 (349)	0.107	0.007
Contrast 2 (CGFM vs. others)	0.04	0.07	[-0.10, 0.18]	0.57 (358)	0.571	0.001	0.01	0.02	[-0.04, 0.06]	0.39 (342)	0.696	0.001
Contrast 3 (SGM vs. others)	0.03	0.12	[-0.22, 0.27]	0.21 (358)	0.831	0.000	0.00	0.04	[-0.09, 0.08]	-0.05 (348)	0.960	0.000

Note. Dependent variables reflect within-subject difference scores (ratings for the student without SEN minus ratings for the student with SEN). *b* = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = *t*-value; *df* = degrees of freedom; *p* = *p*-value; η^2_p = partial eta-squared. Threat Level (-0.5 = low threat condition, 0.5 = high threat condition); Student Gender (-0.5 = boy context, 0.5 = girl context). Contrast 1 (CGMF vs. others: CGMF = +3, CGFM = -1, SGM = -1, SGF = -1); Contrast 2 (CGFM vs. others: CGMF = 0, CGFM = +2, SGM = -1, SGF = -1); Contrast 3 (SGM vs. others: CGMF = 0, CGFM = 0, SGM = +1, SGF = -1). Experimental conditions: CGMF = Cross-Gender Male-Female (boy without SEN vs. girl with SEN); CGFM = Cross-Gender Female-Male (girl without SEN vs. boy with SEN); SGM = Same-Gender Male (boy without SEN vs. boy with SEN); SGF = Same-Gender Female (girl without SEN vs. girl with SEN). Statistical differences are highlighted as follows: ***$p < .001$.

General Discussion

We conducted three experimental studies to examine whether teachers systematically devalue students with SEN who are provided with accommodations, whether student gender and performance level amplify this bias, and whether fairness perceptions moderate teachers' judgments.

Despite methodological variations, all three studies revealed a consistent backlash effect: teachers rated students with SEN lower in grading and competence than their non-SEN peers, regardless of whether participants were pre-service (Study 1) or in-service teachers (Studies 2 and 3). This underscored its resilience across levels of teaching experience. Contrary to our predictions, neither student gender (Studies 1–3) nor performance-based threat (Studies 1 and 2) intensified this bias. Even when directly contrasting female students with SEN to male non-SEN students (Study 3), devaluation remained stable. A consistent theme across all three studies was fairness perceptions regarding accommodations: teachers who viewed the "half-exercise" as fair penalized students with SEN less, thereby mitigating – but never eliminating – the backlash effect.

Taken together, these findings demonstrate that SEN status with accommodations drives teacher evaluations, while gender and performance-based threat play minor roles. At the same time, fairness perceptions consistently moderate SEN evaluations, indicating that teachers' acceptance (or rejection) of accommodations can soften the penalty otherwise directed at students with SEN.

Theoretical Implications

Our findings provide evidence that teachers systematically devalue the achievements of students with SEN when accommodations are perceived as unfair, consistent with a backlash effect (Rudman et al., 2012; Stanczak, Aelenei, et al., 2024). Specifically, a reduced workload accommodation (half-exercise) probably led teachers to attribute success more to external help

than to the student's own ability. This aligns with arguments that accommodations may threaten teachers' meritocratic ideals by appearing to grant undeserved advantages, prompting a psychological correction that undermines recognition of students' actual competence (Brueggemann et al., 2001; Stanczak, Jury, et al., 2024).

These findings further contribute to discussions on how fairness perceptions intersect with meritocratic ideology in education. While meritocracy posits that success reflects effort and talent, our findings indicate that teachers' immediate fairness judgments about accommodations play a more direct role in shaping evaluations. This aligns with justice-based frameworks, which propose that individuals react negatively to perceived imbalances between inputs (e.g., effort) and outcomes (e.g., success; Deutsch, 1975; Rudman et al., 2012). In our context, the accommodation appeared to lower "input" while yielding similar "output", making the student's success appear less earned (Rudman et al., 2012; Stanczak, Aelenei, et al., 2024). Thus, even teachers who support inclusive education may penalize students with SEN if they feel an accommodation violates their standard of equity. By highlighting the influence of situational fairness judgments, our findings refine prior arguments that meritocratic ideals can become ideological barriers to inclusion (Darnon, Smeding, et al., 2018; Stanczak, Jury, et al., 2024). Although we preregistered a moderating role of meritocratic beliefs, these were not consistent predictors across studies (see SM). These results suggest that while meritocratic ideology may form the broader cultural framework, backlash effects are more immediately shaped by how justifiable a given accommodation appears within the classroom context.

Another key implication is the consistent absence of gender differences in backlash patterns. Contrary to expectations that students with SEN might face compounded bias based on gender, we found no evidence that teachers evaluated female and male students with SEN differently in grading or competence. This aligns with research suggesting that biases against

individuals with disabilities often operate independently of gender, reflecting broader ableist patterns rather than gendered stereotypes (Wang et al., 2019). In our study, negative evaluations were driven not by student gender but by teachers' fairness perceptions of accommodations. This suggests that backlash stems primarily from perceived threats to "meritocratic" fairness rather than from gender stereotypes. Thus, our findings refine theory by showing that biases against students with SEN can generalize across gender lines, emphasizing fairness perceptions as the core mechanism behind SEN-related backlash.

Generalizability, Limitations, and Future Directions

While our study provides important insights, several limitations affect generalizability and suggest avenues for future research. One concerns our participant sample, which included pre-service and in-service teachers from a single national context (France), with in-service participants limited to primary-level educators. Given that meritocratic competition intensifies in secondary education (Stanczak, Aelenei, et al., 2024), our findings may not generalize to high school teachers or educational systems with greater academic competition. Future research should examine whether bias toward accommodated students differs in secondary and post-secondary settings or across national education systems.

Another limitation is our vignette methodology, which – while ensuring tight control – does not fully replicate real classroom interactions. In practice, teachers develop long-term perceptions of students, which may attenuate or reinforce bias. For instance, knowing a student with SEN's struggles might increase sympathy or entrench pre-existing biases. Because our vignette was a one-off snapshot, these results may not reflect teachers' day-to-day behavior. To improve ecological validity, future research should employ longitudinal or naturalistic designs to assess whether bias fades with familiarity or persists over time.

The specificity of the SEN diagnosis and accommodation in our study also limits generalizability. We focused on a student labeled with ADHD, a common yet often stereotyped neurodevelopmental disorder. However, SEN encompasses diverse conditions – from learning and intellectual disabilities to physical and sensory impairments – each of which may elicit distinct biases. Research suggests that disability type influences social perceptions and expectations (Rohmer & Louvet, 2011), and that some accommodations (e.g., assistive technology) are perceived as fairer than workload reductions (Stanczak, Aelenei, et al., 2024). In our study, a reduced workload accommodation triggered bias, but it remains unclear whether this extends to other accommodations or SEN profiles. Future research should examine a broader range of SEN categories and support measures.

We must also consider potential self-selection biases. Participation was voluntary, meaning our sample may overrepresent educators supportive of inclusion, while more skeptical teachers may have opted out – potentially underestimating the prevalence of bias. Still, the emergence of a backlash effect in this relatively inclusion-supportive sample suggests such bias may be even more widespread. That said, we used multiple imputation to mitigate missing data, though this cannot fully correct for nonrandom dropout (Enders, 2022). Furthermore, reliance on self-reported judgments introduces social desirability bias – participants know that overtly unfair treatment of a student with SEN is undesirable. Future research should incorporate implicit measures or behavioral indicators, such as whether teachers assign fewer challenges or offer less encouragement to accommodated students who excel.

Implications for Social Policy

Mandating accommodations alone is insufficient for true inclusion – teachers' fairness perceptions shape how accommodations affect students with SEN. Policies should not only ensure accommodations are available but also address how teachers understand and apply them.

Our findings show that teachers who perceive accommodations as fair penalize students with SEN less. Teacher training should explicitly clarify the purpose of accommodations and address concerns about fairness. Workshops using real cases can demonstrate how accommodations fit within standardized evaluation. Reframing accommodations as parallel to commonplace adjustments (e.g., eyeglasses or hearing aids) can help shift perspectives. Empirical evidence – such as findings that extra time does not inflate grades but enables students to demonstrate actual knowledge (Sireci et al., 2005; Vidal Rodeiro & Macinska, 2022) – can reinforce this understanding.

Fairness perceptions also depend on how accommodations are explained. Schools should establish transparent communication protocols clarifying why accommodations exist and how they ensure equal opportunity. Framing them as necessary corrections for structural barriers rather than as special advantages, may decrease skepticism. Standardized messaging at the institutional level can ensure consistency in how accommodations are framed.

Bias – even subtle – can shape student outcomes. Schools should create structured opportunities for teachers to reflect on fairness dilemmas (e.g., “Is it fair to give an easier exam version to a student with SEN?”). Without such discussions, implicit biases may continue to influence decision-making. At a policy level, embedding fairness discussions and bias training into teacher certification and evaluation standards can help ensure inclusive practices are meaningfully implemented.

Conclusion

This study contributes to our understanding of backlash effects in inclusive education. Across three studies, teachers devalued the grades and competence of an accommodated student with SEN, revealing a subtle but systematic bias. This backlash occurred regardless of student gender, suggesting it applies to all students with SEN. Importantly, teachers' fairness perceptions moderated this bias: when accommodations were perceived as fair, devaluation weakened. These findings expand meritocracy-based theories by showing how accommodated success may trigger subtle forms of resistance in evaluators. By identifying when and why this bias occurs, our study offers practical insights for teacher training and education policy. Addressing this may require framing accommodations not as undue advantages, but as equity tools that help students with SEN demonstrate their competence. Ultimately, shifting teacher perceptions may be key to ensuring that students with SEN are evaluated fairly, recognized for their achievements, and truly included in education.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the main author used OpenAI's ChatGPT to assist with statistical code checking in R and for proofreading grammar and spelling. After using this tool, the author reviewed and edited all content as needed and takes full responsibility for the content of the published article.

References

Ainscow, M., Slee, R., & Best, M. (2019). Editorial: The Salamanca Statement: 25 years on. *International Journal of Inclusive Education*, *23*(7–8), 671–676. https://doi.org/10.1080/13603116.2019.1622800

Amor, A. M., Hagiwara, M., Shogren, K. A., Thompson, J. R., Verdugo, M. Á., Burke, K. M., & Aguayo, V. (2019). International perspectives and trends in research on inclusive education: A systematic review. *International Journal of Inclusive Education*, *23*(12), 1277–1295. https://doi.org/10.1080/13603116.2018.1445304

Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*(4), 948–955. https://www.jstor.org/stable/2673599

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., Jetten, J., Van De Werfhorst, H., Darnon, C., & Butera, F. (2023). Belief in school meritocracy and the legitimization of social and income inequality. *Social Psychological and Personality Science*, *14*(5), 621–635. https://doi.org/10.1177/19485506221111017

Benjamin, S. (2002). *The micropolitics of inclusive education: An ethnography*. Open Univ. Press.

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391. https://doi.org/10.1126/science.aah6524

Brauer, M., & McClelland, G. (2005). L'utilisation des contrastes dans l'analyse des données: Comment tester les hypothèses spécifiques dans la recherche en psychologie ? [The use of contrasts ind ata analyses: How to test specific hypotheses in psychological research]. *L'année psychologique*, *105*(2), 273–305. https://doi.org/10.3406/psy.2005.29696

Brueggemann, B. J., White, L. F., Dunn, P. A., Heifferon, B. A., & Cheu, J. (2001). Becoming visible: Lessons in disability. *College Composition and Communication*, *52*(3), 368. https://doi.org/10.2307/358624

Brussino, O. (2020). *Mapping policy approaches and practices for the inclusion of students with special education needs* (OECD Education Working Papers 227). https://doi.org/10.1787/600fbad5-en

Butera, F., Świątkowski, W., & Dompnier, B. (2024). Competition in education. In S. M. Garcia, A. Tor, & A. J. Elliot (Eds.), *The Oxford Handbook of the Psychology of Competition* (1st ed., pp. 569–597). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190060800.013.24

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

Clément-Guillotin, C., Rohmer, O., Forestier, C., Guillotin, P., Deshayes, M., & d'Arripe-Longueville, F. (2018). Implicit and explicit stereotype content associated with people with physical disability: Does sport change anything? *Psychology of Sport and Exercise*, *38*, 192–201. https://doi.org/10.1016/j.psychsport.2018.06.014

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Routledge. https://doi.org/10.4324/9780203771587

Cohen, J., Schiffler, F., Rohmer, O., Louvet, E., & Mollaret, P. (2019). Is disability really an obstacle to success? Impact of a disability simulation on motivation and performance. *Journal of Applied Social Psychology*, *49*(1), 50–59. https://doi.org/10.1111/jasp.12564

Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human*

Resources, *48*(1), 236–264. https://doi.org/10.3368/jhr.48.1.236

Darnon, C., Smeding, A., & Redersdorff, S. (2018). Belief in school meritocracy as an ideological barrier to the promotion of equality. *European Journal of Social Psychology*, *48*(4), 523–534. https://doi.org/10.1002/ejsp.2347

Darnon, C., Wiederkehr, V., Dompnier, B., & Martinot, D. (2018). 'Where there is a will, there is a way': Belief in school meritocracy and the social-class achievement gap. *British Journal of Social Psychology*, *57*(1), 250–262. https://doi.org/10.1111/bjso.12214

De Beco, G. (2018). The right to inclusive education: Why is there so much opposition to its implementation? *International Journal of Law in Context*, *14*(3), 396–415. https://doi.org/10.1017/S1744552317000532

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, *31*(3), 137–149. https://doi.org/10.1111/j.1540-4560.1975.tb01000.x

Dunn, D. S., & Andrews, E. E. (2015). Person-first and identity-first language: Developing psychologists' cultural competence using disability language. *American Psychologist*, *70*(3), 255–264. https://doi.org/10.1037/a0038636

Enders, C. K. (2022). *Applied missing data analysis* (2nd ed., pp. ix, 546). The Guilford Press.

