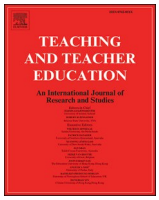
Teaching and Teacher Education 171 (2026) 105304

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate



Research paper

When accommodations are not enough: A multi-study examination of teacher bias toward students with special educational needs across student gender☆

Fabian Müller [a,b], Cristina Aelenei [c], Mickaël Jury [a,d,*]

[a] Laboratoire ACTé, Université Clermont Auvergne, Clermont-Ferrand, France
[b] Université Paris Cité, CNRS, LaPsyDÉ, F-75005 Paris, France
[c] Université Paris Cité, Laboratoire de Psychologie Sociale, Boulogne-Billancourt, France
[d] Institut Universitaire de France, Paris, France

ARTICLE INFO

Keywords:
Teacher bias
Special educational needs
Gender
Fairness perceptions
Meritocracy

ABSTRACT

This research examines whether a student's special educational needs (SEN) status accompanied by accommodations influences teachers' evaluations of performance, and whether 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 were devalued in grades and competence—a *backlash effect*—and whether fairness perceptions moderated this bias. We operationalized an SEN case as an ADHD-labeled student receiving reduced-exercise accommodations. In Studies 1–2, 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 versus 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 support fairness-focused teacher education to promote inclusive evaluation.

1. Introduction

A teacher sits down to assess two student assignments. One student completed all 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 accommodated student's success 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). International policy reviews show that teachers across systems, such as in North America, Asia, and Australia, often report low confidence in evaluating students with SEN and high demand for related training (Brussino, 2020; OECD, 2019). Exam accommodations (e.g., extended test time or adapted instructions) aim to help students reach and demonstrate their potential. Together, these measures promote fairness by addressing individual needs, recognizing that equal treatment does not always mean identical treatment (Deutsch, 1975). Accordingly, we adopt a needs-based perspective on inclusion: our interest lies not in diagnostic labels but in whether students who receive accommodations—regardless of their label—are fairly evaluated. In this view, inclusive practice is about meeting needs, not categorizing students.

☆ We gratefully acknowledge funding from the Clermont Auvergne Métropole. Parts of this research were presented at the LICIE 2025 Conference (Jury, Müller, & Aelenei, 2025) and are planned to be presented at upcoming conferences. A preprint of this manuscript is available on PsyArXiv (Müller et al., 2025a). 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.

* Corresponing author. Laboratoire ACTé, INSPE Clermont-Auvergne, 36 avenue Jean-Jaurès CS 20001, Cedex, 63407, Chamalières, France.
E-mail address: mickael.jury@uca.fr (M. Jury).

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

https://doi.org/10.1016/j.tate.2025.105304
Received 17 June 2025; Received in revised form 2 November 2025; Accepted 12 November 2025
Available online 6 December 2025

However, even when students with SEN are included in general classrooms and benefit from accommodations, they may still face bias or lowered expectations. A critical yet underexplored issue is how teachers evaluate their 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 they perform well with accommodations (Stanczak, Aelenei, et al., 2024). Such biases—often rooted in stereotypes, i.e., generalized beliefs about social groups—risk undermining inclusion by limiting recognition of students with SEN's abilities. This research addresses that concern by examining teachers' evaluative judgments of students with SEN, while also considering student gender and a psychological factor (fairness perceptions) that might influence these judgments.

1.1. 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 stereotyped as less competent (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 shows 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' abilities (Cohen et al., 2019; Krischler & Pit-ten Cate, 2020). These expectations can shape teacher behavior, from the difficulty of material to grading strictness. Classic work on teacher expectations (e.g., the Pygmalion effect) shows biased expectations can become self-fulfilling, affecting student performance and teachers' interpretations of behavior (Jussim & Harber, 2005). If a teacher assumes a student with SEN will struggle, they may attribute success to external support rather than ability—and grade 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 showing 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 discount their success (e.g., lower competence ratings, harsher grading, or attributing success to unfair advantages). Batruch et al. (2017) demonstrated a similar pattern 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, preserving beliefs about who deserves to succeed. By analogy, students with SEN who succeed might face backlash, as they too challenge a stereotype of low competence.

Stanczak, Aelenei, et al. (2024) provide direct empirical evidence for 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. When accommodations seemed more legitimate (e.g., assistive technology), the devaluation was smaller. The authors interpreted this as evidence that teachers may use backlash as an "ideological barrier" to full inclusion—a psychological mechanism preserving the meritocratic status quo. 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. It isn't a real accomplishment."

1.2. Meritocratic ideology and system justification in education

To understand why teachers might engage in biased evaluations, we must consider the broader ideological framework of schooling (Florian, 2014). Modern schooling, particularly in industrialized societies, emphasizes meritocracy (Butera et al., 2024)—the belief that academic success should solely reflect individual talent and effort, assuming equal opportunities for all (Darnon, Wiederkehr, et al., 2018; Mijs, 2016). Ideally, this 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 inequalities by attributing success or failure solely to individual factors like effort or talent (Darnon, Smeding, & Redersdorff, 2018; Stanczak, Jury, et al., 2024). If all students are supposedly given the same initial conditions and only merit counts, then any outcome inequality appears justified—high achievers earned their place, while others 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 students' different needs and starting points. Achieving equity requires differentiated treatment—such as accommodations—rather than uniform standards (Sireci et al., 2005). This clashes with strict meritocratic views of fairness, which assume 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 internalize meritocratic ideals may view accommodations (e.g., extra test time, adjusted standards) as violating equal competition. In this view, accommodations are not equity tools but distortions of merit-based evaluation. A student with SEN's success may then appear less "earned." If teachers think the student had it "easier" due to accommodations, 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 with SEN more favorably, whereas those who saw them as tilting the field (i.e., unfair) were more likely to discount students' success.

