Studies in Educational Evaluation 78 (2023) 101284



Contents lists available at ScienceDirect

Studies in Educational Evaluation

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





Discrimination in grading: A scoping review of studies on teachers' discrimination in school

Giulietta Zanga *, Elena De Gioannis

Università degli studi di Milano (Department of Social and Political Sciences), Italy

ARTICLE INFO

Keywords:
Discrimination
Equal education
Minority group students
Teacher attitudes

ABSTRACT

Studies on teachers' grading suggest that school grades depend not only on students' performance but also on teachers' bias toward specific social categories. Numerous studies tested the existence of discrimination in grading using different strategies and focusing on multiple students' characteristics. This study aims to summarise those studies by identifying (1) the methodologies used, (2) the characteristics on which discrimination is based and (3) the empirical results. We conducted a scoping review where studies were selected blindly by the two authors. The initial search was conducted on ERIC, Education Database and PsycInfo and 37 studies were finally identified. A comparison among the included studies suggests that the main strategies used are experiments and regression analysis on the difference between blind and non-blind scores, while gender, race/ethnicity and migration background are the most frequently tested characteristics. Finally, on average studies confirmed the presence of discrimination in grading still with some exceptions and, sometimes, under specific conditions. To conclude, it is challenging to test teachers' discrimination through grading and to date the methodologies used have some limitations. However, on average, empirical evidence suggests that school grades are affected by teachers' bias.

1. Introduction

Studies on educational inequalities have long emphasised that school contributes to social stratification (Ballarino & Panichella, 2021). In fulfilling their selection function, schools do not sort students into different educational tracks – and thus careers – based solely on meritocracy but also on students' characteristics, e.g. social origin.

Literature about educational discrimination has focused on several issues, e.g. teachers' expectations regarding the school outcomes of students with different ascribed characteristics (Lorenz, 2021; Parks & Kennedy, 2007; Tobisch & Dresel, 2017), teachers' track advices (Barg, 2013; Barone et al., 2018; Bruneau et al., 2020), perceived discrimination by students regarding, among others, punishments given by teachers (Butler-Barnes & Inniss-Thompson, 2020) and teachers' attitude towards students (Auwarter & Aruguete, 2008; Joseph et al., 2016).

Among the identified processes through which the school and its agents act as reproducers of inequalities is teachers' discrimination in the attribution of grades (Argentin & Pavolini, 2020; Dee, 2005). Studies on teachers' grading have, in fact, highlighted that school grades depend not only on students' performance but can also be influenced by teachers' bias toward specific social categories, such as ethnic groups or gender (Malouff & Thorsteinsson, 2016).

This tendency may have several detrimental consequences, both at the individual and the collective level. On the one hand, teachers' discrimination affects students' educational perspectives which are highly correlated with academic performance and their motivation to study (OECD, 2012). On the other hand, it also fosters inequalities among social groups (Argentin & Pavolini, 2020).

Given the importance of this issue for academic research in education, a comprehensive summary of the existing studies on teachers' discrimination through grading could be useful to inform and guide future research on this theme. This scoping review aims to fulfil this objective using a systematic approach.

2. Background

There is not a unique and univocal definition of discrimination, rather several definitions are proposed both by sociologists and psychologists, which can, in some cases, also be in conflict.

Here we adopt the definition proposed by Wenz (2020, p. 56) of

* Correspondence to: Via Conservatorio 7, 20122 Milan, Italy.
E-mail address: giulietta.zanga@unimi.it (G. Zanga).

https://doi.org/10.1016/j.stueduc.2023.101284
Received 14 February 2023; Received in revised form 6 May 2023; Accepted 21 June 2023
Available online 27 June 2023
0191-491X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

discrimination as "the causal effect of information about or a signal sent out by an individual on how this individual is treated by another individual so that any human decision that distinguishes between individuals or groups of individuals constitutes discrimination". Discrimination is thus the act of treating two – otherwise identical individuals – differently based on any attribute, behaviour, or characteristic that allows one to distinguish them.

Economic theories of discrimination usually distinguish between two types of discrimination (Guryan & Charles, 2013), i.e. taste-based discrimination based on agents' preferences and statistical discrimination based on noisy signals about others' abilities. In the context of teachers and the grading process, taste-based discrimination occurs if teachers attribute higher grades to students possessing characteristics they have a personal preference for. For instance, if a teacher has a personal preference toward White people compared to Black people, he/she would give higher grades to a White student – and conversely a lower grade to a Black one – regardless of the student's actual performance. We would have statistical discrimination if the teacher would base his/her judgement of the student's ability not only on his/her performance but also on the teacher's opinion about the average ability of the student's social group, e.g. White students are on average better than Black students in school.

As regards discrimination in the school context, previous studies provided evidence of the detrimental consequences that discrimination has on students. Perceived teachers' discrimination was found to be positively associated with suffering from depression, anger, suicidal thoughts, anxiety and behaviour disorders (Jelsma et al., 2022; Jiang & Dong, 2020; Lopez et al., 2016; Weeks & Sullivan, 2019) and negatively associated with scholastic behaviour, academic outcomes, perceived scholastic utility and school bonding (Assari & Caldwell, 2018; Benner & Graham, 2011; Butler-Barnes & Inniss-Thompson, 2020; D'hondt et al., 2016; Kyere et al., 2022; Thomas et al., 2009; Wang & Huguley, 2012).

Teachers' discrimination has, therefore, a huge impact on students' personal life, academic experience and future, by affecting their scholastic outcomes and their propensity to consider studying useful. If the target of discrimination is those students who are affected the most by educational inequalities and whose scholastic paths are already at risk, the existence of teacher bias is detrimental not just per se but also because it contributes to the starting condition of the educational disadvantage of specific social groups (Davis & Dollard, 1940; Willis, 2006). In the case of grading, for instance, parents with less knowledge about the functioning of the education system trust teachers' judgments more and, concurrently, have less economic and social resources to overcome possible bad scholastic results (Argentin & Pavolini, 2020).

While a few reviews on the theme already exist, none of them systematically and precisely summarise the literature on the theme. Farkas (2003) reviewed studies on racial discrepancies in education and, similarly, Ferguson (2003) reviewed studies explaining the Black-White test score gap. While both reviews include also studies on teachers' grading, they only focus on discrimination based on race. Two meta-analyses (Machts et al., 2016; Südkamp et al., 2012) collect studies measuring the correspondence between teachers' judgements of students' academic achievement and their actual achievement. However, as highlighted by the authors themselves, teachers' judgement accuracy depends on many factors including those that are not related to the object of perception, i.e. the characteristics of teachers, students, judgement and tests (Machts et al., 2016), and thus includes – but it is not limited to – discrimination. Finally, Malouff and Thorsteinsson (2016) conducted a meta-analysis of experimental studies measuring bias in grading. However, they only included studies adopting a specific methodological approach, i.e. experiments in which graders are exposed to information about the student other than his/her performance.

Differently from previous reviews on the theme, here we considered also discrimination based on students' characteristics other than race and we did not impose a constraint on the type of methodology used to measure discrimination. Conversely, one of the aims of the review is to summarise the methods that are currently used to test discrimination through grading.

Table 1
Inclusion and exclusion criteria.