European Commission. Directorate General for Employment, Social Affairs and Inclusion. (2018). *Access to quality education for children with special educational needs.* Publications Office. https://data.europa.eu/doi/10.2767/440746

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Florian, L. (2014). What counts as evidence of inclusive education? *European Journal of Special*

Needs Education, *29*(3), 286–294. https://doi.org/10.1080/08856257.2014.933551

Gernsbacher, M. A., Raimond, A. R., Balinghasay, M. T., & Boston, J. S. (2016). "Special needs" is an ineffective euphemism. *Cognitive Research: Principles and Implications*, *1*(1), 29. https://doi.org/10.1186/s41235-016-0025-4

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, *17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Hafen, C. A., Ruzek, E. A., Gregory, A., Allen, J. P., & Mikami, A. Y. (2015). Focusing on teacher–student interactions eliminates the negative impact of students' disruptive behavior on teacher perceptions. *International Journal of Behavioral Development*, *39*(5), 426–431. https://doi.org/10.1177/0165025415579455

Hibel, J., Farkas, G., & Morgan, P. L. (2010). Who is placed into special education? *Sociology of Education*, *83*(4), 312–332. https://doi.org/10.1177/0038040710383518

INSEE (Institut national de la statistique et des études économiques). (2024). *Classement des prénoms en France depuis 1900 [Ranking of first names in France since 1900]*. https://www.insee.fr/en/statistiques/6536067

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, *75*(4), 649–674. https://doi.org/10.1007/s11336-010-9175-3

Jenifer, J. B., Jaxon, J., Levine, S. C., & Cimpian, A. (2024). "You need to be super smart to do well in math!" Young children's field-specific ability beliefs. *Developmental Science*, *27*(1), e13429. https://doi.org/10.1111/desc.13429

Jost, J. T. (2019). A quarter century of system justification theory: Questions, answers, criticisms, and societal applications. *British Journal of Social Psychology*, *58*(2), 263–314.

https://doi.org/10.1111/bjso.12297

Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, *38*(6), 586–602. https://doi.org/10.1016/S0022-1031(02)00505-X

Jury, M., Perrin, A.-L., Rohmer, O., & Desombre, C. (2021). Attitudes toward inclusive education: An exploration of the interaction between teachers' status and students' type of disability within the French context. *Frontiers in Education*, *6*, 655356. https://doi.org/10.3389/feduc.2021.655356

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3

Kefallinou, A., Symeonidou, S., & Meijer, C. J. W. (2020). Understanding the value of inclusive education and its implementation: A review of the literature. *PROSPECTS*, *49*(3–4), 135–152. https://doi.org/10.1007/s11125-020-09500-2

Khamzina, K., Jury, M., Ducreux, E., & Desombre, C. (2021). The conflict between inclusive education and the selection function of schools in the minds of French teachers. *Teaching and Teacher Education*, *106*, 103454. https://doi.org/10.1016/j.tate.2021.103454

Krämer, S., & Zimmermann, F. (2023). Students with emotional and behavioral disorder and teachers' stereotypes – Effects on teacher judgments. *The Journal of Experimental Education*, *91*(3), 450–471. https://doi.org/10.1080/00220973.2021.1934809

Krischler, M., & Pit-ten Cate, I. M. (2019). Pre- and in-service teachers' attitudes toward students with learning difficulties and challenging behavior. *Frontiers in Psychology*, *10*, 327. https://doi.org/10.3389/fpsyg.2019.00327

Krischler, M., & Pit-ten Cate, I. M. (2020). Inclusive education in Luxembourg: Implicit and explicit

attitudes toward inclusion and students with special educational needs. *International Journal of Inclusive Education*, *24*(6), 597–615. https://doi.org/10.1080/13603116.2018.1474954

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. https://doi.org/10.1126/science.1261375

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Louvet, E., & Rohmer, O. (2016). Évaluation des personnes en situation de handicap en milieu éducatif et professionnel: Approche expérimentale [Evaluation of people with disabilities in educational and professional settings: An experimental approach]. *La nouvelle revue de l'adaptation et de la scolarisation*, *74*(2), 159. https://doi.org/10.3917/nras.074.0159

Lumley, T. (2019). *mitools: Tools for multiple imputation of missing data* (p. 2.4) [Dataset]. https://doi.org/10.32614/CRAN.package.mitools

Martin, J. (2024). Why are females less likely to be diagnosed with ADHD in childhood than males? *The Lancet Psychiatry*, *11*(4), 303–310. https://doi.org/10.1016/S2215-0366(24)00010-5

Mijs, J. J. B. (2016). The unfulfillable promise of meritocracy: Three lessons and their implications for justice in education. *Social Justice Research*, *29*(1), 14–34. https://doi.org/10.1007/s11211-014-0228-0

Ministère de l'Éducation Nationale et de la Jeunesse. (2023). *Évaluations 2023 repères CM1—Premiers résultats [2023 CM1 benchmark assessments—Initial results]*. Ministère de l'Education Nationale et de la Jeunesse. https://www.education.gouv.fr/evaluations-2023-reperes-cm1-premiers-resultats-379866

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019: International results in mathematics and science*. TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/international-results/

Nilholm, C., & Göransson, K. (2017). What is meant by inclusion? An analysis of European and North American journal articles with high impact. *European Journal of Special Needs Education*, *32*(3), 437–451. https://doi.org/10.1080/08856257.2017.1295638

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

OECD. (2005). *Students with disabilities, learning difficulties and disadvantages: Statistics and indicators*. OECD. https://doi.org/10.1787/9789264009813-en

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD. https://doi.org/10.1787/53f23881-en

Rohmer, O., & Louvet, E. (2011). Le stéréotype des personnes handicapées en fonction de la nature de la déficience: Une application des modèles de la bi-dimensionnalité du jugement social [Stereotypes of individuals with disabilities based on the nature of their impairment: An application of bi-dimensional models of social judgment]. *L'Année psychologique*, *111*(1), 69–85. https://doi.org/10.3917/anpsy.111.0069

Rosseel, Y. (2012). lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (1st ed.). John Wiley & Sons.

https://doi.org/10.1002/9780470316696

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, *87*(2), 157–176. https://doi.org/10.1037/0022-3514.87.2.157

Rudman, L. A., Moss-Racusin, C. A., Glick, P., & Phelan, J. E. (2012). Reactions to vanguards. In *Advances in Experimental Social Psychology* (Vol. 45, pp. 167–227). Elsevier. https://doi.org/10.1016/B978-0-12-394286-9.00004-4

Schaeffer, K. (2023, July 24). What federal education data shows about students with disabilities in the U.S. *Pew Research Center*. https://www.pewresearch.org/short-reads/2023/07/24/what-federal-education-data-shows-about-students-with-disabilities-in-the-us/

Shifrer, D. (2013). Stigma of a label: Educational expectations for high school students labeled with learning disabilities. *Journal of Health and Social Behavior*, *54*(4), 462–480. https://doi.org/10.1177/0022146513503346

Shifrer, D. (2016). Stigma and stratification limiting the math course progression of adolescents labeled with a learning disability. *Learning and Instruction*, *42*, 47–57. https://doi.org/10.1016/j.learninstruc.2015.12.001

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, *75*(4), 457–490. https://doi.org/10.3102/00346543075004457

Stanczak, A., Aelenei, C., Pironom, J., Toczek-Capelle, M.-C., Rohmer, O., & Jury, M. (2024). Can students with special educational needs overcome the “success” expectations? *Social Psychology of Education*. https://doi.org/10.1007/s11218-023-09806-x

Stanczak, A., Jury, M., Aelenei, C., Pironom, J., Toczek-Capelle, M.-C., & Rohmer, O. (2024). Special

education and meritocratic inclusion. *Educational Policy*, *38*(1), 85–103. https://doi.org/10.1177/08959048231153606

Thapar, A., Eyre, O., Patel, V., & Brent, D. (2022). Depression in young people. *The Lancet*, *400*(10352), 617–631. https://doi.org/10.1016/S0140-6736(22)01012-1

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3). https://doi.org/10.18637/jss.v045.i03

Verniers, C., Martinot, D., & Dompnier, B. (2016). The feminization of school hypothesis called into question among junior and high school students. *British Journal of Educational Psychology*, *86*(3), 369–381. https://doi.org/10.1111/bjep.12111

Vidal Rodeiro, C., & Macinska, S. (2022). Equal opportunity or unfair advantage? The impact of test accommodations on performance in high-stakes assessments. *Assessment in Education: Principles, Policy & Practice*, *29*(4), 462–481. https://doi.org/10.1080/0969594X.2022.2121680

Vlachou, A., Eleftheriadou, D., & Metallidou, P. (2014). Do learning difficulties differentiate elementary teachers' attributional patterns for students' academic failure? A comparison between Greek regular and special education teachers. *European Journal of Special Needs Education*, *29*(1), 1–15. https://doi.org/10.1080/08856257.2013.830440

Wang, K., Walker, K., Pietri, E., & Ashburn-Nardo, L. (2019). Consequences of confronting patronizing help for people with disabilities: Do target gender and disability type matter? *Journal of Social Issues*, *75*(3), 904–923. https://doi.org/10.1111/josi.12332

Wiederkehr, V., Bonnot, V., Krauth-Gruber, S., & Darnon, C. (2015). Belief in school meritocracy as a system-justifying tool for low status students. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01053

Supplemental Online Material for:

When Accommodations Are Not Enough:

A Multi-Study Examination of Teacher Bias Toward Students with Special Educational Needs Across Student Gender

Fabian Müller[1], Cristina Aelenei[2], Mickaël Jury[1,3]

[1] Laboratoire ACTé, Université Clermont Auvergne, Clermont-Ferrand, France

[2] Université Paris Cité, Laboratoire de Psychologie Sociale, Boulogne-Billancourt, France

[3] Institut Universitaire de France, Paris, France

Table of Contents

List of Tables

Meritocratic Beliefs

Conceptual Background. Across all three studies, we preregistered that belief in school meritocracy (BSM) would moderate backlash effects – particularly under high-threat conditions for female students (Studies 1 and 2) or in the CGMF condition (Study 3). BSM reflects the conviction that school success depends on talent and effort within a fair system (Mijs, 2016; Wiederkehr et al., 2015). As a system-justifying belief, it can legitimize inequality and lead teachers to resist accommodations as violations of equal rules (Butera et al., 2024; Darnon et al., 2018; Mijs, 2016; Stanczak et al., 2024).

However, across all three studies, immediate fairness perceptions – not broader meritocratic beliefs – were the more consistent predictor of teacher judgments. While BSM may provide a cultural frame, fairness appraisals of the specific accommodation shaped evaluations more directly.

Descriptive and Measurement Details. Participants completed the eight-item BSM scale (Wiederkehr et al., 2015; 1 = not at all, 5 = completely agree). Responses were averaged and mean-centered. Scale descriptives and reliability per study are as follows:

- **Study 1:** M = 2.21, SD = 0.66, Range = 1–5, α = .78
- **Study 2:** M = 2.19, SD = 0.57, Range = 1–4, α = .74
- **Study 3:** M = 2.15, SD = 0.60, Range = 1–4, α = .75

Confirmatory Factor Analysis (Study 1) supported a one-factor structure: robust RMSEA = .082, 90% CI [.063, .103]; robust CFI = .915; robust TLI = 1.000; SRMR = .051. A reverse-coded item had low loading (.313), but removing it worsened the model fit (robust RMSEA increased to .091, 90% CI [.068, .116]). It was therefore retained for theoretical consistency, with standardized loadings ranging from .313 to .686. Factor structures in Studies 2 and 3 were revalidated and showed comparable results (see R code on OSF).

Moderation Analyses. Across all studies, BSM did not significantly moderate the effects of condition on grading or competence:

- **Study 1:** All *p*s > .058
- **Study 2:** All *p*s > .164
- **Study 3:** All *p*s > .096

Full model results are presented in Table S3

Summary of Moderation Analyses for Grading and Competence (Study 1, Study 2, and Study 3).

Exploratory Analyses. To explore whether BSM predicted evaluations for students with and without SEN, we conducted disaggregated analyses:

- **Study 1:** Meritocratic beliefs did not significantly predict grading or competence for either student group (all *p*s > .137).
- **Study 2:** A small negative effect was found for perceived competence in non-SEN students (b = -0.14, 95% CI [-0.27, 0.00], SE = 0.07, $t(354)$ = -2.00, p = .046, η_p^2 = 0.011); no other effects reached significance (all *p*s > .098).
- **Study 3:** Meritocratic beliefs negatively predicted grading for non-SEN students (b = -0.23, 95% CI [-0.44, 0.01], SE = 0.10, $t(290)$ = -2.08, p = .039, η_p^2 = 0.013); no significant effects were found for students with SEN (p = .166), and meritocratic beliefs did not predict competence for either group (*p*s > .408).

These findings suggest that fairness perceptions are a more consistent and proximal predictor of teacher evaluations than generalized ideological beliefs. One possible explanation is that meritocratic beliefs serve as a distal cultural frame, while fairness judgments are activated in the moment and tied to the specific accommodation presented. Full exploratory models are reported in Table S4

Summary of Not Preregistered Exploratory Regression Analyses by SEN Status for Grading and Competence (Study 1, Study 2, and Study 3).