1.3. Gender stereotypes and intersection with SEN

While research increasingly explores biases in evaluating students with SEN, a critical gap remains: how student gender intersects with SEN status to influence teacher evaluations, as SEN- and gender-related issues seem interconnected (Brussino, 2020). Most research treats students with SEN as a homogeneous group, overlooking 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 can seem atypical. 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 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, 2023b).

These trends mirror common stereotypes associating mathematical talent with boys and reading abilities with girls—biases that can undermine girls' motivation early on (Bian et al., 2017; Jenifer et al., 2024; Leslie et al., 2015) and may influence teachers' competence

assessments. 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 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 pronounced devaluation. Alternatively, teachers might 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 buffer bias.

In short, teachers' judgments are not always gender-neutral; they may reflect stereotypes such as "girls are better readers" or "boys are naturally better at math," or 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.

1.4. Research objectives and overview

Drawing on the above frameworks, this research addresses two critical gaps. First, while prior work suggests 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 bias intensifies when students with SEN outperform their non-SEN peers—thereby increasing perceptions of unexpected success and creating higher-threat situations (vs. lower-threat, where performance is equal). While Stanczak, Aelenei, et al. (2024) explored this possibility, their findings were mixed, highlighting the need for further testing. 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 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 as a central tension. They called for empirical testing of these contradictions and their consequences for students with SEN. We respond by extending their work in two key ways: (a) examining whether backlash generalizes across student gender, and (b) identifying for whom and under what conditions backlash is strongest by probing teachers' fairness perceptions. This contributes to both theory and practice: providing a more nuanced understanding of teacher biases in inclusive education, and informing interventions that promote fairer recognition of all students' achievements, including those with SEN.

Specifically, we address 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 this bias?
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 among pre-service teachers. Study 2 replicated this design with in-service teachers, assessing whether teaching experience influences these biases. Study 3 refined the design to compare backlash effects when a female student with SEN is evaluated against a male student without SEN, 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). 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 of Université Paris Cité (Comité d'Éthique de la Recherche – CER U-Paris Cité; IRB No.: 00012024-76). Informed consent was obtained; confidentiality, anonymity, GDPR compliance, and full debriefing were ensured.

2. Study 1

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

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 male students (H1b), and under high-threat conditions, where students with SEN outperform 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 one 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.

2.1. Method

2.1.1. Participants and procedure

Participants. Data were collected online via LimeSurvey from pre-service teachers across France. Multiple INSPE[2] centers distributed the study via email. The goal was to maximize responses within a four-week 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"; Table 1 for demographics).

A sensitivity power analysis using G*Power 3.1.9.4 (Faul et al., 2009) determined the minimum detectable effect sizes. For main and interaction effects (student gender, threat level, and their interaction; H1–H2), and with $\alpha = .05$, power = .80, and a final sample of $N = 463$, the study was powered to detect an effect of partial eta squared, $\eta^2_p = .023$. 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 $\eta^2_p = .017$. 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 vs. with SEN) as a within-subject

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

Table 1
Demographic characteristics of pre-service and in-service teachers (Studies 1–3).

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%)
Teacher Role	Specialized Teacher	–	22 (5.7%)	22 (6.0%)
Number of Educational Stages Taught	1	–	298 (80.1%)	281 (80.1%)
	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 ($M = 23.90$, $SD = 5.40$, 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, Trainee Status, and Teacher Role were collected only in Studies 2 and 3. Teacher Role refers to self-identified specialized teachers, i.e., teachers formally designated to support students with SEN in France. 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: Studies 2 and 3 included participants teaching in *Maternelle* (Preschool), *Élémentaire* (Primary), *Collège* (Middle), and *Lycée* (High School). The majority taught a single stage (most commonly *Élémentaire*), with fewer participants teaching across multiple stages (e.g., *Maternelle* and *Élémentaire*).
As is standard in France, preschool (*Maternelle*) and primary (*Élémentaire*) teachers are generalists who teach and assess all core subjects, including mathematics, whereas secondary (*Collège, Lycée*) teachers are typically subject specialists. In these studies, all participants (pre- and in-service) graded the same short fourth-grade math tests requiring only routine grading judgments, not specialist expertise.
The *Diplôme d'Université* is a French university-specific diploma offering targeted professional or academic training, in this case focused on teacher education.

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, 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 intensify 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 the exercises due to ADHD accommodations; no such information was provided for the student without SEN. Reduced-exercise accommodations were selected because they are recognized in official policy frameworks for students with SEN in France (MENESR, 2015), including those with behavioral regulation difficulties. They are also used internationally in national assessments (Guez et al., 2024), making them suitable to operationalize an accommodated SEN case in written evaluations. Consistent with best practice for studying evaluative bias, this vignette grading task isolates accommodation effects while holding performance constant (Aguinis & Bradley, 2014; Krolak-Schwerdt et al., 2018).

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 than physical or less behaviorally disruptive neurodevelopmental disabilities like dyspraxia (Jury et al., 2021; Stanczak, Aelenei, et al., 2024). Importantly, ADHD served purely as an operationalization of an accommodated SEN case. Our approach was needs-based rather than diagnostic.

Student gender was manipulated by assigning traditionally male (Léo, Lucas) or female (Léa, Emma) names. 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 2024 classroom demographics. To mirror classroom norms, the non-SEN student's test was always presented first, as students in ordinary learning conditions represent the majority. 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. 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 on easy items and difficulties only with the most challenging. 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, demographics (e.g., age, gender) were collected. All procedures followed ethical guidelines, ensuring voluntary, anonymous participation and full debriefing.