Characteristic	Inclusion criteria	Exclusion criteria
Participants	- Teachers from primary to upper secondary school - Students from primary to upper secondary school	- University teachers, pre-service teacher - BA or MSc students, or above
Concept	- Assessment of the presence of discrimination through students' grading, i.e. evaluation of performance - Only quantitative studies (ok mixed method)	- Other forms of teachers' discrimination (e.g. expectations, interpersonal relationships, behaviour) - No qualitative studies
Context	Language: English, Italian or Spanish	Language: others
Types of studies	Peer-reviewed studies	No grey literature; no review

2.1. Aim of the study

The study aims to summarise and discuss studies on teachers' discrimination through students' grading. More specifically, the review aims to answer three research questions:

1. Which methodologies are used to test the existence of teachers' discrimination in students' grading?
2. What are the students' characteristics that lead to discrimination?
3. What are the results of the studies, i.e. is the existence of teachers' discrimination confirmed empirically? If so, for which students' characteristics?

2.2. Methods

We followed the guidelines of the JBI Manual for Evidence Synthesis (Peters et al., 2020) to select the studies eligible for being included in this review. More specifically, we, first, developed an a priori research protocol, including the research questions, the inclusion and exclusion criteria on which the decision to include the studies is based and the search strategy, i.e. the search query and the databases where the search query is executed.

Table 1 summarises the inclusion and exclusion criteria established in the first stage of the review. We decided to exclude higher education because of its several differences with lower educational levels, especially in terms of the student-teacher relationship (Hagenauer & Volet, 2014; Snijders et al., 2022) and grading procedure (Lipnevich et al., 2020). Furthermore, since the aim of the review is to summarise studies on discrimination through grading we excluded those investigating other forms of discrimination, e.g. teachers' expectations and attitudes toward students based on specific characteristics (Robinson-Cimpian et al., 2014; Tobisch & Dresel, 2017). For the same reason, studies including teachers' judgments about whether students' ability in a certain discipline is above or below their peers' average (e.g. Campbell, 2015; Kleen & Glock, 2018) were excluded, as they are a generic evaluation of their competences rather than a judgment on performance. As regards the methodology and the type of studies, we excluded those published in journals that are not peer-reviewed, i.e. grey literature and those adopting a qualitative methodology. For what concerns the latter, we decided to exclude studies using this kind of methodology to facilitate the comparison between the studies included in the review.

Both authors conducted the studies' selection independently from October to December 2022 on three databases, i.e. ERIC, Education Database and PsycInfo and then through citation searching. More specifically, the selection occurred in three steps, i.e. the authors first read the titles of all the studies obtained in the original search, they then read the abstract of those whose titles seemed to meet the inclusion criteria

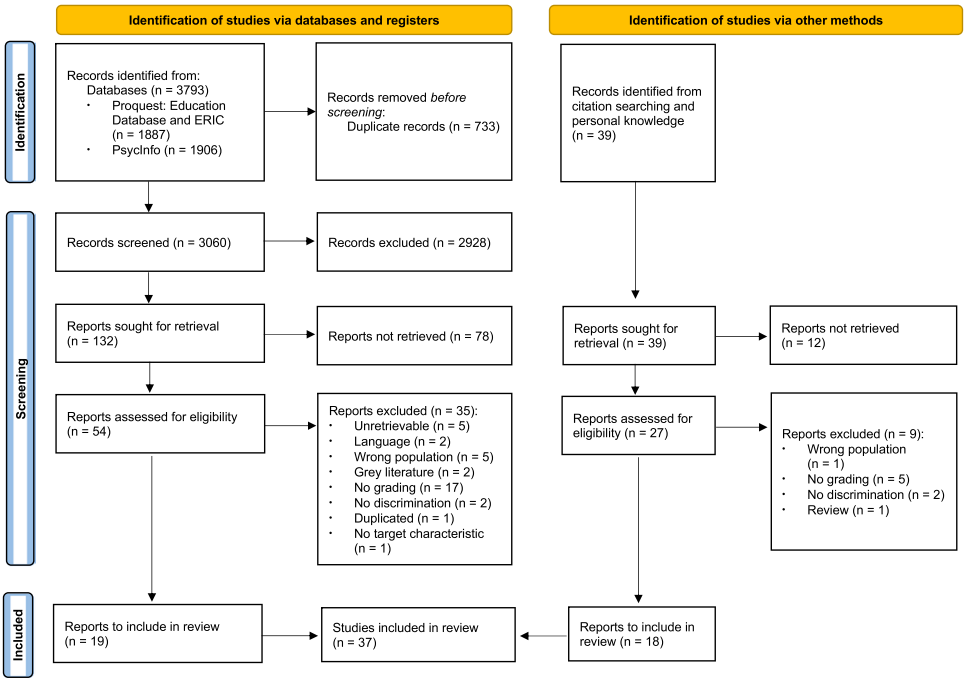


Fig. 1. PRISMA flow chart.

Table 2
Methodology.

Methodology	Number of studies	%
Experiment	19	51%
Regression	14	38%
Natural experiment	2	5%
Correlation	2	5%

and, finally, they read the full text of those whose abstract met the inclusion criteria. The search query used was the following: (bias* OR discrimin* OR prejudice* OR stereotyp*) AND (teacher* OR professor* OR educator* OR instructor*) AND (grad* OR performance* OR scor* OR mark OR rating*) AND (school* OR academic). Conflicts in the two reviewers' choices were discussed at the end of each step of the procedure and resolved by consensus.

3. Results

Fig. 1 reports the JBI's flow chart of the studies' selection procedure. The search query executed in the three above-mentioned databases returned 3793 records, of which 733 were duplicates. Of the 3060 studies resulting from the initial search, we selected 132 records when reading the titles and then, 54 records after reading their abstracts. The full-text revision suggested that 19 of these studies were eligible to be included in the review. We repeated the same procedure on 39 records identified from citation searching and authors' personal knowledge. Of these, 27 passed the abstract revision and 18 resulted eligible after the full-text revision. A total number of 37 studies was, thus, included in the review (see Table A1 in the Appendix).

3.1. Methodology

Table 2 summarises the methodologies used by the included studies, i.e. experiments, natural experiments, regression analysis and correlation analysis.

The experimental strategy is the most frequently adopted among the four (51% of studies). Experiments are conducted mainly in the same way in all studies, i.e. a sample of teachers is asked to evaluate the performance of a student, frequently a writing assignment such as an essay, but in some cases, teachers were asked to hear an oral answer to a set of questions.

The manipulation occurs in the type of information that teachers have on the student and discrimination exists if the information on students' characteristics influences the evaluation given by teachers. Two studies stand out, as they tested the existence of discrimination using a field experiment (Hinnerich et al., 2011, 2015). Both were conducted in Sweden, where compulsory national tests during high school are given in three core subjects. The performance in the test was evaluated by the students' own teacher (non-blind score), but researchers obtained a blind grade by first writing all tests on a word processor and removing the students' identities. These anonymized tests were then regraded blindly by hired teachers. Finally, in two studies (Kaiser et al., 2017; Kramer & Zimmermann, 2021), researchers adopted the simulated classroom paradigm, i.e. a computer simulation of an instructional situation in which participants take over the role of teachers and interact in a virtual classroom with simulated students, whose characteristics, engagement and achievement can be experimentally manipulated.