Preregistered Secondary Analyses – Stratified Analyses for SEN and Non-SEN Students

To assess whether the observed effects were primarily driven by how teachers evaluated students with SEN, we conducted preregistered secondary analyses that stratified regression models by SEN status. Full results are presented in Table S5.

- **Study 1:** Results revealed distinct patterns: Among students without SEN, competence ratings were significantly higher for female students than for male students (b = 0.16, 95% CI [0.03, 0.29], SE = 0.07, $t(454) = 2.46$, p = .014, η_p^2 = 0.013), while grading was unaffected (all ps > .156). For students with SEN, female students received more favorable grading (b = 0.32, 95% CI [0.01, 0.62], SE = 0.16, $t(456) = 2.01$, p = .045, η_p^2 = 0.009), but competence ratings were not significantly affected (all ps > .069). These findings reinforce that the observed backlash primarily penalizes students with SEN, rather than reflecting elevated ratings for non-SEN peers. Although a small gender effect on competence appeared among non-SEN students, its size (η_p^2 = 0.013) was minor relative to the larger backlash effect observed in the primary analysis for all SEN students (η_p^2 = 0.115; see main text). Similarly, the grading advantage for female students with SEN did not extend to competence ratings, suggesting that teachers may slightly reward them in grading without perceiving them as more competent than their male peers.
- **Study 2:** No significant effects were found in either group (all ps > .136), confirming that the backlash effect primarily reflects penalizing SEN students rather than boosting their non-SEN peers, and that no major shifts in non-SEN ratings were evident.
- **Study 3:** For students without SEN, none of the planned contrasts significantly influenced grading or competence ratings (all ps > .060). For students with SEN, competence ratings were significantly influenced by Contrast 1 (b = 0.06, 95% CI [0.02, 0.11], SE = 0.02, $t(354) = 2.77$, p = .006, η_p^2 = 0.021), while grading ratings remained unaffected (ps > .579). These findings align with the previous studies, reinforcing that

backlash effects are specific to students with SEN. Ratings for non-SEN students remained stable across conditions, indicating no general uplift or reevaluation of their performance. The observed shift in competence ratings was limited to students with SEN, suggesting that the backlash reflects a targeted bias, rather than comparative enhancement of their non-SEN peers.

Taken together, these stratified results reinforce the interpretation that the backlash effect reflects a targeted penalty toward students with SEN, rather than a general reevaluation of their non-SEN peers. SEN status – not differential treatment of their non-SEN classmates – consistently emerges as the primary driver of the observed bias.

Preregistered Secondary and Exploratory Analyses – Perceived Effort

Perceived effort was preregistered as a secondary outcome in Studies 1 and 2 and as an exploratory outcome in Study 3, based on prior work highlighting its relevance to competence judgments – particularly in gendered contexts. Across all studies, we examined effort difference scores (ratings for the student without SEN minus ratings for the student with SEN). Full results are reported in Table S6; descriptive statistics and paired *t*-tests comparing effort ratings by SEN status appear in Table S1.

- **Study 1:** No significant effects emerged for the overall effort difference score (*p*s > .123), indicating that students with SEN were not systematically perceived as exerting more or less effort than their non-SEN peers. This pattern held after controlling for participant gender. However, stratified models revealed more nuanced patterns. Among students without SEN, female students were rated as exerting more effort than male peers (*b* = 0.17, 95% CI [0.02, 0.31], *SE* = 0.08, *t*(456) = 2.18, *p* = .030, η_p^2 = 0.010). Among students with SEN, a significant Threat × Student Gender interaction emerged (b = -0.31, 95% CI [-0.62, 0.00], SE = 0.16, t(453) = -1.99, p = .047, ηp2 = 0.009): under high-threat conditions, female students with SEN were rated as exerting less effort than male peers.

Still, because no overall effort difference emerged in the primary analysis, these subgroup effects should be interpreted cautiously.

- **Study 2:** A significant backlash effect was found: students with SEN were rated as exerting less effort than non-SEN peers (b = 0.07, 95% CI [0.03, 0.12], SE = 0.02, t(362) = 3.01, p = .003, η_p^2 = 0.024). This effect became non-significant (p = .070) after adjusting for covariates, though the descriptive pattern remained. Stratified models revealed no significant effects for either group (all ps > .336 for non-SEN, all ps > .542 for SEN), confirming that the backlash reflects lower ratings for SEN students, rather than elevated perceptions of non-SEN peers.
- **Study 3:** In line with preregistered exploratory analyses, we examined effort difference scores using two models:

  Model 1 (contrast-coded) revealed a significant baseline backlash effect (b = 0.057, 95% CI [0.01, 0.10], SE = 0.02, t(328) = 2.42, p = .016, η_p^2 = 0.017), again indicating that students with SEN were seen as exerting less effort. Female students with SEN were rated as exerting more effort (Contrast 1: b = -0.03, 95% CI [-0.06, -0.01], SE = 0.01, t(347) = -2.45, p = .015, η_p^2 = 0.017), while male students with SEN were rated as exerting less effort (Contrast 2: b = 0.05, 95% CI [0.01, 0.09], SE = 0.02, t(350) = 2.53, p = .012, η_p^2 = 0.018) in cross-gender contexts. These effects remained when covariates were added, although the intercept (general backlash) became non-significant (p = .353).

  Model 2 (factorial-coded) confirmed these patterns. Female students with SEN were rated as exerting more effort (b = -0.10, 95% CI [-0.19, -0.01], SE = 0.05, t(346) = -2.07, p = .039, η_p^2 = 0.012). This advantage weakened in cross-gender contexts, as indicated by a significant Student Gender × Gender Contrast interaction (b = -0.26, 95% CI [-0.44, -0.07], SE = 0.09, t(345) = -2.75, p = .006, η_p^2 = 0.021).

  Stratified models of effort ratings supported this interpretation. Among students without SEN, cross-gender contexts boosted effort ratings (b = 0.30, 95% CI [0.12, 0.47],

SE = 0.09, *t*(355) = 3.30, *p* = .001, η_p^2 = 0.030), particularly for female students (interaction: *b* = -0.37, 95% CI [-0.72, -0.01], *SE* = 0.18, *t*(353) = -2.04, *p* = .042, η_p^2 = 0.012). For students with SEN, effort ratings were also slightly higher in cross-gender contexts (*b* = 0.26, 95% CI [0.08, 0.45], *SE* = 0.09, *t*(354) = 2.82, *p* = .005, η_p^2 = 0.022), but no significant interaction with student gender was found (*p* = .566).

Effort perceptions revealed a more complex and less consistent pattern than competence or grading. While Studies 2 and 3 showed some evidence of a backlash – lower perceived effort for students with SEN – Study 1 did not. Gendered and contextual effects further shaped these ratings, particularly in Study 3. Together, these findings suggest that effort judgments are sensitive to context and social cues but do not consistently reflect a generalized bias against students with SEN.

Preregistered Exploratory Analyses – Study 3: Grading and Competence Difference Scores

We conducted preregistered exploratory analyses examining whether gender of the student with SEN, gender contrast, and their interaction predicted difference scores in grading and competence. Full results are reported in Table S7.

The intercepts in this factorial-coded model mirrored those of the contrast-coded model presented in the main text, confirming that students with SEN were rated lower than their non-SEN peers overall (grading: *b* = 0.65, 95% CI [0.48, 0.82], *SE* = 0.09, *t*(358) = 7.54, *p* < .001, η_p^2 = 0.136; competence: *b* = 0.26, 95% CI [0.20, 0.32], *SE* = 0.03, *t*(352) = 8.62, *p* < .001, η_p^2 = 0.173).

No significant main or interaction effects emerged for either outcome:

- Grading: No significant effects of student gender, gender contrast, or their interaction were found (all *p*s > .649).
- Competence: Similarly, neither student gender, gender contrast, nor their interaction significantly influenced competence ratings (all *p*s > .268).

Taken together, these results suggest that the general backlash effect observed in grading and competence ratings was not significantly moderated by the gender of the student with SEN or the gender composition of the comparison. Judgments remained lower for students with SEN across all gender contexts.

Covariate Coding – Participant Gender

Participants reported their gender identity as "female," "male," "self-identify" (e.g., nonbinary, gender non-conforming), or "prefer not to say". For exploratory analyses where participant gender was included as a covariate, responses were contrast-coded: male = -0.5, non-male = 0.5 (including female, gender non-conforming, and "prefer not to say").

This binary coding scheme was selected for both conceptual and statistical reasons: (1) gender identity exists along a spectrum rather than as three discrete categories; (2) creating a separate category for gender non-conforming individuals could inadvertently reinforce othering; (3) the binary structure aligns with the way gender is often socially experienced – as male vs. non-male – in psychological and societal contexts; and (4) small subgroup sizes for gender non-conforming individuals and "prefer not to say" responses posed challenges for multivariate modeling.

Non-Preregistered Exploratory Factor Analysis of a full Fairness Scale – Study 1

An exploratory factor analysis (EFA) was conducted on the non-imputed dataset from Study 1 to examine the dimensionality of a possible fairness of accommodations scale. Results suggested a single-factor structure, accounting for 23% of the variance. However, internal consistency was low ($\alpha = .59$), and model fit indices indicated poor fit: RMSEA = .102, 90% CI [.066, .141]; TLI = .764; $\chi^2(5) = 26.91$, $p < .001$. Factor loadings ranged from .36 to .57, with some items contributing weakly (e.g., "separate room": .36). These findings point to psychometric

limitations in aggregating the five fairness items into a composite score, and support our decision to focus the moderation analysis on the half-exercise item only.

Not Preregistered Exploratory Regression Analysis by SEN Status for Grading and Competence

As mentioned in the main text, we conducted exploratory regressions separately for students with and without SEN to examine whether fairness perceptions interacted with student gender and threat level (Studies 1–2) or gender contrast (Study 3). One significant three-way interaction emerged in Study 1 for grading evaluations of students with SEN (Threat Level × Student Gender × Fairness: $b = 0.56$, 95% CI [0.06, 1.07], $SE = 0.26$, $t(405) = 2.18$, $p < .030$, $\eta_p^2 = 0.011$). This interaction suggested that fairness perceptions were particularly influential when multiple expectancy violations were present – specifically, under high-threat conditions for female students with SEN. In this condition, higher fairness perceptions were most strongly associated with reduced grading bias. No other interactions were significant ($ps > .111$), and this pattern did not replicate in Study 2 (all $ps > .147$) or Study 3 (all $ps > .278$). These analyses were not preregistered and are not central to our theoretical model but are reported here in full for transparency. Full regression results are available in Table S4.

References

Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*(4), 948–955.

Butera, F., Świątkowski, W., & Dompnier, B. (2024). Competition in education. In S. M. Garcia, A. Tor, & A. J. Elliot (Eds.), *The Oxford Handbook of the Psychology of Competition* (1st ed., pp. 569–597). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190060800.013.24

Darnon, C., Smeding, A., & Redersdorff, S. (2018). Belief in school meritocracy as an ideological barrier to the promotion of equality. *European Journal of Social Psychology*, *48*(4), 523–534. https://doi.org/10.1002/ejsp.2347

Enders, C. K. (2022). *Applied missing data analysis* (2nd ed., pp. ix, 546). The Guilford Press.

Mijs, J. J. B. (2016). The unfulfillable promise of meritocracy: Three lessons and their implications for justice in education. *Social Justice Research*, *29*(1), 14–34. https://doi.org/10.1007/s11211-014-0228-0

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (1st ed.). John Wiley & Sons. https://doi.org/10.1002/9780470316696

Stanczak, A., Jury, M., Aelenei, C., Pironom, J., Toczek-Capelle, M.-C., & Rohmer, O. (2024). Special education and meritocratic inclusion. *Educational Policy*, *38*(1), 85–103. https://doi.org/10.1177/08959048231153606

Wiederkehr, V., Bonnot, V., Krauth-Gruber, S., & Darnon, C. (2015). Belief in school meritocracy as a system-justifying tool for low status students. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01053

Table S1

Descriptive Statistics and t-Tests for Effort and Warmth by SEN Status (Study 1, Study 2, and Study 3)

	Without SEN			With SEN			Test Statistics		
Variables	*M (SD)*	Range	α	*M (SD)*	Range	α	*t (df)*	*p*	*d* [95% CI]
Study 1									
Effort	3.93 (0.82)	1-5	0.91	3.89 (0.85)	1-5	0.90	1.60 (461)	.110	0.07 [-0.02, 0.17]
Warmth	3.76 (1.00)	1-5	0.98	3.73 (1.00)	1-5	0.98	2.85 (452)	.005**	0.13 [0.04, 0.23]
Study 2									
Effort	3.68 (0.89)	1-5	0.91	3.61 (0.94)	1-5	0.93	3.11 (379)	.002**	0.16 [0.06, 0.26]
Warmth	3.29 (1.04)	1-5	0.99	3.28 (1.05)	1-5	0.99	0.70 (184)	.486	0.04 [-0.07, 0.15]
Study 3									
Effort	3.79 (0.86)	1-5	0.93	3.73 (0.90)	1-5	0.93	2.37 (349)	.057†	0.18 [0.02, 0.23]
Warmth	3.49 (1.04)	1-5	0.99	3.47 (1.05)	1-5	0.99	1.07 (136)	.286	0.07 [-0.05, 0.19]

Note. Means (*M*) and standard deviations (*SD*) were pooled across 20 imputed datasets using Rubin's Rules (Rubin, 1987), while Cronbach's alphas (α) were averaged across imputations. Paired *t*-tests examined differences by SEN status, with degrees of freedom (*df*) estimated using the Barnard & Rubin (1999) small-sample adjustment. Repeated-measures Cohen's *d* was computed as the mean difference divided by its standard deviation, with confidence intervals (CI) reflecting both within- and between-imputation variance. *SD*s were approximated as the average sample *SD* across imputations (Enders, 2022). Statistical differences are highlighted as follows: †$p < .10$; **$p < .01$. The lower *df* for Warmth in Study 2 and Study 3 reflect limited variability in Warmth ratings between students with and without SEN.