2.1.2. 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

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

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 total; adapted from national evaluations aligned with the French curriculum, MENJ, 2023). Each section included two easy and two difficult items, with difficult ones in bold. Participants rated each test immediately after viewing it. In the version attributed to the student without SEN, all 20 items were completed, with eight errors 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 (8/20 for the student without SEN vs. 4/10 for the student with SEN). The only differences between the two tests within a condition were the number of completed items and error placement, ensuring a controlled comparison.

Perceived Competence. 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 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 into composite competence scores for the student without SEN ($\alpha = .83$) and with SEN ($\alpha = .85$), demonstrating good internal consistency. Descriptive statistics and reliability for the warmth and effort subscales are in Table S1 (Supplementary Material, SM).

Fairness of Accommodations. Participants evaluated the perceived fairness of five SEN accommodations—extra time, oral exams, computer use, 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 ADHD-related barriers. To prevent bias and avoid overemphasizing the half-exercise accommodation, all items were presented equally and in random order. Preregistered moderation analyses focused on the half-exercise item ($M = 3.42$, $SD = 1.22$, Range = 1–5), mean-centered prior to analysis.

2.2. Results

2.2.1. 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 >2.5 MADs below the median (Leys et al., 2013); none 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, warmth), with good to excellent model fit: 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 pooled datasets (via mice), pooled via the mitools package (v2.4; Lumley, 2019). To examine 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), threat level (low vs. high), and 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 competence difference scores.[5] Exploratory analyses tested perceived fairness as an independent predictor of grading and competence, separately for students with and without SEN, to explore whether effects were 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.

2.2.2. Primary analyses

Grading (Math Tests). Results confirmed a backlash effect: students with SEN received lower grades than their non-SEN peers ($b = 0.60$, 95% CI [0.44, 0.77], $SE = 0.08$, $t(456) = 7.20$, $p < .001$, $\eta_p^2 = .102$). However, this effect was unaffected by student gender, threat level, or their interaction (all $ps > .320$).

Perceived Competence. Similarly, students with SEN were judged less competent than their non-SEN peers ($b = 0.22$, 95% CI [0.16, 0.27], $SE = 0.03$, $t(452) = 7.69$, $p < .001$, $\eta_p^2 = .115$). Again, neither student gender, threat level, nor their interaction affected this backlash effect (all $ps > .328$).

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

2.2.3. Moderation analyses

Fairness of Accommodations. Higher fairness ratings predicted smaller grading differences ($b = -0.30$, 95% CI [−0.44, −0.17], $SE = 0.07$, $t(439) = -4.43$, $p < .001$, $\eta_p^2 = .042$) and reduced perceived competence differences ($b = -0.07$, 95% CI [−0.12, −0.03], $SE = 0.02$, $t(425) = -3.09$, $p = .002$, $\eta_p^2 = .021$) between students with and without SEN (Fig. 1). Teachers who viewed accommodations as fair evaluated students with SEN more similarly to their non-SEN peers. However,

[4] While EFA was preregistered to explore factor structures of the Social Judgment Scales, pooling EFA results across multiple imputations is inappropriate. We therefore conducted CFA. An exploratory EFA on one imputed dataset supported these theoretical expectations.

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

Table 2
Descriptive statistics and *t*-tests for grading and competence evaluations by SEN status (Studies 1–3).

Variables	Without SEN		With SEN		Test Statistics		
	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 and 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: ***$p < .001$. Higher ratings for students without SEN reflect potential teacher bias (i.e., backlash against accommodated students with SEN).

Table 3
Summary of primary regression analyses for teacher bias in grading and competence (Studies 1–3).

Variables	Grading						Perceived Competence					
	b	*SE*	[95% CI]	*t (df)*	*p*	η_p^2	*b*	*SE*	[95% CI]	*t (df)*	*p*	η_p^2
Study 1												
Intercept	0.60	0.08	[0.44, 0.77]	7.20 (456)	<.001***	.102	0.22	0.03	[0.16, 0.27]	7.69 (452)	<.001***	.115
Threat Level	0.06	0.17	[−0.27, 0.38]	0.34 (456)	.736	.000	−0.02	0.06	[−0.13, 0.09]	−0.28 (454)	.783	.000
Student Gender	−0.17	0.17	[−0.49, 0.16]	−1.00 (456)	.320	.002	0.05	0.06	[−0.05, 0.16]	0.98 (454)	.328	.002
Threat Level × Student Gender	−0.23	0.33	[−0.89, 0.42]	−0.70 (456)	.485	.001	0.07	0.11	[−0.15, 0.29]	0.63 (455)	.529	.001
Study 2												
Intercept	0.63	0.08	[0.47, 0.79]	7.75 (380)	<.001***	.136	0.19	0.03	[0.14, 0.24]	7.10 (379)	<.001***	.117
Threat Level	−0.04	0.16	[−0.36, 0.28]	−0.26 (380)	.797	.000	−0.02	0.05	[−0.13, 0.08]	−0.44 (373)	.659	.001
Student Gender	0.15	0.16	[−0.17, 0.47]	0.91 (380)	.365	.002	0.05	0.05	[−0.06, 0.15]	0.91 (372)	.363	.002
Threat Level × Student Gender	0.40	0.33	[−0.24, 1.04]	1.22 (380)	.225	.004	0.12	0.11	[−0.09, 0.33]	1.14 (375)	.255	.003
Study 3												
Intercept	0.65	0.09	[0.48, 0.82]	7.54 (358)	<.001***	.136	0.26	0.03	[0.20, 0.32]	8.62 (352)	<.001***	.173
Contrast 1 (CGMF vs. others)	0.00	0.05	[−0.10, 0.09]	−0.01 (358)	.988	.000	−0.03	0.02	[−0.06, 0.01]	−1.61 (349)	.107	.007
Contrast 2 (CGFM vs. others)	0.04	0.07	[−0.10, 0.18]	0.57 (358)	.571	.001	0.01	0.02	[−0.04, 0.06]	0.39 (342)	.696	.001
Contrast 3 (SGM vs. others)	0.03	0.12	[−0.22, 0.27]	0.21 (358)	.831	.000	0.00	0.04	[−0.09, 0.08]	−0.05 (348)	.960	.000