Two studies applied a strategy that could be defined as a natural experiment. In one case (Lavy, 2008), researchers exploited the Israeli high-school matriculation exam system in which the final score in a given subject is the mean of two intermediate scores. One is anonymized and scored by an independent agency (blind score) while the other is scored by the student's teacher (non-blind score). What is interesting for researchers is that the two intermediate exams – despite not being exactly the same – draw their questions from the same bank of questions and are performed in the same environment and under the same conditions. Both facts go "some way to convince us that the exams are testing the same skills and cognitive achievements" (Lavy, 2008, p. 2086). Similarly, in another study (Falch & Naper, 2013) researchers exploited the fact that in Norway, at the end of lower secondary education, students are evaluated both by their teachers and blindly in central exit exams.

In 38% of studies, researchers matched students' school grades (non-blind score) with their performance in a national, e.g. Pruebas de

Table 3
Students' characteristics.

Characteristic	Number of studies	%
Gender	21	57%
Race/ethnicity	15	40%
Migration background	5	14%
Obesity	2	5%
SES	2	5%
Special education	2	5%
Caste	1	3%
Family type	1	3%
Physical attractiveness	1	3%

Table 4
Results.

Characteristic	Result		
	Yes	No	Mixed
Gender	52%	24%	24%
Race/Ethnicity	40%	33%	27%
Migration background	80%	0%	20%
Obesity	100%	0%	0%
SES	100%	0%	0%
Special education	50%	50%	0%
Caste	100%	0%	0%
Family type	0%	100%	0%
Physical attractiveness	0%	100%	0%

evaluación diagnóstica in Spain (Marcenaro-Gutierrez & Vignoles, 2015), or international standardized test (blind score), e.g. the Programme for International Student Assessment (PISA) and the Progress in International Reading Literacy Study (PIRLS) (e.g. Kiss, 2013; Mateju & Smith, 2015; Triventi, 2020). See Table A2 in the Appendix for a complete list. Discrimination is then verified either by regressing students' characteristics on the difference between the blind and non-blind score or by regressing students' characteristics on the non-blind score controlling for the blind score.

Finally, two studies (Carter, 1971; MacCann & Roberts, 2013) used Pearson correlation coefficients and mean differences between school grades and students' characteristics.

3.2. Students' characteristics

Table 3 summarises students' characteristics on which the included studies verified the existence of teachers' discrimination in grading. We identified eight characteristics, i.e. gender, race/ethnicity, migration background, obesity, socioeconomic status (SES), need for special education, caste, family type (parents are divorced vs parents are not divorced) and physical attractiveness.

Gender is the most frequent among the above-mentioned characteristics (57% of studies), followed by race/ethnicity (40% of studies) and migration background (14% of studies). The other characteristics are investigated only in a few studies.

While in most studies the characteristic researchers are interested in is only one, in 32% of studies more than one characteristic is investigated. This is interesting because it allows testing potential interactions between the above-mentioned characteristics. As discussed in the next section, empirical results on main and interaction effects suggest that in some cases the combination of students' characteristics matters.

3.3. Results

Table 4 summarises the main results obtained by the included studies for each characteristic. More specifically, we reported whether empirical results suggested that discrimination existed ("Yes") or that it was not possible to confirm the existence of discrimination ("No"). We classified the result as "Mixed" in those cases in which the existence of discrimination was confirmed only under specific circumstances, e.g. interaction with other factors, teachers' characteristics or subject on which students are evaluated.

In the case of gender, most studies (52%) confirmed the existence of discrimination in grading based on this characteristic. On average, teachers' assessment is higher for female students compared to what is measured by blind test scores and boys tend to receive lower grades than girls in school, even controlling for their performance in standard tests. However, a few studies found discrimination in favour of male students. In an experiment where 80 teachers were asked to evaluate students' performance in a science test, Spear (1984) found that male students received higher grades compared to female students. In a similar experiment conducted with 96 teachers, Kehle et al. (1974) found that male students received higher grades than female students, but only among students who were depicted as unattractive. Similarly, an experiment conducted in Israel (Guttmann & Boudo, 1988) on 241 teachers found that male students received higher evaluations, but only when their family was described as intact, e.g. parents were not divorced.

As regards race and ethnicity, 40% of studies could confirm the existence of discrimination in grading, while 27% found that discrimination occurred only under certain circumstances. The remaining studies (33%) were not able to confirm the existence of discrimination. In the case of a positive result, the advantaged group was always the White or majority group with one exception: in a recent experiment conducted in the U.S. on 70 mixed-ethnicity teachers, Sedlacek (2021) found that White teachers tended to give higher evaluations to White male students, while Hispanic teachers to Latina female students.

Also in the case of migration background, the studies that confirmed the existence of discrimination (80%) found that the children of immigrants are evaluated worse compared to students with native parents. Interestingly, in Spain Marcenaro-Gutierrez and Vignoles (2015) found that teachers' assessment of young immigrant students exceeded these students' scores in a blind standardized national test. However, this was confirmed only when the subject evaluated by the test was reading, while it was not confirmed in the case of math.

As regards other students' characteristics, obesity and socioeconomic status were found to play a significant role in teachers' evaluation of their students' performance. Obese students and those with a low socioeconomic status are evaluated worse than healthy-weight students (Dian & Triventi, 2021; MacCann & Roberts, 2013) and those with a higher SES, respectively (Doyle et al., 2022; Westphal et al., 2016). The same is true in India for students belonging to a lower caste (Hanna & Linden, 2012). On the contrary, it was not possible to confirm the existence of discrimination in grading based on either family type (Guttmann & Boudo, 1988) or physical attractiveness (Kehle et al., 1974). Finally, the two studies testing the influence of being labelled as a "student with special education need" found discordant results. In an experiment with 30 teachers in the U.S., Fogel and Nelson (1983) did not find discrimination against students labelled as needing special education. On the contrary, using the simulated classroom paradigm, Kramer and Zimmermann (2021) in Germany found that teachers on average marked more answers as wrong in the case of students with emotional disorders.

4. Discussion

Teachers' bias and discrimination towards students with specific ascribed characteristics is a pivotal issue in the context of educational inequalities. However, given the numerous and heterogeneous contributions to the theme, a comprehensive review of the existing studies on teachers' discrimination through grading could be useful for future research. This study aimed to summarise those contributions using a systematic approach.

Starting from an initial number of 3793 records, we ended up with 37 empirical studies focusing on discrimination in grading perpetuated by

teachers, from kindergarten to high school. These studies' screening allowed us to answer the three research questions of the scoping review, i.e. (1) which methodologies are used to test the existence of teachers' discrimination in students' grading, (2) what are the students' characteristics that lead to discrimination and (3) which are the results of the studies?

As regards the first question, we found mainly two methodologies, i.e. experiments in which teachers were asked to evaluate a real or fake test of a student who was randomly assigned specific characteristics and a regression analysis testing the effect of students' characteristics on school grades controlling for students' performance as measured by standard, blindly evaluated tests. The most frequently tested characteristics were gender and race/ethnicity, while few studies tested other characteristics, such as migration background, socioeconomic status and obesity. Finally, most studies confirmed the existence of discriminatory behaviour in students' grading based on the tested characteristics. However, especially in the case of gender and race/ethnicity, the presence of discrimination was sometimes confirmed only under specific conditions, e.g. interaction with other factors, teachers' characteristics or the subject on which students were evaluated.

As regards students' characteristics, those most frequently analysed were gender, race/ethnicity and migratory background as they are historical and traditional dimensions of inequality and, at the same time, they can be easily manipulated in experiments. It is, on the contrary, more difficult to infer one's SES by his/her name (Doyle et al., 2022). However, we found differences in how researchers defined the categories for race/ethnicity and migration background. As regards migration background, studies usually compared natives with those with a generic migratory background or they focused on specific countries of origin for the foreign group. For what concerns race, studies conducted in the United States often used the categories "White", "Black" and "Latino", while studies conducted in European countries mainly referred to students' country of origin, e.g. Turkish in Germany. These characteristics were also those for which evidence was more contradictory (see Tables A4-A6 in the Appendix).