Table S2

Summary of Additional Regression Analyses Controlling for Covariates (Study 1, Study 2, and Study 3)

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Study 1												
Intercept	0.50	0.09	[0.32, 0.69]	5.44 (449)	**<.001*** **	0.061	0.19	0.03	[0.13, 0.25]	6.14 (441)	**<.001*** **	0.078
Threat Level	0.04	0.17	[-0.29, 0.36]	0.23 (455)	0.819	0.000	-0.02	0.06	[-0.13, 0.09]	-0.36 (453)	0.722	0.000
Student Gender	-0.18	0.17	[-0.50, 0.15]	-1.07 (455)	0.285	0.003	0.05	0.06	[-0.06, 0.16]	0.93 (454)	0.354	0.002
Participants' Gender	0.45	0.19	[0.08, 0.82]	2.41 (423)	**0.016***	0.013	0.11	0.06	[-0.01, 0.23]	1.74 (411)	**0.083†**	0.007
Threat Level × Student Gender	-0.22	0.33	[-0.88, 0.43]	-0.67 (455)	0.502	0.001	0.07	0.11	[-0.15, 0.29]	0.65 (453)	0.514	0.001
Study 2												
Intercept	0.68	0.14	[0.41, 0.96]	4.89 (370)	**<.001*** **	0.060	0.19	0.05	[0.10, 0.28]	4.12 (365)	**<.001*** **	0.044
Threat Level	-0.04	0.16	[-0.36, 0.28]	-0.24 (378)	0.808	0.000	-0.03	0.05	[-0.13, 0.08]	-0.47 (372)	0.639	0.001
Student Gender	0.14	0.16	[-0.18, 0.46]	0.88 (378)	0.381	0.002	0.05	0.05	[-0.05, 0.16]	0.97 (370)	0.334	0.003
Participants' Gender	-0.04	0.24	[-0.51, 0.44]	-0.15 (367)	0.884	0.000	-0.07	0.08	[-0.23, 0.08]	-0.93 (361)	0.354	0.002
Marginalized Group	0.12	0.21	[-0.29, 0.53]	0.57 (354)	0.568	0.001	-0.10	0.07	[-0.23, 0.04]	-1.41 (354)	0.158	0.006
Threat Level × Student Gender	0.39	0.33	[-0.25, 1.04]	1.20 (378)	0.231	0.004	0.13	0.11	[-0.08, 0.35]	1.26 (373)	0.209	0.004
Study 3												
Intercept	0.61	0.15	[0.31, 0.91]	3.99 (298)	**<.001*** **	0.046	0.25	0.05	[0.15, 0.36]	4.87 (341)	**<.001*** **	0.064
Contrast 1 (CGMF vs. others)	0.00	0.05	[-0.10, 0.10]	0.01 (356)	0.995	0.000	-0.03	0.02	[-0.06, 0.01]	-1.59 (347)	0.114	0.007
Contrast 2 (CGFM vs. others)	0.04	0.07	[-0.10, 0.18]	0.55 (356)	0.585	0.001	0.01	0.02	[-0.04, 0.06]	0.38 (340)	0.703	0.000
Contrast 3 (SGM vs. others)	0.03	0.12	[-0.22, 0.27]	0.22 (356)	0.823	0.000	0.00	0.04	[-0.09, 0.08]	-0.04 (346)	0.968	0.000
Participants'Gender	0.01	0.27	[-0.53, 0.54]	0.02 (301)	0.984	0.000	-0.02	0.09	[-0.20, 0.16]	-0.22 (335)	0.828	0.000
Marginalized Group	-0.10	0.24	[-0.57, 0.37]	-0.43 (241)	0.666	0.001	-0.04	0.08	[-0.20, 0.11]	-0.53 (315)	0.593	0.001

Note. *b* = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = t-value; *df* = degrees of freedom; *p* = *p*-value; η^2_p = partial eta-squared. Threat Level (-0.5 = low threat condition, 0.5 = high threat condition); Student Gender (-0.5 = boy context, 0.5 = girl context). Contrast 1 (CGMF vs. others: CGMF = +3, CGFM = -1, SGM = -1, SGF = -1); Contrast 2 (CGFM vs. others: CGMF = 0, CGFM = +2, SGM = -1, SGF = -1); Contrast 3 (SGM vs. others: CGMF = 0, CGFM = 0, SGM = +1, SGF = -1). Experimental conditions: CGMF = Cross-Gender Male-Female (boy without SEN vs. girl with SEN); CGFM = Cross-Gender Female-Male (girl without SEN vs. boy with SEN); SGM = Same-Gender Male (boy without SEN vs. boy with SEN); SGF = Same-Gender Female (girl without SEN vs. girl with SEN). Participants' Gender (-0.5 = male participants, 0.5 = participants identifying as non-male, including women and gender-diverse individuals); Marginalized Group (-0.5 = participants not identifying as marginalized, 0.5 = participants identifying as part of a group facing discrimination in France). Statistical differences are highlighted as follows: *$p < .05$; ***$p < .001$.

Table S3

Summary of Moderation Analyses for Grading and Competence (Study 1, Study 2, and Study 3)

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Study 1												
Fairness												
Intercept	0.59	0.08	[0.43, 0.75]	7.24 (452)	**<.001*** **	0.104	0.21	0.03	[0.16, 0.27]	7.67 (447)	**<.001*** **	0.116
Threat Level	0.07	0.16	[-0.25, 0.4]	0.45 (451)	0.651	0.000	-0.02	0.06	[-0.12, 0.09]	-0.28 (449)	0.778	0.000
Student Gender	-0.16	0.16	[-0.48, 0.16]	-0.99 (451)	0.321	0.002	0.05	0.06	[-0.06, 0.16]	0.94 (449)	0.347	0.002
Fairness	-0.30	0.07	[-0.44, -0.17]	-4.43 (439)	**<.001*** **	0.042	-0.07	0.02	[-0.12, -0.03]	-3.09 (425)	**0.002** **	0.021
Threat Level × Student Gender	-0.34	0.33	[-0.98, 0.30]	-1.04 (452)	0.301	0.002	0.05	0.11	[-0.17, 0.26]	0.41 (450)	0.679	0.000
Threat Level × Fairness	0.03	0.14	[-0.24, 0.30]	0.23 (444)	0.819	0.000	-0.03	0.05	[-0.12, 0.06]	-0.67 (429)	0.505	0.001
Student Gender × Fairness	0.15	0.14	[-0.12, 0.42]	1.06 (427)	0.288	0.003	-0.02	0.05	[-0.11, 0.07]	-0.4 (421)	0.688	0.000
Threat Level × Student Gender × Fairness	-0.43	0.27	[-0.97, 0.11]	-1.56 (437)	0.120	0.005	-0.09	0.09	[-0.27, 0.09]	-0.97 (422)	0.330	0.002
Meritocratic Beliefs												
Intercept	0.60	0.08	[0.43, 0.76]	7.14 (452)	**<.001*** **	0.101	0.21	0.03	[0.16, 0.27]	7.65 (447)	**<.001*** **	0.004
Threat Level	0.07	0.17	[-0.26, 0.40]	0.41 (452)	0.680	0.000	-0.02	0.06	[-0.13, 0.09]	-0.31 (450)	0.758	0.007
Student Gender	-0.18	0.17	[-0.51, 0.14]	-1.09 (452)	0.274	0.003	0.05	0.06	[-0.06, 0.16]	0.92 (450)	0.357	0.001
Meritocracy	-0.25	0.13	[-0.51, 0.01]	-1.9 (376)	**0.058**†	0.009	0.04	0.04	[-0.04, 0.13]	0.94 (440)	0.347	0.002
Threat Level × Student Gender	-0.20	0.33	[-0.85, 0.46]	-0.59 (452)	0.556	0.001	0.07	0.11	[-0.15, 0.29]	0.63 (451)	0.526	0.000
Threat Level × Meritocracy	0.10	0.26	[-0.41, 0.61]	0.38 (415)	0.702	0.000	0.15	0.09	[-0.02, 0.32]	1.73 (427)	**0.084**†	0.007
Student Gender × Meritocracy	-0.08	0.26	[-0.59, 0.44]	-0.29 (400)	0.769	0.000	0.07	0.09	[-0.10, 0.24]	0.85 (435)	0.398	0.002
Threat Level × Student Gender × Meritocracy	0.31	0.53	[-0.72, 1.34]	0.59 (405)	0.558	0.001	0.10	0.17	[-0.24, 0.45]	0.59 (414)	0.554	0.001
Study 2												
Fairness												
Intercept	0.63	0.08	[0.47, 0.79]	7.73 (376)	**<.001*** **	0.137	0.19	0.03	[0.14, 0.24]	7.13 (375)	**<.001*** **	0.119
Threat Level	-0.04	0.16	[-0.36, 0.28]	-0.24 (376)	0.814	0.000	-0.02	0.05	[-0.13, 0.08]	-0.42 (369)	0.678	0.001

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Student Gender	0.16	0.16	[-0.16, 0.48]	0.99 (376)	0.325	0.003	0.05	0.05	[-0.06, 0.15]	0.92 (368)	0.358	0.002
Fairness	-0.16	0.08	[-0.32, -0.01]	-2.10 (353)	**0.036***	0.012	-0.04	0.02	[-0.09, 0.00]	-1.80 (371)	**0.073†**	0.009
Threat Level × Student Gender	0.41	0.33	[-0.23, 1.06]	1.27 (376)	0.205	0.004	0.12	0.11	[-0.09, 0.33]	1.15 (371)	0.253	0.004
Threat Level × Fairness	0.05	0.15	[-0.26, 0.35]	0.30 (361)	0.761	0.000	0.09	0.05	[-0.01, 0.19]	1.76 (371)	**0.079†**	0.008
Student Gender × Fairness	0.03	0.15	[-0.28, 0.33]	0.17 (359)	0.867	0.000	-0.05	0.05	[-0.14, 0.05]	-0.92 (374)	0.358	0.002
Threat Level × Student Gender × Fairness	0.00	0.31	[-0.61, 0.60]	0.00 (361)	0.997	0.000	0.07	0.10	[-0.13, 0.26]	0.68 (371)	0.497	0.001
Meritocratic Beliefs												
Intercept	0.63	0.08	[0.47, 0.79]	7.64 (376)	**<.001*** **	0.134	0.19	0.03	[0.14, 0.24]	7.14 (375)	**<.001*** **	0.012
Threat Level	-0.04	0.16	[-0.37, 0.28]	-0.27 (375)	0.790	0.000	-0.02	0.05	[-0.13, 0.08]	-0.43 (368)	0.665	0.005
Student Gender	0.14	0.16	[-0.18, 0.46]	0.87 (377)	0.384	0.002	0.06	0.05	[-0.05, 0.16]	1.02 (368)	0.307	0.000
Meritocracy	-0.20	0.15	[-0.49, 0.08]	-1.40 (358)	0.164	0.005	-0.02	0.05	[-0.11, 0.07]	-0.38 (371)	0.708	0.000
Threat Level × Student Gender	0.38	0.33	[-0.27, 1.02]	1.15 (377)	0.250	0.003	0.12	0.11	[-0.09, 0.33]	1.09 (371)	0.276	0.000
Threat Level × Meritocracy	-0.24	0.29	[-0.80, 0.33]	-0.81 (365)	0.417	0.002	0.12	0.09	[-0.06, 0.31]	1.32 (374)	0.188	0.005
Student Gender × Meritocracy	0.12	0.29	[-0.45, 0.69]	0.41 (364)	0.682	0.001	0.04	0.09	[-0.14, 0.23]	0.45 (371)	0.652	0.001
Threat Level × Student Gender × Meritocracy	-0.31	0.58	[-1.46, 0.83]	-0.54 (353)	0.592	0.001	0.10	0.19	[-0.26, 0.47]	0.56 (373)	0.578	0.001
Study 3												
Fairness												
Intercept	0.65	0.09	[0.48, 0.82]	7.60 (354)	**<.001*** **	0.140	0.26	0.03	[0.20, 0.32]	8.64 (348)	**<.001*** **	0.175
Contrast 1 (CGMF vs. others)	0.00	0.05	[-0.10, 0.09]	-0.09 (354)	0.931	0.000	-0.03	0.02	[-0.06, 0.01]	-1.64 (344)	0.102	0.008
Fairness	-0.12	0.08	[-0.27, 0.04]	-1.51 (350)	0.133	0.006	-0.04	0.03	[-0.10, 0.01]	-1.60 (342)	0.110	0.007
Contrast 2 (CGFM vs. others)	0.03	0.07	[-0.11, 0.17]	0.46 (354)	0.648	0.001	0.01	0.02	[-0.04, 0.06]	0.36 (338)	0.717	0.000
Contrast 3 (SGM vs. others)	-0.01	0.12	[-0.25, 0.24]	-0.06 (354)	0.953	0.000	-0.01	0.04	[-0.10, 0.07]	-0.29 (345)	0.772	0.000
Contrast 1 × Fairness	0.01	0.04	[-0.08, 0.09]	0.12 (341)	0.907	0.000	0.01	0.01	[-0.01, 0.04]	1.00 (350)	0.319	0.003
Fairness × Contrast 2	0.14	0.07	[0.01, 0.27]	2.12 (354)	**0.035***	0.012	0.02	0.02	[-0.02, 0.07]	0.93 (331)	0.351	0.003
Fairness × Contrast 3	0.13	0.11	[-0.1, 0.35]	1.13 (354)	0.260	0.004	0.02	0.04	[-0.06, 0.10]	0.55 (350)	0.584	0.001
Meritocratic Beliefs												