Note. Dependent variables reflect within-subject difference scores (ratings for the student without SEN minus ratings for the student with SEN), indexing teacher bias (backlash effect). *b* = unstandardized regression coefficient; *SE* = standard error; CI = confidence interval; *t* = *t*-value; *df* = degrees of freedom; *p* = *p*-value; η_p^2 = 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: ***$p < .001$.

fairness did not interact with student gender, threat level, or their combination to moderate the backlash effect (all *p*s > .120; Table S3 in SM for full results).

2.2.4. 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, 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], $SE = 0.03$, $t(438) = 3.47$, $p = .001$, $\eta_p^2 = .026$) but not grading ($p = .705$). For students with SEN, fairness predicted both grading ($b = 0.32$, 95% CI [0.19, 0.45], $SE = 0.06$, $t(420) = 5.00$, $p < .001$, $\eta_p^2 = .054$) and perceived competence ($b = 0.17$, 95% CI [0.11, 0.22], $SE = 0.03$, $t(425) = 5.77$, $p < .001$, $\eta_p^2 = .071$). See Table S4 and corresponding section in SM for exploratory interaction results not central to our hypotheses.

2.3. Discussion

The goal of Study 1 was to test whether students with SEN receive lower grading and competence evaluations (i.e., a "backlash effect") compared to 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 fairness perceptions moderate the backlash (H3).

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

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

Exploratory analyses indicated that fairness perceptions influenced judgments for both students with and without SEN, but effects were stronger for students with SEN. This suggests fairness concerns are more salient—and perhaps more consequential—in inclusive education

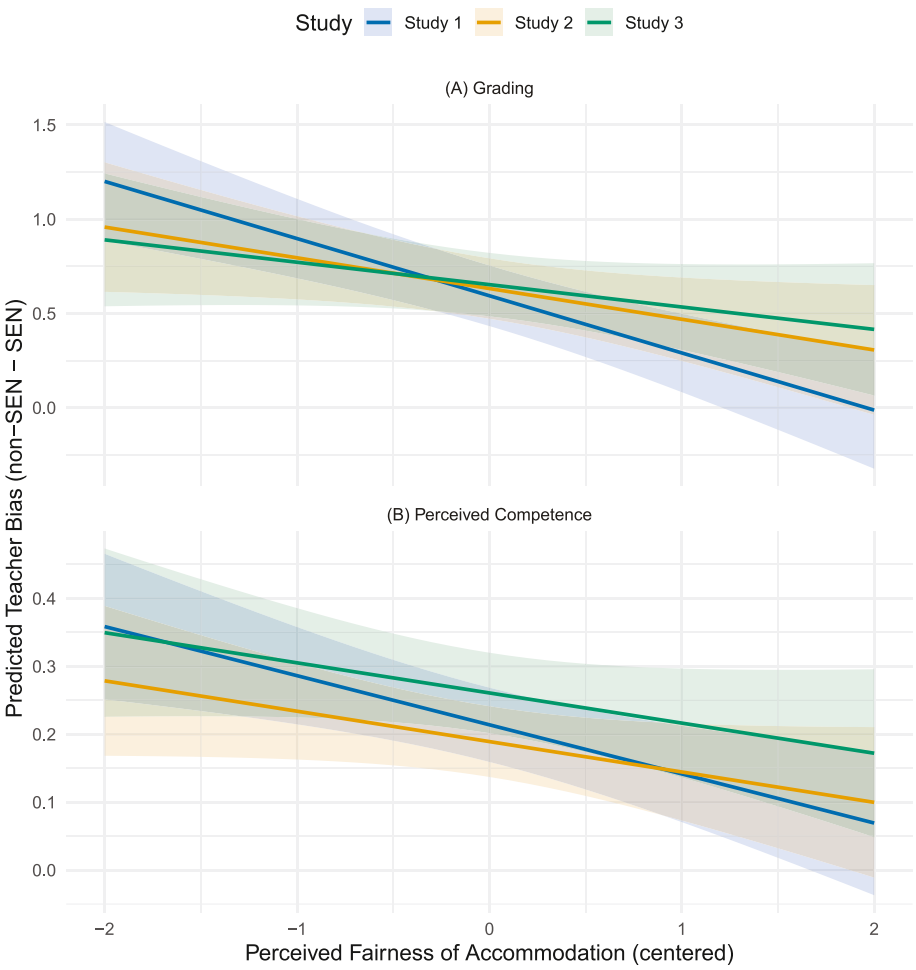


Fig. 1. Predicted teacher bias by perceived fairness across Studies 1–3 (pre-service and in-service teachers).
Note. Lines depict model-implied difference scores (non-SEN − SEN; higher values indicate greater devaluation of students with SEN) with shaded 95% confidence intervals, estimated separately for each study and outcome (Panel A: Grading; Panel B: Perceived Competence). Predictions were pooled across 20 imputations using Rubin's rules. For Studies 1–2, predictions are evaluated at Threat Level = 0 and Student Gender = 0 (the average of −0.5/+0.5 coding); for Study 3, predictions are evaluated at Contrast 1 = 0, Contrast 2 = 0, and Contrast 3 = 0 (reference profile). Perceived fairness was mean-centered; thus $x = 0$ corresponds to the sample mean. These difference scores operationalize teacher bias (backlash effect), such that more negative slopes indicate less bias/backlash at higher fairness. The fairness–bias association was significant for grading and perceived competence in Study 1, significant for grading only in Study 2, and nonsignificant in Study 3 (direction consistent across studies).