In the case of experiments, most of them were conducted in a similar way but some specificities in the design vary from one experiment to the other. This is the case of how students' characteristics were made salient for teachers. In some cases, mainly when gender or race was the characteristic of interest, researchers specified the students' names or added their photos on top of the test (e.g. Kehle et al., 1974), while in other cases, teachers were given a report card of the student (e.g. Guttmann & Boudo, 1988) or a brief bio summarising his/her educational characteristics (e.g. Spear, 1984). Differently from the other studies, in an experiment whose aim was to investigate ethnic discrimination, teachers listened to students' oral answers and could infer their ethnicity by their accent (Crowl & MacGinitie, 1974). Another source of heterogeneity was the location in which the experiment was conducted, either a laboratory – in most cases – or online, with teachers answering from their homes.

This heterogeneity in the experimental design should be kept in mind when comparing the studies' results. Differently from the other methodologies, the experimental one is, in fact, characterized by the most heterogeneous results (see Table A3 in the Appendix). This heterogeneity could derive from the variation in the experimental design. For instance, students' names alone do not necessarily provide enough information about students' characteristics, e.g. it is easier to infer students' gender or ethnicity than to infer their socioeconomic status, still, we need to assume that teachers paid attention to the student's name printed on the test and that they associated it with the desired characteristic. Nevertheless, the provision of further information beyond students' names could make teachers suspicious about the true aim of the experiment, thus resulting in social desirability bias. Therefore, one of the limitations of the experimental methodology is that it is difficult to distinguish whether a null result is due to the absence of discrimination or a failure in the manipulation. A possible solution to this problem is the strategy adopted by Quinn (2020) and van Ewijk (2011), where the names of the students and their relatives were included in the essays. In this way, it is unlikely both that teachers do not read students' names and that they understand the true aim of the experiment.

Compared to experiments, a regression analysis on students' characteristics and grades – controlling for standardized tests' performance – can solve the problems related to the size and heterogeneity of the sample. However, this methodology has its limitations, that can put into doubt its efficacy in measuring teachers' bias discrimination. Standardized tests and school grades are different measures of performance (Lindahl, 2007) and, thus, they do not necessarily measure the same abilities. Standardized tests are generally designed to evaluate students' overall abilities on a specific subject or general knowledge. For this reason, they are often designed such that all students could answer the questions, even without previous knowledge of the theme. On the contrary, teachers assess students' knowledge about one (or more) topics of a specific subject. Moreover, standardized tests provide a one-time snapshot of students' knowledge and test-taking skills, while teachers' grades reflect a more holistic picture of students' performance (Rauschenberg, 2014).

Therefore, this type of analysis is reliable if we assume that the tests are unbiased, otherwise we cannot fully rule out the possibility that part of the discrepancy between the blind and non-blind scores is due to the test design (Bonesronning, 2004; Bonesrønning, 2008). In this sense, a better performance in standardized tests could be related to other factors rather than ability. Numerous studies, for instance, found that women and men differ in their attitudes toward competition, with women being more likely to shy away from competitive settings. This has repercussions on the ability of national, standardized tests of measuring abilities. As stated by Niederle and Vesterlund (2010, p. 130), "the competitive pressures associated with test taking may result in performances that do not reflect those of less-competitive settings". Furthermore, standardized tests are not typically part of the students' final grade – sometimes they do not even know their score – and this can influence students' attitudes toward these tests (Falch and Naper, 2013). We must, thus, be cautious to assert that the differences between standardized tests and teachers' grades are entirely attributable to teachers' discrimination. Nevertheless, differently from studies adopting an experimental design, those using a regression analysis report more homogenous results, pointing at the presence of teachers' discrimination (see Table A3 in the Appendix).

In this regard, the studies using a natural experiment as a research design represent an interesting case. In those studies, researchers compared blind and non-blind scores that students received on high school exams. This allows us to solve many of the above-mentioned issues, as the differences between the two assessments are minimized or even eliminated. Obviously, this method is difficult to be applied worldwide, as it depends on the characteristics of countries' educational systems. Still, future research could simulate the conditions of this natural experiment in more artificial and controlled settings.

The scoping review focused on how the research was conducted and what was found, while we did not discuss why discrimination in grading occurs. However, some of the included studies also tested potential explanations, which is worth mentioning. As stated by Petersen (2005, p. 678), in general discrimination is mainly driven by three motives, i.e. prejudice – an irrational dislike for some social groups, stereotypes based on "erroneous information or an incorrect assessment of the situation" and statistical discrimination, that occurs when "groups differ statistically in their distributions on characteristics relevant in a given situation". However, only one of the included studies tested the hypothesis of stereotypes (Alesina et al., 2018) and one that of statistical discrimination (Burgess & Greaves, 2013).

Alesina et al. (2018) found that teachers' endorsement of stereotypes moderated discrimination in grading, i.e. teachers with stronger implicit stereotypes gave lower grades to immigrants compared to natives with the same performance. Burgess and Greaves (2013) found that members

Table A1
Studies included in the review.

Article	Country	Methodology	Characteristic	Results
Marcenaro-Gutierrez & Vignoles, 2015	Spain	Regression	Gender, migration background	Gender: Mixed Migration background: Yes
Hinnerich et al., 2011	Sweden	Experiment	Gender	No
Mateju & Smith, 2015	Czech Republic	Regression	Gender	Yes
Triventi, 2020	Italy	Regression	Migration background	Yes
Kiss, 2013	Germany	Regression	Gender, migration background	Migration background: Mixed Gender: Mixed
Hinnerich et al., 2015	Sweden	Experiment	Migration background	Yes
Hanna & Linden, 2012	India	Experiment	Gender, caste	Caste: yes Gender: no
Sprietsma, 2013	Germany	Experiment	Ethnicity	Yes
Lavy, 2008	Israel	Natural experiment	Gender	Yes
Falch & Naper, 2013	Norway	Natural experiment	Gender	Yes
Shepherd, 2020	USA	Experiment	Gender, race	Gender: No Race: Yes
Shepherd, 2011	USA	Experiment	Race, gender	Race: Yes Gender: Yes
Chase, 1986	USA	Experiment	Gender, race	Gender: Mixed Race: Mixed
Sedlacek, 2021	USA	Experiment	Gender, ethnicity	Gender: mixed Ethnicity: mixed
Quinn, 2020	USA	Experiment	Race	Mixed
Protivínský & Münich, 2018	Czech Republic	Regression	Gender	Yes
Kisfalusi et al., 2021	Hungary	Regression	Gender, ethnicity	Ethnicity: Yes Gender: Yes
Di Liberto et al., 2022	Italy	Regression	Gender	Yes
Carter, 1952	USA	Correlation	Gender	Yes
Bennett et al., 1993	USA	Regression	Gender	No
MacCann & Roberts, 2013	USA	Correlation	Obesity	Yes
Botelho et al., 2015	Brasil	Regression	Race	Yes
Alesina et al., 2018	Italy	Regression	Migration background	Yes
Doyle et al., 2022	UK	Experiment	SES, race	SES: Yes Race: No
van Ewijk, 2011	Netherlands	Experiment	Ethnicity	No
Kramer & Zimmermann, 2021	Germany	Experiment	Special education	Yes
Kehle et al., 1974	USA	Experiment	Gender, race, physical attractiveness	Gender: Yes Race: No Physical attractiveness: No
Copur-Gencturk et al., 2020	USA	Experiment	Gender, race	Gender: No Race: No
Guttmann & Boudo, 1988	Israel	Experiment	Gender, family type	Gender: Mixed Family type: No
Burgess & Greaves, 2013	UK	Regression	Race	Mixed
Spear, 1984	UK	Experiment	Gender	Yes
Bonesrønning, 2008	Norway	Regression	Gender	Yes
Fogel & Nelson, 1983	USA	Experiment	Special education	No
Kaiser et al., 2017	Germany	Experiment	Race	No
Crowl & MacGinitie, 1974	USA	Experiment	Race	Yes
Westphal et al., 2016	Germany	Regression	SES	Yes
Dian & Triventi, 2021	Germany	Regression	Obesity	Yes