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Intercept	0.66	0.09	[0.49, 0.83]	7.59 (354)	**<.001*** **	0.139	0.26	0.03	[0.20, 0.32]	8.66 (348)	**<.001*** **	0.176
Contrast 1 (CGMF vs. others)	0.00	0.05	[-0.1, 0.09]	-0.08 (354)	0.933	0.000	-0.03	0.02	[-0.06, 0.00]	-1.67 (346)	**0.096**†	0.008
Meritocracy	-0.08	0.15	[-0.39, 0.22]	-0.54 (285)	0.588	0.001	-0.02	0.05	[-0.13, 0.08]	-0.46 (300)	0.644	0.001
Contrast 2 (CGFM vs. others)	0.04	0.07	[-0.10, 0.17]	0.50 (354)	0.618	0.001	0.01	0.02	[-0.04, 0.06]	0.36 (340)	0.721	0.000
Contrast 3 (SGM vs. others)	0.03	0.13	[-0.21, 0.28]	0.26 (354)	0.795	0.000	0.00	0.04	[-0.09, 0.08]	-0.04 (344)	0.965	0.000
Contrast 1 × Meritocracy	0.03	0.08	[-0.13, 0.20]	0.36 (339)	0.717	0.000	0.02	0.03	[-0.04, 0.07]	0.57 (343)	0.567	0.001
Meritocracy × Contrast 2	-0.01	0.12	[-0.25, 0.23]	-0.08 (270)	0.938	0.000	-0.03	0.04	[-0.11, 0.05]	-0.73 (261)	0.469	0.002
Meritocracy × Contrast 3	-0.24	0.23	[-0.68, 0.21]	-1.03 (337)	0.303	0.003	-0.06	0.08	[-0.22, 0.09]	-0.81 (347)	0.417	0.002

Note. Dependent variables reflect within-subject difference scores (ratings for the student without SEN minus ratings for the student with SEN). *b* = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = t-value; *df* = degrees of freedom; *p* = *p*-value; η^2_p = partial eta-squared. Threat Level (-0.5 = low threat condition, 0.5 = high threat condition); Student Gender (-0.5 = boy context, 0.5 = girl context). Contrast 1 (CGMF vs. others: CGMF = +3, CGFM = -1, SGM = -1, SGF = -1); Contrast 2 (CGFM vs. others: CGMF = 0, CGFM = +2, SGM = -1, SGF = -1); Contrast 3 (SGM vs. others: CGMF = 0, CGFM = 0, SGM = +1, SGF = -1). Experimental conditions: CGMF = Cross-Gender Male-Female (boy without SEN vs. girl with SEN); CGFM = Cross-Gender Female-Male (girl without SEN vs. boy with SEN); SGM = Same-Gender Male (boy without SEN vs. boy with SEN); SGF = Same-Gender Female (girl without SEN vs. girl with SEN). Fairness and Meritocracy variables were centered at their means. Statistical differences are highlighted as follows: †$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

Table S4

Summary of Not Preregistered Exploratory Regression Analyses by SEN Status for Grading and Competence (Study 1, Study 2, and Study 3)

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Study 1												
Fairness												
Students Without SEN												
Intercept	6.21	0.05	[6.11, 6.32]	118.10 (453)	**<.001*** **	0.968	3.67	0.03	[3.61, 3.74]	113.35 (451)	**<.001*** **	0.966
Threat Level	-0.03	0.11	[-0.24, 0.17]	-0.32 (453)	0.749	0.000	0.04	0.06	[-0.08, 0.17]	0.68 (450)	0.497	0.001
Student Gender	0.15	0.11	[-0.06, 0.35]	1.38 (453)	0.168	0.004	0.16	0.06	[0.03, 0.29]	2.44 (450)	**0.015***	0.013
Fairness	0.02	0.04	[-0.07, 0.10]	0.38 (411)	0.705	0.000	0.09	0.03	[0.04, 0.15]	3.47 (438)	**0.001***	0.026
Threat Level × Student Gender	-0.01	0.21	[-0.42, 0.40]	-0.04 (453)	0.971	0.000	-0.15	0.13	[-0.41, 0.10]	-1.20 (452)	0.232	0.003
Threat Level × Fairness	-0.04	0.09	[-0.21, 0.14]	-0.41 (421)	0.681	0.000	-0.03	0.05	[-0.14, 0.08]	-0.55 (434)	0.581	0.001
Student Gender × Fairness	-0.06	0.09	[-0.24, 0.11]	-0.69 (426)	0.488	0.001	0.01	0.05	[-0.09, 0.12]	0.20 (447)	0.840	0.000
Threat Level × Student Gender × Fairness	0.13	0.18	[-0.21, 0.48]	0.76 (441)	0.447	0.001	0.05	0.11	[-0.17, 0.26]	0.43 (444)	0.667	0.000
Students With SEN												
Intercept	5.62	0.08	[5.47, 5.77]	74.09 (452)	**<.001*** **	0.924	3.46	0.03	[3.39, 3.53]	101.02 (451)	**<.001*** **	0.958
Threat Level	-0.11	0.15	[-0.41, 0.19]	-0.71 (450)	0.477	0.001	0.06	0.07	[-0.07, 0.19]	0.87 (447)	0.385	0.002
Student Gender	0.31	0.15	[0.01, 0.61]	2.03 (451)	**0.043***	0.009	0.11	0.07	[-0.03, 0.24]	1.54 (447)	0.125	0.005
Fairness	0.32	0.06	[0.19, 0.45]	5.00 (420)	**<.001*** **	0.054	0.17	0.03	[0.11, 0.22]	5.77 (425)	**<.001*** **	0.071
Threat Level × Student Gender	0.33	0.30	[-0.26, 0.93]	1.09 (452)	0.274	0.003	-0.20	0.14	[-0.47, 0.07]	-1.47 (452)	0.143	0.005
Threat Level × Fairness	-0.07	0.13	[-0.32, 0.18]	-0.53 (428)	0.594	0.001	0.00	0.06	[-0.11, 0.11]	0.02 (423)	0.984	0.000
Student Gender × Fairness	-0.21	0.13	[-0.46, 0.05]	-1.60 (373)	0.111	0.006	0.03	0.06	[-0.08, 0.14]	0.52 (425)	0.606	0.001
Threat Level × Student Gender × Fairness	0.56	0.26	[0.06, 1.07]	2.18 (405)	**0.030***	0.011	0.14	0.12	[-0.09, 0.36]	1.20 (426)	0.232	0.003

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Meritocratic Beliefs												
Students Without SEN												
Intercept	6.21	0.05	[6.11, 6.31]	118.09 (453)	**<.001*** **	0.968	3.68	0.03	[3.61, 3.74]	112.10 (451)	**<.001*** **	0.965
Threat Level	-0.03	0.11	[-0.23, 0.18]	-0.25 (453)	0.804	0.000	0.04	0.07	[-0.09, 0.17]	0.60 (450)	0.550	0.001
Student Gender	0.15	0.11	[-0.06, 0.35]	1.41 (453)	0.159	0.004	0.16	0.07	[0.03, 0.29]	2.47 (450)	**0.014***	0.013
Meritocracy	-0.06	0.08	[-0.22, 0.10]	-0.78 (430)	0.433	0.001	0.03	0.05	[-0.07, 0.13]	0.66 (441)	0.509	0.001
Threat Level × Student Gender	-0.01	0.21	[-0.42, 0.40]	-0.05 (453)	0.959	0.000	-0.19	0.13	[-0.45, 0.06]	-1.46 (452)	0.144	0.005
Threat Level × Meritocracy	-0.10	0.16	[-0.42, 0.22]	-0.63 (438)	0.530	0.001	0.05	0.10	[-0.15, 0.24]	0.44 (426)	0.658	0.001
Student Gender × Meritocracy	-0.01	0.16	[-0.33, 0.31]	-0.07 (430)	0.946	0.000	0.12	0.10	[-0.08, 0.32]	1.18 (440)	0.239	0.003
Threat Level × Student Gender × Meritocracy	0.03	0.33	[-0.61, 0.68]	0.10 (418)	0.919	0.000	-0.09	0.20	[-0.49, 0.31]	-0.45 (431)	0.649	0.001
Students With SEN												
Intercept	5.61	0.08	[5.46, 5.77]	71.55 (452)	**<.001*** **	0.919	3.46	0.04	[3.39, 3.53]	97.29 (451)	**<.001*** **	0.954
Threat Level	-0.10	0.16	[-0.40, 0.21]	-0.61 (452)	0.544	0.001	0.06	0.07	[-0.08, 0.20]	0.79 (448)	0.429	0.001
Student Gender	0.33	0.16	[0.02, 0.64]	2.11 (452)	**0.035***	0.010	0.11	0.07	[-0.03, 0.25]	1.55 (448)	0.123	0.005
Meritocracy	0.19	0.13	[-0.06, 0.44]	1.49 (333)	0.137	0.006	-0.01	0.05	[-0.11, 0.10]	-0.13 (438)	0.896	0.000
Threat Level × Student Gender	0.19	0.31	[-0.43, 0.80]	0.59 (452)	0.553	0.001	-0.26	0.14	[-0.54, 0.02]	-1.85 (452)	**0.065†**	0.007
Threat Level × Meritocracy	-0.20	0.25	[-0.69, 0.29]	-0.81 (377)	0.416	0.002	-0.11	0.11	[-0.32, 0.11]	-0.95 (414)	0.345	0.002
Student Gender × Meritocracy	0.07	0.25	[-0.42, 0.55]	0.27 (388)	0.788	0.000	0.05	0.11	[-0.17, 0.26]	0.42 (418)	0.678	0.000
Threat Level × Student Gender × Meritocracy	-0.27	0.51	[-1.27, 0.72]	-0.54 (335)	0.588	0.001	-0.20	0.22	[-0.63, 0.24]	-0.89 (429)	0.376	0.002
Study 2												
Fairness												
Students Without SEN												
Intercept	6.21	0.06	[6.09, 6.34]	100.05 (375)	**<.001*** **	0.964	3.45	0.04	[3.37, 3.52]	90.06 (375)	**<.001*** **	0.956

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Threat Level	-0.02	0.12	[-0.26, 0.22]	-0.16 (375)	0.873	0.000	-0.02	0.08	[-0.17, 0.13]	-0.26 (373)	0.798	0.000
Student Gender	0.09	0.12	[-0.15, 0.33]	0.71 (375)	0.475	0.001	0.00	0.08	[-0.15, 0.16]	0.06 (374)	0.949	0.000
Fairness	0.08	0.06	[-0.04, 0.19]	1.33 (365)	0.183	0.005	0.12	0.04	[0.05, 0.19]	3.27 (343)	**0.001****	0.029
Threat Level × Student Gender	-0.14	0.25	[-0.62, 0.35]	-0.54 (375)	0.586	0.001	0.15	0.15	[-0.15, 0.45]	0.98 (374)	0.326	0.003
Threat Level × Fairness	0.11	0.12	[-0.12, 0.34]	0.93 (363)	0.353	0.002	0.09	0.07	[-0.05, 0.24]	1.28 (320)	0.202	0.005
Student Gender × Fairness	0.11	0.12	[-0.12, 0.34]	0.97 (365)	0.332	0.003	0.01	0.07	[-0.14, 0.15]	0.09 (334)	0.932	0.000
Threat Level × Student Gender × Fairness	-0.34	0.23	[-0.80, 0.12]	-1.45 (362)	0.147	0.006	-0.10	0.15	[-0.39, 0.19]	-0.66 (327)	0.512	0.001
Students With SEN												
Intercept	5.58	0.08	[5.42, 5.75]	66.22 (376)	**<.001*****	0.921	3.26	0.04	[3.18, 3.33]	81.86 (375)	**<.001*****	0.947
Threat Level	0.02	0.17	[-0.31, 0.35]	0.11 (376)	0.912	0.000	0.00	0.08	[-0.15, 0.16]	0.03 (370)	0.975	0.000
Student Gender	-0.07	0.17	[-0.40, 0.26]	-0.43 (376)	0.668	0.000	-0.04	0.08	[-0.20, 0.11]	-0.56 (373)	0.578	0.001
Fairness	0.24	0.08	[0.08, 0.40]	3.02 (358)	**0.003****	0.024	0.16	0.04	[0.09, 0.24]	4.31 (338)	**<.001*****	0.049
Threat Level × Student Gender	-0.55	0.34	[-1.21, 0.11]	-1.63 (376)	0.104	0.007	0.03	0.16	[-0.28, 0.34]	0.18 (373)	0.858	0.000
Threat Level × Fairness	0.06	0.16	[-0.25, 0.37]	0.39 (360)	0.698	0.000	0.01	0.08	[-0.14, 0.16]	0.09 (328)	0.929	0.000
Student Gender × Fairness	0.09	0.16	[-0.22, 0.40]	0.55 (361)	0.583	0.001	0.05	0.08	[-0.10, 0.20]	0.69 (343)	0.494	0.001
Threat Level × Student Gender × Fairness	-0.34	0.32	[-0.96, 0.28]	-1.07 (366)	0.286	0.003	-0.16	0.15	[-0.46, 0.13]	-1.08 (332)	0.282	0.003
Meritocratic Beliefs												
Students Without SEN												
Intercept	6.21	0.06	[6.09, 6.33]	99.74 (375)	**<.001*****	0.963	3.45	0.04	[3.37, 3.52]	89.28 (376)	**<.001*****	0.955
Threat Level	-0.03	0.12	[-0.28, 0.21]	-0.26 (375)	0.797	0.000	-0.02	0.08	[-0.18, 0.13]	-0.31 (374)	0.756	0.000
Student Gender	0.10	0.12	[-0.15, 0.34]	0.77 (375)	0.440	0.002	0.03	0.08	[-0.12, 0.18]	0.40 (374)	0.690	0.000
Meritocracy	-0.13	0.11	[-0.34, 0.09]	-1.15 (372)	0.249	0.004	-0.14	0.07	[-0.27, 0.00]	-2.00 (354)	**0.046***	0.011
Threat Level × Student Gender	-0.14	0.25	[-0.63, 0.34]	-0.58 (375)	0.565	0.001	0.15	0.16	[-0.16, 0.45]	0.96 (372)	0.340	0.002
Threat Level × Meritocracy	0.05	0.22	[-0.38, 0.48]	0.24 (371)	0.808	0.000	0.23	0.14	[-0.04, 0.50]	1.65 (355)	0.100	0.007