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 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, with more classroom experience and exposure to institutional norms and inclusion practices, show similar or divergent patterns of bias and moderation.

3. Study 2

The goal of Study 2 was to replicate Study 1 with in-service teachers to examine whether the effects observed among pre-service teachers generalize to experienced educators. Methodology, hypotheses, and experimental design were identical to Study 1, with the only difference being the participant sample.

3.1. Method

3.1.1. Participants and procedure

Participants. Data were collected from in-service teachers across France via LimeSurvey in autumn 2024. To ensure geographic diversity and address low response rates, 18,000 preschool and primary school principals were randomly invited to participate (if they taught) or 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 recipients is unknown, the 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 ($\eta_p^2 = .040$; based on Stanczak, Aelenei, et al., 2024) with 80% power at $\alpha = .05$. The final sample ($N = 387$, post-imputation) exceeded this threshold. A sensitivity power analysis confirmed sufficient power for small effects ($\eta_p^2 = .028$ for main/interaction effects; $\eta_p^2 = .020$ for moderation models). Table 1 presents demographics.

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, and all study protocols mirrored

Study 1.

3.1.2. 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.

3.2. Results

3.2.1. Analysis strategy

Analyses followed Study 1. Multiple imputation was performed via mice, and results pooled using mitools. CFA revalidated the Social Judgement Scales, yielding similar fit and 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 and sensitivity analyses followed Study 1.

3.2.2. Primary analyses

Grading (Math Tests). Results replicated the backlash effect: students with SEN received lower gradings than their non-SEN peers ($b = 0.63$, 95% CI [0.47, 0.79], $SE = 0.08$, $t(380) = 7.75$, $p < .001$, $\eta_p^2 = .136$). As in Study 1, this effect was unaffected by student gender, threat level, or their interaction (all $ps > .225$).

Perceived Competence. Similarly, students with SEN were judged as less competent ($b = 0.19$, 95% CI [0.14, 0.24], $SE = 0.03$, $t(379) = 7.10$, $p < .001$, $\eta_p^2 = .117$), with no effects of student gender, threat level, or their interaction (all $ps > .255$).

Findings mirrored Study 1 (Table 3). Controlling for participants' gender or marginalized group status did not alter findings (Table S2 in SM).

3.2.3. Moderation analyses

Fairness of Accommodations. Higher fairness perceptions predicted again smaller grading differences ($b = -0.16$, 95% CI [−0.32, −0.01], $SE = 0.08$, $t(353) = -2.10$, $p = .036$, $\eta_p^2 = .012$), but not competence differences ($p = .073$), although the descriptive pattern was consistent (Fig. 1). As in Study 1, fairness did not interact with student gender, threat level, or their combination (all $ps > .073$; Table S3 in SM).

3.2.4. 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 predicted higher grading ($b = 0.24$, 95% CI [0.08, 0.40], $SE = 0.08$, $t(358) = 3.02$, $p = .003$, $\eta_p^2 = .024$) and greater perceived competence ($b = 0.16$, 95% CI [0.09, 0.24], $SE = 0.04$, $t(338) = 4.31$, $p < .001$, $\eta_p^2 = .049$). For non-SEN students, fairness was more weakly associated with perceived competence ($b = 0.12$, 95% CI [0.05, 0.19], $SE = 0.04$, $t(343) = 3.27$, $p = .001$, $\eta_p^2 = .029$), but not with grading ($p = .183$; Table S4 in SM).

3.3. Discussion

The goal of Study 2 was to replicate Study 1's findings among in-service teachers with more classroom experience. Students with SEN were again devalued in both grading and competence ratings—a robust backlash effect (H1a) 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.

As in Study 1, fairness perceptions were 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 showed that fairness perceptions influenced grading for students with SEN but not for their non-SEN peers. While fairness 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 more strongly influenced competence ratings for students with SEN, mirroring Study 1. This underscores that fairness concerns shape how teachers evaluate accommodated students, suggesting they may rely on fairness heuristics in both grading and competence judgments.

Together, these findings highlight that SEN status—and how fair accommodations are perceived—constitutes the central axis of bias in teachers' judgments. Study 3 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 conditions under which bias toward students with SEN is amplified or mitigated by gender-related assumptions.

4. Study 3

Study 3 refined the design from Studies 1 and 2 by eliminating threat manipulations (which yielded no effects) and introducing a gender contrast manipulation. While prior studies manipulated student gender by evaluating same-gender student pairs, Study 3 focused on gender contrast (same-gender vs. cross-gender pairs) to assess whether backlash is 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 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.

4.1. Method

4.1.1. Participants and procedure

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

Given the anticipated small effects in gender-related biases in

education, the preregistered power analysis (G*Power 3.1.9.4) for the planned contrast in H1 targeted effect sizes of η_p^2 = .020–.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. A sensitivity power analysis confirmed sufficient power for small effects (η_p^2 = .021 for main effects; η_p^2 = .029 for moderation models). Table 1 presents demographics.