of ethnic groups that teachers experienced to perform better in school tended to be on average over assessed than students belonging to ethnic groups which teachers experienced to perform worse. Note that, according to Avitzour et al. (2021) and van den Bergh et al. (2010), implicit stereotypes, more than explicit ones, are correlated with teachers' grading practices. Future research could replicate the design proposed by Alesina et al. (2018) and use instruments such as the Implicit Association Test (Greenwald et al., 1998) to measure teachers' implicit stereotypes.

Another possible explanation that has been widely investigated in the included studies regards "behaviour", as to say the fact that students with different ascribed characteristics show different attitudes towards school and behaviours in class. The hypothesis is that, on equal performance, teachers would reward with higher grades students who behave better. This hypothesis could be especially true for what concerns gender, as girls on average show more discipline than boys (Bennett et al., 1993). The included studies, however, came to discordant conclusions: some found that behaviour is a moderating factor in teachers' assignment of grades, while others could not confirm this relationship. However, teachers' evaluation of students' behaviour could in part be affected by the stereotypes regarding a student's social group. If this is the case, as confirmed in some studies, teachers' judgment about pupils' behaviour could be biased per se. Teachers' stereotypes could therefore partly explain also bias in grades due to differentials in students' behaviour.

Another potential explanation for teachers' discrimination regards their own characteristics, i.e. the fact that teachers could unconsciously favour students of their own gender, ethnicity or social class – a tendency known as in-group favouritism (Everett et al., 2015). Also in this case, empirical evidence is not concordant with this assumption, as some studies found that teachers' characteristics could moderate their grading procedure but not necessarily in favour of students with the same characteristics, while others did not find evidence of this tendency (van Ewijk, 2011). For example, Falch and Naper (2013) found that girls received the highest grades in reading when assessed by male teachers, while they received the highest grades in math when assessed by inexperienced teachers. These heterogeneous results are coherent with what is generally found in studies investigating the hypothesis that teachers

Table A2
Standardized tests and related papers.

Standardized test	Country	Papers
INVALSI	Italy	(Alesina, Carlana, La Ferrara, & Pinotti, 2018; Di Liberto, Casula, & Pau, 2022; Triventi, 2020)
Sistema de Avaliaç ão de Rendimento Escolar do Estado de São Paulo	Brasil	(Botelho et al., 2015)
Panel data from the Norwegian lower secondary school	Norway	(Bonesrønning, 2008)
Key stages tests by English National Curriculum	UK	(Burgess and Greaves, 2013)
SCIO entrance exam to high school	Czech Republic	(Protivínský and Münich, 2018)
National Assessment of Basic Competences	Hungary	(Kisfalusi et al., 2021)
Einstein Assessment of School-Related Skills	International	(Bennett et al., 1993)
German National Educational Panel Study	Germany	(Dian and Triventi, 2021)
Andalusian Educational Authority assessment programme	Spain	(Marcenaro-Gutierrez and Vignoles, 2015)
PIRLS	International	(Kiss, 2013)
PISA	International	(Kiss, 2013; Mateju & Smith, 2015)
TIMSS	International	(Westphal et al., 2016)

favour students of their own gender (e.g. Eble & Hu, 2020; Hoffmann & Oreopoulos, 2009) or race/ethnicity (Dee, 2004; Nguyen, 2018). In short, the assumption about teachers' in-group favouritism has not been confirmed yet, thus future research could shed some light on this mechanism which could explain the presence of positive discrimination, especially for what concerns students' race and ethnicity.

To conclude, it is challenging to test teachers' discrimination through grading and to date the methodologies used have some limitations, still, the natural experiment is a promising strategy. The decision to exclude qualitative studies necessarily means a lack of results that could be useful for better understanding the mechanisms of discrimination and their consequences on students' psychological/aptitude aspects (e.g. Jiang & Dong, 2020; Lorenz, 2021). Furthermore, it is possible that, because of publication bias, studies with negative results have not been published and therefore are not included in this review: some examples of grey literature, of course not representative of the entire field, show however the existence of teachers' discrimination through grading (Terrier, 2020). On average, therefore, empirical evidence on the presence of teachers' discrimination suggests that school grades are affected by teachers' bias, even if with some exceptions and, sometimes, under specific conditions. Furthermore, the literature on the topic of perceived discrimination, which is not inlcuded in this review, underlines the importance of other moderating variables related to students themselves, such as student belonging and cognitive engagement with school (Baysu et al., 2023; Jelsma et al., 2022).

Sharing these results also with schools and teachers could be an effective strategy to improve students' experience and the scholastic environment. As shown by Alesina et al. (2018), teachers who after the study were informed about their discriminatory behaviour increased the grades given to immigrant students. At the same time, school actions specifically intended to overcome discrimination, e.g. culturally responsive teaching and a focus on anti-discrimination in teacher education, are needed.

Grading is an educational practice that is relevant for students for many reasons. They represent a signal for pupils and their parents, they have an impact on students' motivation and also consequences on their choices (Ballarino & Panichella, 2021). Evaluating teachers' discrimination in grading could, therefore, help in alleviating the educational inequalities that affect certain social groups.

Table A3
Results by methodology.

Methodology	Result		
	Yes	No for at least one characteristic	Mixed
Experiment	31%	53%	16%
Regression	79%	7%	14%
Natural experiment	100%	0%	0%
Correlation	100%	0%	0%

Table A4
Methodology by result (gender).

Methodology	Result		
	Yes	No for at least one characteristic	Mixed
Experiment	14%	19%	14%
Regression	24%	5%	9,5%
Natural experiment	9,5%	0%	0%
Correlation	5%	0%	0%

Table A5
Results by methodology (race/ethnicity).

Methodology	Result		
	Yes	No for at least one characteristic	Mixed
Experiment	27%	33%	20%
Regression	13%	0%	7%
Natural experiment	0%	0%	0%
Correlation	0%	0%	0%

Table A6
Results by methodology (migratory background).

Methodology	Result		
	Yes	No for at least one characteristic	Mixed
Experiment	20%	0%	0%
Regression	60%	0%	20%
Natural experiment	0%	0%	0%
Correlation	0%	0%	0%

Table A7
Results by country.

Countries	Result		
	Yes	No for at least one characteristic	Mixed
United States of America	27%	46%	27%
Germany	66%	17%	17%
United Kingdom	33,3%	33,3%	33,3%
Italy	100%	0%	0%
Sweden	50%	50%	0%
Israel	50%	50%	0%
Norway	100%	0%	0%
Czech Republic	100%	0%	0%
Spain	0%	0%	100%
Netherlands	0%	100%	0%
Hungary	100%	0%	0%
Brazil	100%	0%	0%
India	0%	100%	0%

Funding

No funding.