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Student Gender × Meritocracy	-0.27	0.22	[-0.69, 0.16]	-1.22 (374)	0.223	0.004	-0.01	0.14	[-0.28, 0.26]	-0.04 (351)	0.969	0.000
Threat Level × Student Gender × Meritocracy	-0.57	0.44	[-1.43, 0.29]	-1.31 (371)	0.191	0.005	0.03	0.27	[-0.50, 0.57]	0.13 (357)	0.899	0.000
Students With SEN												
Intercept	5.58	0.09	[5.42, 5.75]	65.44 (376)	**<.001*** **	0.919	3.26	0.04	[3.18, 3.34]	79.79 (375)	**<.001*** **	0.944
Threat Level	0.01	0.17	[-0.32, 0.35]	0.07 (376)	0.946	0.000	0.00	0.08	[-0.16, 0.16]	-0.01 (372)	0.993	0.000
Student Gender	-0.05	0.17	[-0.38, 0.29]	-0.27 (377)	0.786	0.000	-0.02	0.08	[-0.18, 0.14]	-0.30 (372)	0.766	0.000
Meritocracy	0.08	0.15	[-0.22, 0.37]	0.51 (364)	0.610	0.001	-0.12	0.07	[-0.26, 0.02]	-1.66 (358)	**0.098†**	0.008
Threat Level × Student Gender	-0.52	0.34	[-1.19, 0.15]	-1.53 (377)	0.128	0.006	0.03	0.16	[-0.29, 0.35]	0.19 (372)	0.852	0.000
Threat Level × Meritocracy	0.29	0.30	[-0.30, 0.88]	0.96 (364)	0.340	0.003	0.10	0.14	[-0.18, 0.39]	0.71 (357)	0.477	0.001
Student Gender × Meritocracy	-0.39	0.30	[-0.98, 0.21]	-1.27 (362)	0.204	0.004	-0.05	0.15	[-0.33, 0.24]	-0.33 (350)	0.743	0.000
Threat Level × Student Gender × Meritocracy	-0.26	0.61	[-1.45, 0.93]	-0.43 (359)	0.669	0.001	-0.07	0.29	[-0.64, 0.50]	-0.24 (360)	0.809	0.000
Study 3												
Model A: Contrast-Based Fairness												
Students Without SEN												
Intercept	6.35	0.06	[6.22, 6.47]	102.92 (354)	**<.001*** **	0.967	3.58	0.04	[3.51, 3.65]	97.76 (344)	**<.001*** **	0.965
Contrast 1	-0.03	0.03	[-0.09, 0.04]	-0.73 (354)	0.463	0.002	0.04	0.02	[0.00, 0.08]	1.75 (348)	**0.081†**	0.009
Fairness	0.06	0.06	[-0.05, 0.17]	1.01 (351)	0.315	0.003	0.08	0.03	[0.02, 0.15]	2.48 (340)	**0.014***	0.018
Contrast 2	0.05	0.05	[-0.05, 0.15]	0.93 (354)	0.350	0.002	0.05	0.03	[0.00, 0.11]	1.82 (343)	**0.070†**	0.009
Contrast 3	-0.04	0.09	[-0.21, 0.13]	-0.45 (354)	0.656	0.001	-0.05	0.05	[-0.15, 0.05]	-0.94 (349)	0.346	0.003
Contrast 1 × Fairness	0.03	0.03	[-0.03, 0.09]	1.09 (338)	0.278	0.004	0.02	0.02	[-0.02, 0.05]	0.92 (346)	0.357	0.002
Fairness × Contrast 2	0.02	0.05	[-0.07, 0.11]	0.37 (354)	0.713	0.000	-0.02	0.03	[-0.08, 0.03]	-0.86 (345)	0.392	0.002
Fairness × Contrast 3	0.10	0.08	[-0.06, 0.26]	1.17 (354)	0.242	0.004	0.05	0.05	[-0.05, 0.14]	0.94 (349)	0.348	0.003

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Students With SEN												
Intercept	5.69	0.08	[5.53, 5.86]	68.20 (354)	**<.001*** **	0.929	3.32	0.04	[3.24, 3.39]	83.26 (347)	**<.001*** **	0.952
Contrast 1	-0.02	0.05	[-0.11, 0.07]	-0.45 (353)	0.651	0.001	0.06	0.02	[0.02, 0.11]	2.85 (348)	**0.005** **	0.023
Fairness	0.18	0.08	[0.02, 0.33]	2.28 (342)	**0.023***	0.015	0.13	0.04	[0.06, 0.20]	3.50 (346)	**0.001** **	0.034
Contrast 2	0.02	0.07	[-0.12, 0.15]	0.22 (354)	0.826	0.000	0.05	0.03	[-0.02, 0.11]	1.38 (337)	0.167	0.006
Contrast 3	-0.03	0.12	[-0.27, 0.20]	-0.27 (354)	0.788	0.000	-0.04	0.06	[-0.15, 0.08]	-0.64 (345)	0.522	0.001
Contrast 1 × Fairness	0.03	0.04	[-0.06, 0.11]	0.67 (305)	0.504	0.002	0.00	0.02	[-0.04, 0.04]	0.09 (346)	0.927	0.000
Fairness × Contrast 2	-0.12	0.06	[-0.25, 0.00]	-1.91 (354)	**0.057**†	0.010	-0.05	0.03	[-0.11, 0.01]	-1.51 (347)	0.133	0.006
Fairness × Contrast 3	-0.03	0.11	[-0.25, 0.19]	-0.30 (354)	0.768	0.000	0.02	0.05	[-0.08, 0.13]	0.45 (350)	0.656	0.001
Meritocratic Beliefs												
Students Without SEN												
Intercept	6.34	0.06	[6.22, 6.46]	103.13 (354)	**<.001*** **	0.968	3.57	0.04	[3.50, 3.65]	96.34 (344)	**<.001*** **	0.964
Contrast 1	-0.03	0.03	[-0.09, 0.04]	-0.77 (354)	0.442	0.002	0.03	0.02	[-0.01, 0.08]	1.67 (347)	**0.095**†	0.008
Meritocracy	-0.23	0.11	[-0.44, -0.01]	-2.08 (290)	**0.039***	0.013	-0.04	0.06	[-0.17, 0.08]	-0.66 (339)	0.507	0.001
Contrast 2	0.05	0.05	[-0.05, 0.15]	1.00 (354)	0.320	0.003	0.06	0.03	[0.00, 0.12]	1.84 (343)	**0.066**†	0.010
Contrast 3	-0.01	0.09	[-0.18, 0.16]	-0.11 (354)	0.915	0.000	-0.06	0.05	[-0.16, 0.05]	-1.09 (348)	0.278	0.003
Contrast 1 × Meritocracy	0.02	0.06	[-0.10, 0.14]	0.36 (277)	0.719	0.001	-0.01	0.04	[-0.08, 0.06]	-0.36 (347)	0.720	0.000
Meritocracy × Contrast 2	0.15	0.09	[-0.02, 0.31]	1.73 (284)	**0.085**†	0.009	0.01	0.05	[-0.08, 0.11]	0.29 (334)	0.769	0.000
Meritocracy × Contrast 3	-0.12	0.16	[-0.44, 0.20]	-0.72 (330)	0.470	0.002	-0.03	0.10	[-0.22, 0.15]	-0.36 (341)	0.719	0.000
Students With SEN												
Intercept	5.68	0.08	[5.52, 5.85]	67.35 (354)	**<.001*** **	0.927	3.31	0.04	[3.23, 3.39]	81.46 (347)	**<.001*** **	0.950
Contrast 1	-0.02	0.05	[-0.12, 0.07]	-0.47 (354)	0.636	0.001	0.06	0.02	[0.02, 0.11]	2.78 (350)	**0.006** **	0.021
Meritocracy	-0.14	0.15	[-0.43, 0.14]	-0.98 (331)	0.330	0.003	-0.02	0.07	[-0.15, 0.12]	-0.25 (345)	0.802	0.000
Contrast 2	0.01	0.07	[-0.12, 0.15]	0.21 (354)	0.833	0.000	0.05	0.03	[-0.02, 0.11]	1.41 (338)	0.161	0.006
Contrast 3	-0.04	0.12	[-0.28, 0.20]	-0.34 (354)	0.730	0.000	-0.06	0.06	[-0.17, 0.06]	-0.95 (346)	0.341	0.003

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Contrast 1 × Meritocracy	-0.01	0.08	[-0.17, 0.15]	-0.10 (317)	0.921	0.000	-0.03	0.04	[-0.11, 0.05]	-0.75 (345)	0.451	0.002
Meritocracy × Contrast 2	0.16	0.11	[-0.06, 0.38]	1.39 (345)	0.166	0.006	0.05	0.05	[-0.06, 0.15]	0.83 (328)	0.408	0.002
Meritocracy × Contrast 3	0.12	0.22	[-0.31, 0.55]	0.53 (348)	0.593	0.001	0.03	0.11	[-0.18, 0.24]	0.28 (349)	0.781	0.000
Model B: Factor-Based												
Fairness												
Students Without SEN												
Intercept	6.35	0.06	[6.22, 6.47]	102.92 (354)	**<.001*** **	0.967	3.58	0.04	[3.51, 3.65]	97.76 (344)	**<.001*** **	0.965
Gender (SEN Students)	-0.06	0.12	[-0.30, 0.18]	-0.48 (354)	0.635	0.001	0.07	0.07	[-0.08, 0.21]	0.92 (347)	0.356	0.002
Gender Contrast	0.04	0.12	[-0.20, 0.28]	0.35 (354)	0.728	0.000	0.18	0.07	[0.04, 0.32]	2.47 (343)	**0.014***	0.017
Fairness	0.06	0.06	[-0.05, 0.17]	1.01 (351)	0.315	0.003	0.08	0.03	[0.02, 0.15]	2.48 (340)	**0.014***	0.018
Gender (SEN Students) × Gender Contrast	-0.28	0.25	[-0.76, 0.21]	-1.12 (354)	0.264	0.004	-0.06	0.15	[-0.35, 0.22]	-0.44 (350)	0.663	0.001
Gender (SEN Students) × Fairness	-0.05	0.11	[-0.27, 0.18]	-0.41 (351)	0.684	0.000	0.01	0.07	[-0.12, 0.14]	0.18 (345)	0.860	0.000
Gender (SEN Students) × Fairness	0.10	0.11	[-0.12, 0.32]	0.90 (350)	0.367	0.002	-0.01	0.07	[-0.14, 0.12]	-0.22 (349)	0.829	0.000
Gender (SEN Students) × Gender Contrast × Fairness	0.29	0.23	[-0.15, 0.74]	1.29 (350)	0.196	0.005	0.21	0.13	[-0.06, 0.47]	1.54 (346)	0.123	0.007
Students With SEN												
Intercept	5.69	0.08	[5.53, 5.86]	68.20 (354)	**<.001*** **	0.929	3.32	0.04	[3.24, 3.39]	83.26 (347)	**<.001*** **	0.952
Gender (SEN Students)	-0.03	0.17	[-0.35, 0.30]	-0.15 (354)	0.879	0.000	0.12	0.08	[-0.04, 0.28]	1.50 (346)	0.133	0.006
Gender Contrast	-0.01	0.17	[-0.34, 0.31]	-0.08 (354)	0.939	0.000	0.22	0.08	[0.06, 0.38]	2.73 (337)	**0.007** **	0.021
Fairness	0.18	0.08	[0.02, 0.33]	2.28 (342)	**0.023***	0.015	0.13	0.04	[0.06, 0.20]	3.50 (346)	**0.001** **	0.034
Gender (SEN Students) × Gender Contrast	-0.18	0.33	[-0.83, 0.47]	-0.54 (354)	0.590	0.001	0.09	0.16	[-0.22, 0.40]	0.58 (347)	0.562	0.001
Gender (SEN Students) × Fairness	0.21	0.15	[-0.09, 0.51]	1.37 (342)	0.170	0.005	0.03	0.07	[-0.12, 0.17]	0.35 (347)	0.724	0.000
Gender (SEN Students) × Fairness	-0.19	0.15	[-0.49, 0.12]	-1.20 (344)	0.229	0.004	-0.09	0.07	[-0.23, 0.05]	-1.21 (349)	0.227	0.004