Procedure. A 2 × 2 × 2 mixed-method 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. Materials and measures matched Studies 1 and 2, except that the threat manipulation was removed, retaining only the low-threat condition (same error rate on easy and difficult items). Gender contrast examined whether backlash effects are amplified when a female student with SEN is compared to a male student without SEN. Student names were counterbalanced. To heighten gender salience, instructions included more gendered pronouns and reframed the accommodation description to emphasize student agency (e.g., "the student completed only the first two exercises" vs. "the teacher instructed the student to complete…").

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

4.1.2. 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.

4.2. Results

4.2.1. Analysis strategy

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), we specified planned orthogonal contrasts (Brauer & McClelland, 2005). The primary contrast (Contrast 1) compared the CGMF condition (boy without SEN vs. girl with SEN), 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 remaining variance: Contrast 2 contrasted the CGFM condition (girl without SEN vs. boy with SEN; coded +2) against 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 moderation (H2), perceived fairness 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 before, with sensitivity analyses confirming robustness.

4.2.2. Primary analyses

Grading (Math Tests). Results confirmed the backlash effect: students with SEN received lower grades than their non-SEN peers (b = 0.65, 95% CI [0.48, 0.82], SE = 0.09, $t(358)$ = 7.54, p < .001, η_p^2 = .136). However, Contrast 1 (CGMF vs. other conditions) was not significant (p = .988), indicating the grading gap did not vary by condition.

Perceived Competence. Similarly, students with SEN were rated less competent (b = 0.26, 95% CI [0.20, 0.32], SE = 0.03, $t(352)$ = 8.62, p < .001, η_p^2 = .173). Again, Contrast 1 (CGMF vs. other conditions) was not significant (p = .107), suggesting competence judgments were similarly unaffected by gender contrast manipulations.

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

4.2.3. Moderation analyses

Fairness of Accommodations. Fairness perceptions were not significantly associated with overall grading (p = .133) or competence differences (p = .110), although the descriptive pattern matched Studies 1 and 2 (Fig. 1). No significant interactions with Contrast 1 emerged for either grading (p = .907) or competence (p = .319; Table S3 in SM).[6]

4.2.4. 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 predicted higher grading ratings (b = 0.18, 95% CI [0.02, 0.33], SE = 0.08, $t(342)$ = 2.28, p = .023, η_p^2 = .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 = .034). For students without SEN, fairness perceptions predicted perceived competence (b = 0.08, 95% CI [0.02, 0.15], SE = 0.03, $t(340)$ = 2.48, p = .014, η_p^2 = .018), but not grading (p = .315; Table S4 in SM).

4.3. Discussion

Study 3 refined our design by removing threat manipulations (no effect in earlier studies) and introducing gender contrast (same-gender vs. cross-gender). We hypothesized that the cross-gender male–female (CGMF) pairing (boy without SEN vs. girl with SEN) would elicit the strongest backlash (H1), and fairness perceptions would moderate this effect (H2).

Consistent with Studies 1 and 2, students with SEN were again devalued in grading and competence, confirming a robust backlash effect across conditions. Contrary to H1, however, the CGMF condition did not elicit a stronger backlash than 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.

Exploratory analyses indicated that fairness perceptions more strongly influenced evaluations of students with SEN than those without, echoing the pattern 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 gender pairings. Fairness perceptions remain critical 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 gender does not intensify bias against students with SEN in teacher judgments.

5. General discussion

We conducted three experimental studies to examine whether teachers devalue students with SEN who receive 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

[6] A significant interaction emerged for residual 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, η_p^2 = .012). When fairness perceptions were low, boys with SEN received lower grades; when fairness was high, this gap decreased. This contrast partitioned residual variance; no hypothesis was stated, and the result is reported for completeness.

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 predictions, neither student gender (Studies 1–3) nor performance-based threat (Studies 1 and 2) intensified this bias. Even when 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: teachers who viewed the "half-exercise" as fair penalized students with SEN less, thereby mitigating—but never eliminating—the backlash effect.

Together, these findings suggest that, within this very context, SEN status with accommodations can drive teacher evaluations, while gender and performance-based threat play minor roles. Notably, fairness perceptions consistently moderate SEN evaluations, indicating that teachers' acceptance (or rejection) of accommodations can soften the penalty otherwise directed at students with SEN.

5.1. Theoretical implications

Our findings show 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). A reduced workload accommodation (half-exercise) probably led teachers to attribute success more to external help than to 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 immediate fairness judgments about accommodations more directly shape 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). Even teachers who support inclusion may penalize students with SEN if they feel the accommodation violates their standard of equity. By highlighting the influence of situational fairness judgments, our findings refine claims that meritocratic ideals can become ideological barriers to inclusion (Darnon, Smeding, & Redersdorff, 2018; Stanczak, Jury, et al., 2024). Although we preregistered a moderating role of meritocratic beliefs, they did not consistently predict evaluations (see SM). Thus, while meritocratic ideology may form the broader cultural framework, backlash effects are more immediately shaped by how justifiable a given accommodation appears in the classroom.

Another 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. This supports research suggesting that disability-related biases 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. Backlash thus stems primarily from perceived threats to "meritocratic" fairness than from gender stereotypes. Our findings refine theory by showing that biases against students with SEN can generalize across gender, emphasizing fairness perceptions as the core mechanism behind SEN-related backlash.

5.2. 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 evaluating fourth-grade mathematics performance. This extends prior findings in literacy (Stanczak, Aelenei, et al., 2024), but limits generalizability to other stages and countries with different teacher training systems. 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 international samples with diverse inclusive education policies.