Declaration of Competing Interest

None.

Acknowledgements

The authors are grateful to Prof. Gabriele Ballarino (Università degli Studi di Milano) for his useful comments and suggestions.

Appendix

Table A1, Table A2, Table A3, Table A4, Table A5, Table A6 and Table A7.

References

Alesina, A., Carlana, M., La Ferrara, E., & Pinotti, P. (2018). Revealing stereotypes: Evidence from immigrants in schools. *Institute of Labor Economics (IZA)*. ⟨https://ideas.repec.org/p/iza/izadps/dp11981.html⟩.

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology, 85*(2), 347–356. https://doi.org/10.1037/0022-0663.85.2.347

Bonesrønning, H. (2008). The effect of grading practices on gender differences in academic performance. *Bulletin of Economic Research, 60*(3), 245–264. https://doi.org/10.1111/j.1467-8586.2008.00278.x

Botelho, F., Madeira, R. A., & Rangel, M. A. (2015). Racial discrimination in grading: Evidence from Brazil. *American Economic Journal: Applied Economics, 7*(4), 37–52. https://doi.org/10.1257/app.20140352

Burgess, S., & Greaves, E. (2013). Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities. *Journal of Labor Economics, 31*(3), 535–576. https://doi.org/10.1086/669340

Carter, R. S. (1952). How invalid are marks assigned by teachers. *Journal of Educational Psychology, 43*(4), 218–228. https://doi.org/10.1037/h0061688

Chase, C. I. (1986). ESSAY TEST SCORING: INTERACTION OF RELEVANT VARIABLES. *Journal of Educational Measurement, 23*(1), 33–41. https://doi.org/10.1111/j.1745-3984.1986.tb00232.x

Copur-Gencturk, Y., Cimpian, J. R., Lubienski, S. T., & Thacker, I. (2020). Teachers' Bias Against the Mathematical Ability of Female, Black, and Hispanic Students. *Educational Researcher, 49*(1), 30–43. https://doi.org/10.3102/0013189X19890577

Di Liberto, A., Casula, L., & Pau, S. (2022). Grading practices, gender bias and educational outcomes: Evidence from Italy. *Education Economics, 30*(5), 481–508. https://doi.org/10.1080/09645292.2021.2004999

Dian, M., & Triventi, M. (2021). The weight of school grades: Evidence of biased teachers' evaluations against overweight students in Germany. *PLoS ONE, 16*(2). https://doi.org/10.1371/journal.pone.0245972

Falch, T., & Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review, 36*, 12–25. https://doi.org/10.1016/j.econedurev.2013.05.002

Fogel, L. S., & Nelson, R. O. (1983). The effects of special education labels on teachers' behavioral observations, checklist scores, and grading of academic work. *Journal of School Psychology, 21*(3), 241–251. https://doi.org/10.1016/0022-4405%2883%2990019-5

Guttmann, J., & Boudo, M. (1988). Teachers' evaluations of pupils' performance as a function of pupils' sex, family type and past school performance. *Educational Review, 40*(1), 105–113. https://doi.org/10.1080/0013191880400108

Hanna, R. N., & Linden, L. L. (2012). Discrimination in Grading. *American Economic Journal: Economic Policy, 4*(4), 146–168. https://doi.org/10.1257/pol.4.4.146

Hinnerich, B. T., Höglin, E., & Johannesson, M. (2011). Are boys discriminated in Swedish high schools. *Economics of Education Review, 30*(4), 682–690. https://doi.org/10.1016/j.econedurev.2011.02.007

Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology, 109*(6), 871–888. https://doi.org/10.1037/edu0000156

Kisfalusi, D., Janky, B., & Takács, K. (2021). Grading in Hungarian Primary Schools: Mechanisms of Ethnic Discrimination against Roma Students. *European Sociological Review, 37*(6), 899–917. https://doi.org/10.1093/esr/jcab023

Kiss, D. (2013). Are immigrants and girls graded worse? Results of a matching approach. *Education Economics, 21*(5), 447–463. https://doi.org/10.1080/09645292.2011.585019

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics, 92*(10–11), 2083–2105. https://doi.org/10.1016/j.jpubeco.2008.02.009

Marcenaro-Gutierrez, O., & Vignoles, A. (2015). A Comparison of Teacher and Test-Based Assessment for Spanish Primary and Secondary Students. In *Educational Research, 57* pp. 1–21). Education Collection,. https://doi.org/10.1080/00131881.2014.983720

Mateju, P., & Smith, M. L. (2015). Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools. *British Journal of Sociology of Education, 36*(6), 871–895. https://doi.org/10.1080/01425692.2013.874278

Protivínský, T., & Münich, D. (2018). Gender Bias in teachers' grading: What is in the grade. *Studies in Educational Evaluation, 59*, 141–149. https://doi.org/10.1016/j.stueduc.2018.07.006

Shepherd, M. A. (2011). Effects of Ethnicity and Gender on Teachers' Evaluation of Students' Spoken Responses. *Urban Education, 46*(5), 1011–1028. https://doi.org/10.1177/0042085911400325

Shepherd, M. A. (2020). Effects of race/ethnicity, gender, and intonation on secondary science teachers' evaluation of spoken responses. *Urban Education, 55*(5), 730–752. https://doi.org/10.1177/0042085916660346

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics, 45*(1), 523–538. https://doi.org/10.1007/s00181-012-0609-x

Triventi, M. (2020). Are Children of Immigrants Graded Less Generously by their Teachers than Natives, and Why? Evidence from Student Population Data in Italy. *The International Migration Review, 54*(3), 765–795. https://doi.org/10.1177/0197918319878104

Westphal, A., Becker, M., Vock, M., Maaz, K., Neumann, M., & McElvany, N. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: The role of classroom behavior, motivation, and teacher characteristics. *Contemporary Educational Psychology, 46*, 218–227. https://doi.org/10.1016/j.cedpsych.2016.06.004

Argentin, G., & Pavolini, E. (2020). How schools directly contribute to the reproduction of social inequalities. Evidence of tertiary effects, taken from Italian. *Research Politiche Sociali, 1*, 149–176. https://doi.org/10.7389/97339

Assari, S., & Caldwell, C. H. (2018). Teacher discrimination reduces school performance of African American youth: Role of gender. *Brain Sciences, 8*(10), 183. https://doi.org/10.3390/brainsci8100183

Auwarter, A. E., & Aruguete, M. S. (2008). Effects of Student Gender and Socioeconomic Status on Teacher Perceptions. *The Journal of Educational Research, 101*(4), 242–246. https://doi.org/10.3200/JOER.101.4.243-246

Avitzour, E., Choen, A., Joel, D., & Lavy, V. (2021). *On the Origins of Gender-Biased Behavior: The Role of Explicit and Implicit Stereotypes* (No. 27818; NBER Working Paper Series. *National Bureau of Economic Research*, 19. https://www.nber.org/system/files/working_papers/w27818/w27818.pdf.

Ballarino, G., & Panichella, N. (2021). *Sociologia dell'istruzione il Mulino*.