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Gender (SEN Students) × Gender Contrast × Fairness	0.29	0.31	[-0.31, 0.90]	0.95 (344)	0.343	0.003	0.15	0.15	[-0.14, 0.43]	1.00 (348)	0.316	0.003
Meritocratic Beliefs												
Students Without SEN												
Intercept	6.34	0.06	[6.22, 6.46]	103.13 (354)	**<.001*** **	0.968	3.57	0.04	[3.50, 3.65]	96.34 (344)	**<.001*** **	0.964
Gender (SEN Students)	-0.09	0.12	[-0.33, 0.15]	-0.76 (354)	0.446	0.002	0.07	0.07	[-0.07, 0.22]	0.97 (347)	0.331	0.003
Gender Contrast	0.05	0.12	[-0.19, 0.29]	0.38 (354)	0.706	0.000	0.18	0.07	[0.04, 0.33]	2.44 (342)	**0.015***	0.017
Meritocracy	-0.23	0.11	[-0.44, -0.01]	-2.08 (290)	**0.039***	0.013	-0.04	0.06	[-0.17, 0.08]	-0.66 (339)	0.507	0.001
Gender (SEN Students) × Gender Contrast	-0.23	0.25	[-0.71, 0.26]	-0.92 (354)	0.359	0.002	-0.09	0.15	[-0.38, 0.20]	-0.60 (349)	0.551	0.001
Gender (SEN Students) × Meritocracy	0.01	0.22	[-0.42, 0.45]	0.07 (260)	0.947	0.000	-0.01	0.13	[-0.25, 0.24]	-0.04 (340)	0.967	0.000
Gender (SEN Students) × Meritocracy	0.34	0.21	[-0.08, 0.76]	1.58 (323)	0.116	0.007	0.00	0.13	[-0.24, 0.25]	0.03 (345)	0.976	0.000
Gender (SEN Students) × Gender Contrast × Meritocracy	-0.44	0.43	[-1.29, 0.40]	-1.02 (318)	0.308	0.003	-0.15	0.25	[-0.65, 0.35]	-0.59 (338)	0.558	0.001
Students With SEN												
Intercept	5.68	0.08	[5.52, 5.85]	67.35 (354)	**<.001*** **	0.927	3.31	0.04	[3.23, 3.39]	81.46 (347)	**<.001*** **	0.950
Gender (SEN Students)	-0.02	0.17	[-0.35, 0.31]	-0.10 (354)	0.917	0.000	0.14	0.08	[-0.02, 0.30]	1.68 (347)	**0.095†**	0.008
Gender Contrast	-0.02	0.17	[-0.35, 0.31]	-0.10 (354)	0.924	0.000	0.22	0.08	[0.06, 0.38]	2.70 (339)	**0.007****	0.021
Meritocracy	-0.14	0.15	[-0.43, 0.14]	-0.98 (331)	0.330	0.003	-0.02	0.07	[-0.15, 0.12]	-0.25 (345)	0.802	0.000
Gender (SEN Students) × Gender Contrast	-0.20	0.34	[-0.86, 0.46]	-0.60 (354)	0.547	0.001	0.05	0.16	[-0.27, 0.37]	0.30 (349)	0.767	0.000
Gender (SEN Students) × Meritocracy	-0.29	0.29	[-0.86, 0.28]	-1.00 (343)	0.318	0.003	-0.13	0.14	[-0.41, 0.14]	-0.95 (339)	0.341	0.003

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Gender (SEN Students) × Meritocracy	0.30	0.29	[-0.28, 0.87]	1.01 (335)	0.312	0.003	0.03	0.14	[-0.24, 0.30]	0.23 (344)	0.819	0.000
Gender (SEN Students) × Gender Contrast × Meritocracy	-0.11	0.58	[-1.26, 1.03]	-0.19 (338)	0.849	0.000	-0.15	0.28	[-0.69, 0.40]	-0.54 (343)	0.591	0.001

Note. b = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = t-value; *df* = degrees of freedom; *p* = *p*-value; η^2_p = partial eta-squared. Threat Level (-0.5 = low threat condition, 0.5 = high threat condition); Student Gender (-0.5 = boy context, 0.5 = girl context). Contrast 1 (CGMF vs. others: CGMF = +3, CGFM = -1, SGM = -1, SGF = -1); Contrast 2 (CGFM vs. others: CGMF = 0, CGFM = +2, SGM = -1, SGF = -1); Contrast 3 (SGM vs. others: CGMF = 0, CGFM = 0, SGM = +1, SGF = -1). Experimental conditions: CGMF = Cross-Gender Male-Female (boy without SEN vs. girl with SEN); CGFM = Cross-Gender Female-Male (girl without SEN vs. boy with SEN); SGM = Same-Gender Male (boy without SEN vs. boy with SEN); SGF = Same-Gender Female (girl without SEN vs. girl with SEN). Gender (SEN students): -0.5 = boy with SEN (i.e., experimental conditions SGM and CGFM), 0.5 = girl with SEN (i.e., experimental conditions SGF and CGMF); Gender Contrast: -0.5 = same-gender comparisons (i.e., experimental conditions SGM and SGF), 0.5 = cross-gender comparisons (i.e., experimental conditions CGMF and CGFM). Gender (-0.5 = male participants, 0.5 = participants identifying as non-male, including women and gender-diverse individuals); Marginalized Group (-0.5 = participants not identifying as marginalized, 0.5 = participants identifying as part of a group facing discrimination in France). Fairness and Meritocracy variables were centered at their means. Statistical differences are highlighted as follows: $^{\dagger}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Table S5

Summary of Preregistered Secondary Regression Analyses by SEN Status for Grading and Competence (Study 1, Study 2, and Study 3)

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Study 1												
Students Without SEN												
Intercept	6.21	0.05	[6.11, 6.31]	118.69 (457)	**<.001*** **	0.968	3.67	0.03	[3.61, 3.74]	112.49 (455)	**<.001*** **	0.965
Threat Level	-0.03	0.10	[-0.23, 0.18]	-0.27 (457)	0.786	0.000	0.04	0.07	[-0.08, 0.17]	0.68 (454)	0.495	0.001
Student Gender	0.15	0.10	[-0.06, 0.35]	1.42 (457)	0.156	0.004	0.16	0.07	[0.03, 0.29]	2.46 (454)	**0.014***	0.013
Threat Level × Student Gender	-0.01	0.21	[-0.42, 0.40]	-0.07 (457)	0.948	0.000	-0.19	0.13	[-0.44, 0.07]	-1.43 (456)	0.152	0.004
Students With SEN												
Intercept	5.61	0.08	[5.45, 5.76]	71.66 (456)	**<.001*** **	0.918	3.46	0.04	[3.39, 3.53]	97.61 (455)	**<.001*** **	0.954
Threat Level	-0.08	0.16	[-0.39, 0.22]	-0.54 (456)	0.588	0.001	0.06	0.07	[-0.08, 0.20]	0.85 (452)	0.398	0.002
Student Gender	0.32	0.16	[0.01, 0.62]	2.01 (456)	**0.045***	0.009	0.11	0.07	[-0.03, 0.25]	1.49 (453)	0.136	0.005
Threat Level × Student Gender	0.22	0.31	[-0.39, 0.83]	0.70 (456)	0.483	0.001	-0.26	0.14	[-0.54, 0.02]	-1.82 (456)	**0.069†**	0.007
Study 2												
Students Without SEN												
Intercept	6.21	0.06	[6.09, 6.34]	99.93 (379)	**<.001*** **	0.963	3.44	0.04	[3.37, 3.52]	89.13 (379)	**<.001*** **	0.954
Threat Level	-0.02	0.12	[-0.27, 0.22]	-0.17 (379)	0.863	0.000	-0.02	0.08	[-0.17, 0.13]	-0.24 (377)	0.810	0.000
Student Gender	0.10	0.12	[-0.15, 0.34]	0.78 (379)	0.436	0.002	0.02	0.08	[-0.13, 0.17]	0.23 (378)	0.816	0.000
Threat Level × Student Gender	-0.11	0.25	[-0.60, 0.38]	-0.45 (379)	0.656	0.001	0.18	0.15	[-0.13, 0.48]	1.14 (377)	0.256	0.003
Students With SEN												
Intercept	5.58	0.08	[5.41, 5.75]	65.67 (380)	**<.001*** **	0.919	3.26	0.04	[3.18, 3.34]	80.09 (379)	**<.001*** **	0.944
Threat Level	0.02	0.17	[-0.31, 0.35]	0.12 (380)	0.903	0.000	0.01	0.08	[-0.15, 0.17]	0.06 (375)	0.951	0.000
Student Gender	-0.05	0.17	[-0.38, 0.28]	-0.30 (380)	0.763	0.000	-0.03	0.08	[-0.19, 0.13]	-0.38 (376)	0.705	0.000
Threat Level × Student Gender	-0.51	0.34	[-1.17, 0.16]	-1.49 (380)	0.136	0.006	0.05	0.16	[-0.27, 0.37]	0.33 (377)	0.739	0.000

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Study 3												
Model A: Contrast-Based												
Students Without SEN												
Intercept	6.34	0.06	[6.22, 6.46]	103.23 (358)	**<.001*** **	0.967	3.57	0.04	[3.50, 3.64]	97.35 (349)	**<.001*** **	0.964
Contrast 1 (CGMF vs. others)	-0.03	0.03	[-0.09, 0.04]	-0.74 (358)	0.461	0.002	0.04	0.02	[-0.01, 0.08]	1.71 (352)	**0.088†**	0.008
Contrast 2 (CGFM vs. others)	0.05	0.05	[-0.05, 0.15]	1.04 (358)	0.300	0.003	0.06	0.03	[0.00, 0.12]	1.89 (347)	**0.060†**	0.010
Contrast 3 (SGM vs. others)	-0.04	0.09	[-0.21, 0.13]	-0.46 (358)	0.646	0.001	-0.06	0.05	[-0.16, 0.04]	-1.17 (353)	0.244	0.004
Students With SEN												
Intercept	5.69	0.08	[5.53, 5.85]	67.90 (358)	**<.001*** **	0.928	3.31	0.04	[3.23, 3.39]	82.28 (352)	**<.001*** **	0.95
Contrast 1 (CGMF vs. others)	-0.02	0.05	[-0.12, 0.07]	-0.53 (358)	0.599	0.001	0.06	0.02	[0.02, 0.11]	2.77 (354)	**0.006****	0.021
Contrast 2 (CGFM vs. others)	0.01	0.07	[-0.12, 0.15]	0.18 (358)	0.858	0.000	0.05	0.03	[-0.02, 0.11]	1.42 (341)	0.157	0.006
Contrast 3 (SGM vs. others)	-0.07	0.12	[-0.30, 0.17]	-0.56 (358)	0.579	0.001	-0.06	0.06	[-0.17, 0.05]	-1.02 (349)	0.309	0.003
Model B: Factor-Based												
Students Without SEN												
Intercept	6.34	0.06	[6.22, 6.46]	103.23 (358)	**<.001*** **	0.967	3.57	0.04	[3.50, 3.64]	97.35 (349)	**<.001*** **	0.964
Gender (SEN Students)	-0.06	0.12	[-0.30, 0.18]	-0.51 (358)	0.608	0.001	0.08	0.07	[-0.07, 0.22]	1.03 (351)	0.303	0.003
Gender Contrast	0.05	0.12	[-0.19, 0.29]	0.43 (358)	0.668	0.001	0.18	0.07	[0.04, 0.33]	2.51 (347)	**0.013***	0.018
Gender (SEN Students) × Gender Contrast	-0.29	0.25	[-0.77, 0.19]	-1.17 (358)	0.242	0.004	-0.09	0.15	[-0.38, 0.19]	-0.65 (354)	0.519	0.001
Students With SEN												
Intercept	5.69	0.08	[5.53, 5.85]	67.90 (358)	**<.001*** **	0.928	3.31	0.04	[3.23, 3.39]	82.28 (352)	**<.001*** **	0.950
Gender (SEN Students)	0.00	0.17	[-0.32, 0.33]	0.03 (358)	0.978	0.000	0.14	0.08	[-0.02, 0.30]	1.71 (350)	**0.087†**	0.008
Gender Contrast	-0.03	0.17	[-0.35, 0.30]	-0.15 (358)	0.880	0.000	0.22	0.08	[0.06, 0.38]	2.71 (342)	**0.007****	0.021

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Gender (SEN Students) × Gender Contrast	-0.26	0.34	[-0.92, 0.40]	-0.77 (358)	0.441	0.002	0.04	0.16	[-0.28, 0.36]	0.25 (353)	0.804	0.000