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. 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. We also used a fixed vignette order (non-SEN first) to mirror classroom norms and facilitate comparison, which, however, may have introduced primacy/anchoring effects. Further, classroom judgments involve more complex dynamics—such as ongoing relationships, time pressure, and accountability—than our vignette could capture. A related limitation concerns the fairness moderator, which relied on a single item referring to the "half-exercise" accommodation. This preregistered choice prioritized conceptual clarity, and our exploratory factor analysis (see SM) indicated insufficient internal consistency for a multi-item scale. Still, fairness perceptions vary across accommodations and may elicit distinct bias patterns (Jury, Stanzcak, et al., 2025). Accordingly, the moderation should be interpreted within this specific context. Field/mixed-methods (e.g., anonymized grading of authentic scripts with/without labels; classroom observations; experience-sampling) could better test ecological generalizability. Future research should counterbalance vignette order, employ multi-item fairness measures across different accommodations, and use longitudinal or naturalistic designs to assess whether bias fades with familiarity or persists over time.

The specificity of the SEN diagnosis and accommodation also limits generalizability. We focused on a student labeled with ADHD, a common yet often stereotyped neurodevelopmental disorder. Our focus, however, is needs-based: the proposed backlash and fairness mechanisms concern responses to accommodations rather than diagnostic labels. That said, SEN encompasses diverse conditions—from learning and intellectual disabilities to physical and sensory impairments—each potentially eliciting distinct biases, competence perceptions (Krämer & Zimmermann, 2025; Krischler & Pit-ten Cate, 2019; Schell et al., 2024), and broader social perceptions (Rohmer & Louvet, 2011). Some accommodations (e.g., assistive technology) are also 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 SEN profiles or accommodations (e.g., assistive technology, sensory or behavioral supports) across diverse school contexts. 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 unfair treatment of a student with SEN is undesirable. Future research should incorporate implicit or behavioral indicators, such as whether teachers assign fewer challenges or offer less encouragement to accommodated students who excel.

5.3. Implications for social policy

Our results echo global calls for improved teacher preparedness and fairness in evaluating students with SEN (Brussino, 2020; OECD, 2019; UNESCO, 2020) and align with concerns raised across education systems such as those in Africa, Asia, Europe, Australia, and North America (Chisala, 2025; Jang & Wong, 2025; Marsili, 2024; Page et al., 2024; Scott et al., 2014). Mandating accommodations alone is insufficient for inclusion—teachers' fairness perceptions could 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 clarify the purpose of accommodations and address fairness concerns. Workshops using real cases can demonstrate how accommodations fit within standardized evaluation. Reframing accommodations as parallel to commonplace adjustments (e.g., eyeglasses, hearing aids) can help shift perspectives. Empirical evidence, such as findings showing that extra time enables students to demonstrate actual knowledge without inflating grades (Sireci et al., 2005; Vidal Rodeiro & Macinska, 2022), can reinforce this understanding. For instance, Chile's *Programa de Integración Escolar* links accommodations to collaborative teacher planning and ongoing professional support, embedding fairness into everyday practice (Guthrie et al., 2019).

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 corrections for structural barriers rather than as special advantages may decrease skepticism. Standardized institutional messaging can ensure consistency in how accommodations are framed. For instance, France's *Pour une école inclusive* reform established local support units and mandatory training to clarify the purpose of accommodations (MENJ, 2019).

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 the policy level, embedding fairness discussions and bias training into teacher certification and evaluation standards can promote inclusive practices. Finland's teacher education, for instance, integrates inclusive pedagogy and reflective training, fostering attention to fairness and equity in professional practice from the outset (Brussino, 2021).

Taken together, although drawn from specific national contexts, these examples illustrate transferable mechanisms—collaborative planning, transparent fairness communication, and reflective professional learning—that can be applied more broadly to foster inclusive assessment practices (see also OECD, 2023a; Wakeman et al., 2022).

5.4. 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 in a controlled setting. This backlash occurred regardless of student gender. 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 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 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.

CRediT authorship contribution statement

Fabian Müller: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cristina Aelenei:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Mickaël Jury:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

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. No AI-generated content was included in the scientific reporting, interpretation, or conclusions.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mickaël Jury reports financial support was provided by Clermont Auvergne Métropole. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tate.2025.105304.

Data availability

The datasets and materials supporting this study are openly available on the Open Science Framework at https://osf.io/ckgh7 (Müller et al., 2025b), under the project title "When accommodations are not enough: Preregistrations, open materials, data, and analysis scripts".

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*(4), 351–371. https://doi.org/10.1177/1094428114547952

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. http://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 University 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 in data 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 & 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 No. 227). https://doi.org/10.1787/600fbad5-en.

Brussino, O. (2021). *Building capacity for inclusive teaching: Policies and practices to prepare all teachers for diversity and inclusion* (OECD Education Working Papers No. 256). https://doi.org/10.1787/57fe6a38-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

Chisala, M. (2025). Exploring inclusivity of standardized assessments for learners with disabilities at lower primary special schools in Zambia. *International Journal of Research and Innovation in Social Science, IX*(XIV), 652–669. https://doi.org/10.47772/IJRISS.2025.914MG0051

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.)*. L. Erlbaum Associates.

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.)* (2nd ed.). 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

Guez, A., Ketan, & Piacentini, M. (2024). *Mapping study for the integration of accommodations for students with special needs (SEN) in PISA* (OECD Education Working Papers No. 308). https://doi.org/10.1787/ed03c717-en.