Barg, K. (2013). The influence of students' social background and parental involvement on teachers' school track choices: Reasons and Consequences. *European Sociological Review, 29*(3), 565–579. https://doi.org/10.1093/esr/jcr104

Barone, C., Assirelli, G., Abbiati, G., Argentin, G., & De Luca, D. (2018). *Social origins, Relative risk aversion and Track Choice Acta Sociologica,, 61*(4), 441–459. JSTOR. https://www.jstor.org/stable/48595443.

Baysu, G., Agirdag, O., & De Leersnyder, J. (2023). The Association Between Perceived Discriminatory Climate in School and Student Performance in Math and Reading: A Cross-National Analysis Using PISA 2018. *Journal of Youth and Adolescence, 52*(3), 619–636. https://doi.org/10.1007/s10964-022-01712-3

Benner, A. D., & Graham, S. (2011). Latino Adolescents' Experiences of Discrimination Across the First 2 Years of High School: Correlates and Influences on Educational Outcomes: Latino Adolescents' Experiences of Discrimination. *Child Development, 82*(2), 508–519. https://doi.org/10.1111/j.1467-8624.2010.01524.x

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology, 85*(2), 347–356. https://doi.org/10.1037/0022-0663.85.2.347

Bonesronning, H. (2004). Can effective teacher behavior be identified? (Education Collection) *Economics of Education Review, 23*(3), 237–247. https://doi.org/10.1016/j.econedurev.2003.07.002.

Bonesrønning, H. (2008). The effect of grading practices on gender differences in academic performance. *Bulletin of Economic Research, 60*(3), 245–264. https://doi.org/10.1111/j.1467-8586.2008.00278.x

Bruneau, E., Szekeres, H., Kteily, N., Tropp, L. R., & Kende, A. (2020). Beyond dislike: Blatant dehumanization predicts teacher discrimination. *Group Processes & Intergroup Relations, 23*(4), 560–577. https://doi.org/10.1177/1368430219845462

Burgess, S., & Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics, 31*(3), 535–576. https://doi.org/10.1086/669340

Butler-Barnes, S. T., & Inniss-Thompson, M. N. (2020). "My Teacher Doesn't Like Me": Perceptions of Teacher Discrimination and School Discipline among African-American and Caribbean Black Adolescent Girls. *Education Sciences, 10*(2), 44. https://doi.org/10.3390/educsci10020044

Campbell, T. (2015). Stereotyped at Seven? Biases in Teacher Judgement of Pupils' Ability and Attainment. *Journal of Social Policy, 44*(3), 517–547. https://doi.org/10.1017/S0047279415000227

Carter, T. P. (1971). *School Discrimination: The Mexican American Case* (64300286; ED048969). 1–20. *Education Collection*. https://www.proquest.com/scholarly-journals/school-discrimination-mexican-american-case/docview/64300286/se-2?accountid=12459.

Crowl, T. K., & MacGinitie, W. H. (1974). The influence of students' speech characteristics on teachers' evaluations of oral answers. *Journal of Educational Psychology, 66*(3), 304–308. https://doi.org/10.1037/h0036498

Davis, A., & Dollard, J. (1940). Children of Bondage: The Personality Development of Negro Youth in the Urban South. *American Council on Education*.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics, 86*(1), 195–210. https://doi.org/10.1162/003465304323023750

Dee, T. S. (2005). A Teacher like Me: Does Race, Ethnicity, or Gender Matter? *The American Economic Review, 95*(2), 158–165 (JSTOR) ⟨http://www.jstor.org/stable/4132809⟩.

D'hondt, F., Eccles, J. S., Houtte, M., & Stevens, P. A. J. (2016). Perceived ethnic discrimination by teachers and ethnic minority students' academic futility: Can parents prepare their youth for better or for worse? *Journal of Youth and Adolescence, 45*(6), 1075–1089. https://doi.org/10.1007/s10964-016-0428-z

Doyle, L., Easterbrook, M. J., & Harris, P. R. (2022). Roles of socioeconomic status, ethnicity and teacher beliefs in academic grading. *British Journal of Educational Psychology, No-Specified*. https://doi.org/10.1111/bjep.12541

Eble, A., & Hu, F. (2020). Child beliefs, societal beliefs, and teacher-student identity match. *Economics of Education Review, 77*, Article 101994. https://doi.org/10.1016/j.econedurev.2020.101994

Everett, J. A. C., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00015

Farkas, G. (2003). Racial Disparities and Discrimination in Education: What Do We know, how Do We Know It, and What Do We Need to Know? *Teachers College Record: The Voice of Scholarship in Education, 105*(6), 1119–1146. https://doi.org/10.1111/1467-9620.00279

Ferguson, R. F. (2003). Teachers' Perceptions and Expectations and the Black-White Test Score Gap. *Urban Education, 38*(4), 460–507. https://doi.org/10.1177/0042085903038004006

Fogel, L. S., & Nelson, R. O. (1983). The effects of special education labels on teachers' behavioral observations, checklist scores, and grading of academic work. *Journal of School Psychology, 21*(3), 241–251. https://doi.org/10.1016/0022-4405%2883%2990019-5

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Guryan, J., & Charles, K. K. (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal, 123*(572), F417–F432. https://doi.org/10.1111/ecoj.12080

Hagenauer, G., & Volet, S. E. (2014). Teacher–student relationship at university: An important yet under-researched field. *Oxford Review of Education, 40*(3), 370–388. https://doi.org/10.1080/03054985.2014.921613

Hinnerich, B. T., Höglin, E., & Johannesson, M. (2015). Discrimination against students with foreign backgrounds: Evidence from grading in Swedish public high schools (Education Collection) *Education Economics, 23*(6), 660–676. https://doi.org/10.1080/09645292.2014.899562.

Hoffmann, F., & Oreopoulos, P. (2009). A Professor like Me: The influence of instructor gender on college achievement. *The Journal of Human Resources, 44*(2), 479–494 (JSTOR) ⟨http://www.jstor.org/stable/20648905⟩.

Jelsma, E., Chen, S., & Varner, F. (2022). Working harder than others to prove yourself: High-effort coping as a buffer between teacher-perpetrated racial discrimination and mental health among Black American adolescents. *Journal of Youth and Adolescence, 51*(4), 694–707. https://doi.org/10.1007/s10964-021-01563-4

Jiang, S., & Dong, L. (2020). The effects of teacher discrimination on depression among migrant adolescents: Mediated by school engagement and moderated by poverty status. *Journal of Affective Disorders, 275*, 260–267. https://doi.org/10.1016/j.jad.2020.07.029

Joseph, N. M., Viesca, K. M., & Bianco, M. (2016). Black Female Adolescents and Racism in Schools: Experiences in a Colorblind. *Society The High School Journal,, 100*(1), 4–25. JSTOR. http://www.jstor.org/stable/44077596.

Kehle, T. J., Bramble, W. J., & Mason, E. J. (1974). Teachers' expectations: Ratings of student performance as biased by student characteristics. *Journal of Experimental Education, 43*(1), 54–60. https://doi.org/10.1080/00220973.1974.10806612

Kleen, H., & Glock, S. (2018). A further look into ethnicity: The impact of stereotypical expectations on teachers' judgments of female ethnic minority students. *Social Psychology of Education, 21*(4), 759–773. https://doi.org/10.1007/s11218-018-9451-0

Kramer, S., & Zimmermann, F. (2021). Students with emotional and behavioral disorder and teachers' stereotypes—Effects on teacher judgments. *Journal of Experimental Education, Abele, A. E., Wojciszke, B. (2007). Agency and communion from the perspective of self* versus *others. Journal of Personality and Social Psychology, 93*(5), *751–763*. https://doi.org/10.1037/0022-3514.93.5.751, No-Specified. https://doi.org/10.1080/00220973.2021.1934809.