Note. b = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = t-value; *df* = degrees of freedom; *p* = *p*-value; η^2_p = partial eta-squared. Threat Level (-0.5 = low threat condition, 0.5 = high threat condition); Student Gender (-0.5 = boy context, 0.5 = girl context). Contrast 1 (CGMF vs. others: CGMF = +3, CGFM = -1, SGM = -1, SGF = -1); Contrast 2 (CGFM vs. others: CGMF = 0, CGFM = +2, SGM = -1, SGF = -1); Contrast 3 (SGM vs. others: CGMF = 0, CGFM = 0, SGM = +1, SGF = -1). Experimental conditions: CGMF = Cross-Gender Male-Female (boy without SEN vs. girl with SEN); CGFM = Cross-Gender Female-Male (girl without SEN vs. boy with SEN); SGM = Same-Gender Male (boy without SEN vs. boy with SEN); SGF = Same-Gender Female (girl without SEN vs. girl with SEN). Gender (SEN students): -0.5 = boy with SEN (i.e., experimental conditions SGM and CGFM), 0.5 = girl with SEN (i.e., experimental conditions SGF and CGMF); Gender Contrast: -0.5 = same-gender comparisons (i.e., experimental conditions SGM and SGF), 0.5 = cross-gender comparisons (i.e., experimental conditions CGMF and CGFM). Statistical differences are highlighted as follows: $^{\dagger}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Table S6

Summary of Preregistered Secondary and Exploratory Regression Analyses for Effort (Study 1, Study 2, and Study 3)

	Perceived Effort					
Variables	*b*	*SE*	[95% CI]	*t*	*p*	η^2_p
Study 1						
Difference Score						
Intercept	0.04	0.03	[-0.01, 0.10]	1.55 (451)	0.123	0.005
Threat Level	0.01	0.06	[-0.1, 0.12]	0.26 (453)	0.793	0.000
Student Gender	0.02	0.06	[-0.09, 0.13]	0.34 (454)	0.733	0.000
Threat Level × Student Gender	0.06	0.11	[-0.16, 0.28]	0.53 (453)	0.599	0.001
Difference Score, Covariates						
Intercept	0.03	0.03	[-0.03, 0.09]	0.91 (445)	0.361	0.002
Threat Level	0.01	0.06	[-0.1, 0.12]	0.21 (452)	0.831	0.000
Student Gender	0.02	0.06	[-0.09, 0.13]	0.31 (452)	0.757	0.000
Gender	0.07	0.06	[-0.05, 0.19]	1.09 (432)	0.277	0.003
Threat Level × Student Gender	0.06	0.11	[-0.16, 0.28]	0.54 (451)	0.589	0.001
Students Without SEN						
Intercept	3.93	0.04	[3.86, 4.00]	103.51 (456)	**<.001*****	0.959
Threat Level	0.02	0.08	[-0.13, 0.17]	0.22 (456)	0.828	0.000
Student Gender	0.17	0.08	[0.02, 0.31]	2.18 (456)	**0.030***	0.010
Threat Level × Student Gender	-0.26	0.15	[-0.55, 0.04]	-1.68 (456)	**0.093**†	0.006
Students With SEN						
Intercept	3.89	0.04	[3.81, 3.96]	98.46 (455)	**<.001*****	0.955
Threat Level	0.00	0.08	[-0.15, 0.16]	0.02 (453)	0.981	0.000
Student Gender	0.15	0.08	[-0.01, 0.30]	1.85 (455)	**0.065**†	0.007
Threat Level × Student Gender	-0.31	0.16	[-0.62, 0.00]	-1.99 (453)	**0.047***	0.009
Study 2						
Difference Score						
Intercept	0.07	0.02	[0.03, 0.12]	3.01 (362)	**0.003****	0.024
Threat Level	-0.04	0.05	[-0.13, 0.06]	-0.81 (367)	0.420	0.002
Student Gender	0.00	0.05	[-0.09, 0.10]	0.06 (365)	0.952	0.000
Threat Level × Student Gender	0.15	0.10	[-0.04, 0.34]	1.54 (370)	0.125	0.006
Difference Score, Covariates						
Intercept	0.08	0.04	[-0.01, 0.16]	1.82 (310)	**0.070**†	0.010
Threat Level	-0.04	0.05	[-0.13, 0.06]	-0.82 (366)	0.415	0.002
Student Gender	0.00	0.05	[-0.09, 0.10]	0.08 (362)	0.937	0.000
Gender	-0.04	0.08	[-0.19, 0.11]	-0.49 (262)	0.622	0.001
Marginalized Group	-0.03	0.06	[-0.15, 0.09]	-0.55 (369)	0.583	0.001
Threat Level × Student Gender	0.15	0.10	[-0.04, 0.34]	1.59 (370)	0.113	0.007
Students Without SEN						
Intercept	3.68	0.05	[3.59, 3.77]	80.70 (378)	**<.001*****	0.945
Threat Level	0.02	0.09	[-0.16, 0.20]	0.22 (377)	0.825	0.000
Student Gender	-0.05	0.09	[-0.23, 0.13]	-0.50 (376)	0.614	0.001
Threat Level × Student Gender	0.18	0.18	[-0.18, 0.53]	0.96 (377)	0.336	0.002
Students With SEN						
Intercept	3.61	0.05	[3.52, 3.71]	74.27 (375)	**<.001*****	0.936

	Perceived Effort					
Variables	*b*	*SE*	[95% CI]	*t*	*p*	η^2_p
Threat Level	0.06	0.10	[-0.13, 0.25]	0.61 (374)	0.542	0.001
Student Gender	-0.05	0.10	[-0.24, 0.14]	-0.50 (375)	0.614	0.001
Threat Level × Student Gender	0.03	0.19	[-0.35, 0.41]	0.14 (377)	0.888	0.000
Study 3						
Model A: Contrast-Based						
Difference Score						
Intercept	0.06	0.02	[0.01, 0.10]	2.42 (328)	**0.016***	0.017
Contrast 1 (CGMF vs. others)	-0.03	0.01	[-0.06, -0.01]	-2.45 (347)	**0.015***	0.017
Contrast 2 (CGFM vs. others)	0.05	0.02	[0.01, 0.09]	2.52 (350)	**0.012***	0.018
Contrast 3 (SGM vs. others)	-0.02	0.03	[-0.08, 0.05]	-0.47 (344)	0.636	0.001
Difference Score, Covariates						
Intercept	0.04	0.04	[-0.04, 0.12]	0.93 (288)	0.353	0.003
Contrast 1 (CGMF vs. others)	-0.03	0.01	[-0.06, -0.01]	-2.40 (346)	**0.017***	0.016
Contrast 2 (CGFM vs. others)	0.05	0.02	[0.01, 0.08]	2.46 (345)	**0.014***	0.017
Contrast 3 (SGM vs. others)	-0.02	0.03	[-0.08, 0.05]	-0.45 (341)	0.652	0.001
Gender	0.01	0.07	[-0.14, 0.15]	0.08 (264)	0.934	0.000
Marginalized Group	-0.05	0.07	[-0.18, 0.08]	-0.78 (206)	0.435	0.003
Students Without SEN						
Intercept	3.78	0.04	[3.70, 3.87]	84.66 (355)	**<.001****	0.953
Contrast 1 (CGMF vs. others)	0.02	0.03	[-0.03, 0.07]	0.80 (356)	0.425	0.002
Contrast 2 (CGFM vs. others)	0.13	0.04	[0.06, 0.20]	3.48 (355)	**0.001****	0.033
Contrast 3 (SGM vs. others)	-0.10	0.06	[-0.22, 0.03]	-1.48 (351)	0.139	0.006
Students With SEN						
Intercept	3.73	0.05	[3.64, 3.82]	79.50 (352)	**<.001****	0.947
Contrast 1 (CGMF vs. others)	0.05	0.03	[0.00, 0.10]	1.99 (357)	**0.047***	0.011
Contrast 2 (CGFM vs. others)	0.08	0.04	[0.00, 0.15]	2.07 (354)	**0.039***	0.012
Contrast 3 (SGM vs. others)	-0.08	0.07	[-0.21, 0.05]	-1.18 (353)	0.237	0.004
Model B: Factor-Based						
Difference Score						
Intercept	0.06	0.02	[0.01, 0.10]	2.42 (328)	**0.016***	0.017
Gender (SEN Students)	-0.10	0.05	[-0.19, -0.01]	-2.07 (346)	**0.039***	0.012
Gender Contrast	0.03	0.05	[-0.06, 0.12]	0.67 (351)	0.501	0.001
Gender (SEN Students) × Gender Contrast	-0.26	0.09	[-0.44, -0.07]	-2.75 (345)	**0.006****	0.021
Difference Score, Covariates						
Intercept	0.04	0.04	[-0.04, 0.12]	0.93 (288)	0.353	0.003
Gender (SEN Students)	-0.10	0.05	[-0.19, 0.00]	-2.04 (344)	**0.042***	0.012
Gender Contrast	0.03	0.05	[-0.06, 0.12]	0.66 (347)	0.508	0.001
Gender	0.01	0.07	[-0.14, 0.15]	0.08 (264)	0.934	0.000
Marginalized Group	-0.05	0.07	[-0.18, 0.08]	-0.78 (206)	0.435	0.003
Gender (SEN Students) × Gender Contrast	-0.25	0.09	[-0.44, -0.07]	-2.68 (341)	**0.008****	0.020
Students Without SEN						
Intercept	3.78	0.04	[3.70, 3.87]	84.66 (355)	**<.001****	0.953
Gender (SEN Students)	0.01	0.09	[-0.17, 0.18]	0.10 (354)	0.922	0.000
Gender Contrast	0.30	0.09	[0.12, 0.47]	3.30 (355)	**0.001****	0.030
Gender (SEN Students) × Gender Contrast	-0.37	0.18	[-0.72, -0.01]	-2.04 (353)	**0.042***	0.012

	Perceived Effort					
Variables	*b*	*SE*	[95% CI]	*t*	*p*	η^2_p
Students With SEN						
Intercept	3.73	0.05	[3.64, 3.82]	79.50 (352)	**<.001*** **	0.947
Gender (SEN Students)	0.11	0.09	[-0.08, 0.29]	1.13 (354)	0.260	0.004
Gender Contrast	0.26	0.09	[0.08, 0.45]	2.82 (354)	**0.005** **	0.022
Gender (SEN Students) × Gender Contrast	-0.11	0.19	[-0.47, 0.26]	-0.58 (355)	0.566	0.001

Note. Difference scores were computed as: ratings for the student without SEN minus ratings for the student with SEN. *b* = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = t-value; *df* = degrees of freedom; *p* = *p*-value; η^2_p = partial eta-squared. Threat Level (-0.5 = low threat condition, 0.5 = high threat condition); Student Gender (-0.5 = boy context, 0.5 = girl context). Contrast 1 (CGMF vs. others: CGMF = +3, CGFM = -1, SGM = -1, SGF = -1); Contrast 2 (CGFM vs. others: CGMF = 0, CGFM = +2, SGM = -1, SGF = -1); Contrast 3 (SGM vs. others: CGMF = 0, CGFM = 0, SGM = +1, SGF = -1). Experimental conditions: CGMF = Cross-Gender Male-Female (boy without SEN vs. girl with SEN); CGFM = Cross-Gender Female-Male (girl without SEN vs. boy with SEN); SGM = Same-Gender Male (boy without SEN vs. boy with SEN); SGF = Same-Gender Female (girl without SEN vs. girl with SEN). Gender (SEN students): -0.5 = boy with SEN (i.e., experimental conditions SGM and CGFM), 0.5 = girl with SEN (i.e., experimental conditions SGF and CGMF); Gender Contrast: -0.5 = same-gender comparisons (i.e., experimental conditions SGM and SGF), 0.5 = cross-gender comparisons (i.e., experimental conditions CGMF and CGFM). Gender (-0.5 = male participants, 0.5 = participants identifying as non-male, including women and gender-diverse individuals); Marginalized Group (-0.5 = participants not identifying as marginalized, 0.5 = participants identifying as part of a group facing discrimination in France). Statistical differences are highlighted as follows: $^{\dagger}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Table S7

Summary of Preregistered Exploratory Regression Analyses for Grading and Competence (Study 3)

	Grading						Perceived Competence					
Variables	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p	*b*	*SE*	[95% CI]	*t (df)*	*p*	η^2_p
Factor-Based Model												
Intercept	0.65	0.09	[0.48, 0.82]	7.54 (358)	**<.001*** **	0.136	0.26	0.03	[0.20, 0.32]	8.62 (352)	**<.001*** **	0.173
Gender (SEN Students)	-0.07	0.17	[-0.40, 0.27]	-0.39 (358)	0.694	0.000	-0.06	0.06	[-0.18, 0.06]	-1.04 (345)	0.301	0.003
Gender Contrast	0.08	0.17	[-0.26, 0.42]	0.45 (358)	0.649	0.001	-0.04	0.06	[-0.15, 0.08]	-0.59 (347)	0.553	0.001
Gender (SEN Students) × Gender Contrast	-0.03	0.34	[-0.70, 0.64]	-0.09 (358)	0.932	0.000	-0.13	0.12	[-0.37, 0.10]	-1.11 (347)	0.268	0.004

Note. Dependent variables reflect within-subject difference scores (ratings for the student without SEN minus ratings for the student with SEN). *b* = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = t-value; *df* = degrees of freedom; *p* = *p*-value; η^2_p = partial eta-squared. Gender (SEN students): -0.5 = boy with SEN (i.e., experimental conditions SGM and CGFM), 0.5 = girl with SEN (i.e., experimental conditions SGF and CGMF); Gender Contrast: -0.5 = same-gender comparisons (i.e., experimental conditions SGM and SGF), 0.5 = cross-gender comparisons (i.e., experimental conditions CGMF and CGFM). Experimental conditions: CGMF = Cross-Gender Male-Female (boy without SEN vs. girl with SEN); CGFM = Cross-Gender Female-Male (girl without SEN vs. boy with SEN); SGM = Same-Gender Male (boy without SEN vs. boy with SEN); SGF = Same-Gender Female (girl without SEN vs. girl with SEN).Statistical differences are highlighted as follows: ***$p < .001$.