Guthrie, C., Andersson, H., Cerna, L., & Borgonovi, F. (2019). *Strength through diversity: Country spotlight report for Chile* (OECD Education Working Papers No. 210). https://doi.org/10.1787/058bc849-en.

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

Institut national de la statistique et des études économiques [INSEE]. (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

Jang, H., & Wong, M. E. (2025). Never the twain shall meet? Considering an inclusive meritocracy in Singapore. *Asia Pacific Journal of Education, 45*(1), 314–332. https://doi.org/10.1080/02188791.2024.2416501

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), Article 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., Müller, F., & Aelenei, C. (2025, July). *Experimental evidence regarding backlash effects against students with SEN in inclusive education* [Conference presentation]. London, UK: 1st London International Conference on Inclusive Education (LICIE).

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*, Article 655356. https://doi.org/10.3389/feduc.2021.655356

Jury, M., Stanzcak, A., Huron, C., Müller, F., Aelenei, C., & Sireci, S. (2025). *Perceived fairness of exam accommodations for students with special educational needs* [Preprint] PsyArXiv. https://doi.org/10.31219/osf.io/fcnz4_v1.

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*, Article 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

Krämer, S., & Zimmermann, F. (2025). Teachers' perceptions of students with different disabilities through the lens of the stereotype content model. *Social Psychology of Education, 28*(1), 82. https://doi.org/10.1007/s11218-025-10046-4

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

Krolak-Schwerdt, S., Hörstermann, T., Glock, S., & Böhmer, I. (2018). Teachers' assessments of students' achievements: The ecological validity of studies using case vignettes. *The Journal of Experimental Education, 86*(4), 515–529. https://doi.org/10.1080/00220973.2017.1370686

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* (Version 2.4) [R package]. The Comprehensive R Archive Network (CRAN). https://doi.org/10.32614/CRAN.package.mitools.

Marsili, F. (2024). La gifted education tra meritocrazia e inclusione: Tensioni paradigmatiche e implicazioni pedagogiche [Gifted education between meritocracy and inclusion: Paradigmatic tensions and pedagogical implications]. *Ital. j. spec. educ. incl., 12*(1), 233–243. https://doi.org/10.7346/sipes-01-2024-22

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, de l'Enseignement supérieur et de la Recherche [MENESR]. (2015). *Bulletin officiel n° 5 du 29 janvier 2015 : Plan d'accompagnement*

personnalisé (PAP) [Official bulletin no. 5 of January 29, 2015: Personalized support plan (PAP)]. https://www.education.gouv.fr/sites/default/files/imported_files/documents/BO5_MEN_29_1_2015_387160.pdf.

Ministère de l'Éducation nationale et de la Jeunesse [MENJ]. (2019). *Bulletin officiel n°23 du 6 juin 2019: Circulaire de rentrée 2019 – École inclusive* [Official bulletin no. 23 of June 6, 2019: 2019 start-of-school-year circular—inclusive education]. https://www.education.gouv.fr/bo/19/Hebdo23/MENE1915816C.htm.

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

Müller, F., Aelenei, C, & Jury, M. (2025a). When accommodations are not enough: A multi-study examination of teacher bias toward students with special educational needs across student gender [Preprint]. PsyArXiv https://doi.org/10.31234/osf.io/fx7jc_v1.

Müller, F., Aelenei, C., & Jury, M. (2025b). *When accommodations are not enough: Preregistrations, open materials, data, and analysis scripts* [Dataset]. OSF. https://osf.io/ckgh7.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *Timss 2019: International results in mathematics and science*. Boston College: TIMSS & PIRLS International Study Center. 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 Publishing. https://doi.org/10.1787/9789264009813-en

OECD. (2019). *TALIS 2018 results (volume I): Teachers and School Leaders as Lifelong learners*. OECD Publishing. https://doi.org/10.1787/1d0bc92a-en

OECD. (2023a). *Equity and inclusion in education: Finding strength through diversity*. OECD Publishing. https://doi.org/10.1787/e9072e21-en

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

Page, A., Barr, M., Rendoth, T., Roche, L., Foggett, J. L., Leonard, C., & Duncan, J. (2024). Making reasonable adjustments for students with disability in Australian mainstream classrooms: A scoping review. *Australas. J. Spec. Educ., 48*(1), 46–63. https://doi.org/10.1017/jsi.2024.1

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 P. Devine, & A. Plant (Eds.), *45. Advances in Experimental Social Psychology* (pp. 167–227). Academic Press. https://doi.org/10.1016/B978-0-12-394286-9.00004-4.

Schaeffer, K. (2023). *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/.

Schell, C. S., Dignath, C., Kleen, H., John, N., & Kunter, M. (2024). Judging a book by its cover? Investigating pre-service teacher's stereotypes towards pupils with special educational needs. *Teaching and Teacher Education, 142*, Article 104526. https://doi.org/10.1016/j.tate.2024.104526

Scott, S., Webber, C. F., Lupart, J. L., Aitken, N., & Scott, D. E. (2014). Fair and equitable assessment practices for all students. *Assessment in Education: Principles, Policy & Practice, 21*(1), 52–70. https://doi.org/10.1080/0969594X.2013.776943

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, 27*, 687–708. 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

UNESCO. (2020). *Global Education Monitoring Report 2020: Inclusion and education: All means all.*. https://doi.org/10.54676/JJNK6989.

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

Wakeman, S. Y., Thurlow, M., Reyes, E., & Kearns, J. (2022). Fair and equitable grading for ALL students in inclusive settings. *Inclusive Practices, 1*(4), 127–131. https://doi.org/10.1177/27324745211055398

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