Kyere, E., Rudd, S. E., & Fukui, S. (2022). The Role of Racial–Ethnic Identity in Understanding Depressive Symptoms in the Context of Racial Discrimination Among African American Youth. *Journal of the Society for Social Work and Research, 13*(2), 261–279. https://doi.org/10.1086/711700

Lindahl, E. (2007). *Comparing teachers' assessments and national test results – evidence from Sweden* (Working Paper Series No. 24; p. 20). IFAU - Institute for Evaluation of Labour Market and Education Policy. *Working Paper Series*.

Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice, 27*(5), 480–500. https://doi.org/10.1080/0969594X.2020.1799190

Lopez, W. D., LeBrón, A. M. W., Graham, L. F., & Grogan-Kaylor, A. (2016). Discrimination and Depressive Symptoms Among Latina/o Adolescents of Immigrant Parents. *International Quarterly of Community Health Education, 36*(2), 131–140. https://doi.org/10.1177/0272684X16628723

Lorenz, G. (2021). Subtle discrimination: Do stereotypes among teachers trigger bias in their expectations and widen ethnic achievement gaps?. In *Social Psychology of Education: An International Journal, 24* pp. 537–571) Education Collection,. https://doi.org/10.1007/s11218-021-09615-0

MacCann, C., & Roberts, R. D. (2013). Just as smart but not as successful: Obese students obtain lower school grades but equivalent test scores to nonobese students. *International Journal of Obesity, 37*(1), 40–46. https://doi.org/10.1038/ijo.2012.47

Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85–103. https://doi.org/10.1016/j.edurev.2016.06.003

Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education, 60*(3), 245–256. https://doi.org/10.1177/0004944116664618

Mateju, P., & Smith, M. L. (2015). Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools. *British Journal of Sociology of Education, 36*(6), 871–895. https://doi.org/10.1080/01425692.2013.874278

Nguyen, M. (2018). *The Relationship between Race-Congruent Students and Teachers: Does Racial Discrimination Exist?* (Departmental Working Papers). *Department of Economics*. Louisiana State University,. ⟨https://ideas.repec.org/p/lsu/lsuwpp/2018-06.html⟩.

Niederle, M., & Vesterlund, L. (2010). Explaining the Gender Gap in Math Test Scores: The Role of Competition. *Journal of Economic Perspectives, 24*(2), 129–144. ⟨https://ideas.repec.org/a/aea/jecper/v24y2010i2p129-44.html⟩.

OECD. (2012). Grade Expectations: How Marks and Education Policies Shape Students' Ambitions. *OECD*. https://doi.org/10.1787/9789264187528-en

Parks, F. R., & Kennedy, J. H. (2007). The Impact of Race, Physical Attractiveness, and Gender On Education Majors' and Teachers' Perceptions of Student Competence. In *Journal of Black Studies, 37* pp. 936–943). Education Collection,. https://doi.org/10.1177/0021934705285955

Peters, M., Godfrey, C., McInerney, P., Munn, Z., Trico, A., & Khalil, H. (2020). Scoping Reviews. In E. Aromataris, & Z. Munn (Eds.), *JBI Manual for Evidence Synthesis*. JBI. https://doi.org/10.46658/JBIMES-20-12.

Petersen, T. (2005). Discrimination, Measurement in. *Encyclopedia of Social Measurement* (pp. 677–684). Elsevier,. https://doi.org/10.1016/B0-12-369398-5/00544-2

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: {The} role of grading scales. *Educational Evaluation and Policy Analysis, 42*(3), 375–392. https://doi.org/10.3102/0162373720932188

Rauschenberg, S. (2014). How Consistent Are Course Grades? An Examination of Differential Grading. In *Education Policy Analysis Archives, 22* pp. 1–41). Education Collection,. https://www.proquest.com/scholarly-journals/how-consistent-are-course-grades-examination/docview/1697495514/se-2?accountid=12459.

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' Perceptions of Students' Mathematics Proficiency May Exacerbate Early Gender Gaps in Achievement. *Developmental Psychology, 50*(4), 1262–1281. Education Collection. https://doi.org/10.1037/a0035073.

Sedlacek, Q. C. (2021). Experimental evidence of disparities in biology teachers' responses to student writing. In *International Journal of Science Education, 43* pp. 2351–2373). Education Collection,. https://doi.org/10.1080/09500693.2021.1963499

Snijders, I., Wijnia, L., Kuiper, R. M., Rikers, R. M. J. P., & Loyens, S. M. M. (2022). Relationship quality in higher education and the interplay with student engagement and loyalty. *British Journal of Educational Psychology, 92*(2), 425–446. https://doi.org/10.1111/bjep.12455

Spear, M. G. (1984). The biasing influence of pupil sex in a science marking exercise. *Research in Science & Technological Education, 2*(1), 55–60. https://doi.org/10.1080/0263514840020107

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–762. https://doi.org/10.1037/a0027627

Terrier, C. (2020). Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review, 77*, Article 101981. https://doi.org/10.1016/j.econedurev.2020.101981

Thomas, O. N., Caldwell, C. H., Faison, N., & Jackson, J. S. (2009). Promoting academic achievement: The role of racial identity in buffering perceptions of teacher discrimination on academic achievement among African American and Caribbean Black adolescents. *Journal of Educational Psychology, 101*(2), 420–431. https://doi.org/10.1037/a0014578

Tobisch, A., & Dresel, M. (2017). Negatively or positively biased? Dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds. *Social Psychology of Education, 20*(4), 731–752. https://doi.org/10.1007/s11218-017-9392-z

Triventi, M. (2020). Are Children of Immigrants Graded Less Generously by their Teachers than Natives, and Why? Evidence from Student Population Data in Italy. *The International Migration Review, 54*(3), 765–795. https://doi.org/10.1177/0197918319878104

van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The Implicit Prejudiced Attitudes of Teachers: Relations to Teacher Expectations and the Ethnic Achievement Gap. *American Educational Research Journal, 47*(2), 497. Education Collection. https://www.proquest.com/scholarly-journals/implicit-prejudiced-attitudes-teachers-relations/docview/356883505/se-2?accountid=12459.

van Ewijk, R. (2011). Same Work, Lower Grade? Student Ethnicity and Teachers' Subjective Assessments. *Economics of Education Review, 30*(5), 1045–1058. https://doi.org/10.1016/j.econedurev.2011.05.008

Wang, M.-T., & Huguley, J. P. (2012). Parental Racial Socialization as a Moderator of the Effects of Racial Discrimination on Educational Success Among African American Adolescents. *Child Developmen), 1716*. Education Collection. https://doi.org/10.1111/j.1467-8624.2012.01808.xt, 83(5.

Weeks, M. R., & Sullivan, A. L. (2019). Discrimination Matters: Relations of perceived discrimination to student mental health. *School Mental Health, 11*(3), 425–437. https://doi.org/10.1007/s12310-019-09309-1

Wenz, S. E. (2020). Discrimination in Education: Methodology, theory, and empirics of teachers' stereotypes, prejudice, and discriminatory behavior. *GESIS-Schriftenreihe*. https://doi.org/10.21241/SSOAR.67307

Willis, P. E. (2006). *Learning to labour: How working class kids get working class jobs (Repr)*. Ashgate.