



## OPEN ACCESS

## EDITED BY

Chris Krägeloh,  
Auckland University of Technology,  
New Zealand

## REVIEWED BY

Niclas Schaper,  
University of Paderborn, Germany  
Mariana Cernicova-Buca,  
Politehnica University of Timișoara, Romania

## \*CORRESPONDENCE

Jun Zhu  
✉ zhujun@bfsu.edu.cn  
Tingting Han  
✉ han\_psy@aliyun.com

RECEIVED 15 May 2023

ACCEPTED 18 January 2024

PUBLISHED 02 February 2024

## CITATION

Chai F, Ma J, Wang Y, Zhu J and Han T (2024)  
Grading by AI makes me feel fairer? How  
different evaluators affect college students'  
perception of fairness.  
*Front. Psychol.* 15:1221177.  
doi: 10.3389/fpsyg.2024.1221177

## COPYRIGHT

© 2024 Chai, Ma, Wang, Zhu and Han. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness

Fangyuan Chai<sup>1</sup>, Jiajia Ma<sup>1</sup>, Yi Wang<sup>1</sup>, Jun Zhu<sup>1\*</sup> and Tingting Han<sup>2\*</sup>

<sup>1</sup>Graduate School of Education, Beijing Foreign Studies University, Beijing, China, <sup>2</sup>School of Marxism, Hubei University of Economics, Wuhan, Hubei, China

**Introduction:** In the field of education, new technologies have enhanced the objectivity and scientificity of educational evaluation. However, concerns have been raised about the fairness of evaluators, such as artificial intelligence (AI) algorithms. This study aimed to assess college students' perceptions of fairness in educational evaluation scenarios through three studies using experimental vignettes.

**Methods:** Three studies were conducted involving 172 participants in Study 1, 149 in Study 2, and 145 in Study 3. Different evaluation contexts were used in each study to assess the influence of evaluators on students' perception of fairness. Information transparency and explanations for evaluation outcomes were also examined as potential moderators.

**Results:** Study 1 found that different evaluators could significantly influence the perception of fairness under three evaluation contexts. Students perceived AI algorithms as fairer evaluators than teachers. Study 2 revealed that information transparency was a mediator, indicating that students perceived higher fairness with AI algorithms due to increased transparency compared with teachers. Study 3 revealed that the explanation of evaluation outcomes moderated the effect of evaluator on students' perception of fairness. Specifically, when provided with explanations for evaluation results, the effect of evaluator on students' perception of fairness was lessened.

**Discussion:** This study emphasizes the importance of information transparency and comprehensive explanations in the evaluation process, which is more crucial than solely focusing on the type of evaluators. It also draws attention to potential risks like algorithmic hegemony and advocates for ethical considerations, including privacy regulations, in integrating new technologies into educational evaluation systems. Overall, this study provides valuable theoretical insights and practical guidance for conducting fairer educational evaluations in the era of new technologies.

## KEYWORDS

higher education evaluation, AI algorithm, fairness perception, information transparency, explanation

## Introduction

Imagine at the end of a public elective course you are taking, the teacher grades each student with a final score consisting of everyone's daily class performance and a final exam. However, you are discontented with your score and have some questions. Is the score given by the teacher fair? Is the marking process transparent? Why did I get this score? Is there an explanation? On second thoughts, if it is not the teacher who is doing the grading, but the

artificial intelligence (AI) evaluation system, which could record students' class behaviors with the help of intelligent technologies such as face and body movement recognition, evaluates your class performance and forms your usual performance using the algorithm logic under set rules, would you think the evaluation process and the grades you get are fairer than the teacher's evaluation?

In education, the fairness of educational evaluation has always been widely concerned. In the field of higher education, a variety of educational evaluations guides the development direction of colleges and universities and students like a baton. Therefore, the fairness of evaluation will have a direct bearing on the quality of student training. Moreover, with the rapid and iterative development of big data and artificial intelligence technologies, the digital transformation of education has become the future direction. The UN Transforming Education Summit [United Nations \(2022\)](#) pointed out that global education faced serious challenges so education transformation is urgently needed and the power of digital teaching and learning must be fully harnessed. In addition, the [Ministry of Education of The people's Republic of China \(2023\)](#) strongly called for the international community to promote the digital transformation of education as well.

Thus, AI algorithms has been gradually applied in the field of educational evaluation, including the evaluation of students' ability and knowledge level, personality and mental health, and so on. For instance, researchers have utilized natural language processing technology to develop models for detecting syntax errors and recognizing argumentation structures, thereby achieving automated and accurate composition scoring ([Ruiji et al., 2018](#)). Besides, machine learning techniques are also applied to design an automatic classroom assessment scoring system for the classroom environment, which could observe and record teacher-student interactions while generating instant feedback ([James et al., 2018](#)). Hence, there is no doubt that the application of new technologies, such as AI algorithms, has provided new ideas and solutions for educational evaluation reform. Nevertheless, as the traditional evaluators of educational evaluation have always been teachers or educational institutions, students have become familiar with these subjective evaluation habits. Given it, it remains to be tested whether students can feel that the process and the scores are fairer when the evaluator is replaced by an artificial intelligence system. The research claimed that since it was still humans who set the rules for algorithmic evaluation, the subjective bias of the evaluator might still be hidden in the algorithm ([Yang, 2022](#)). In other words, the opaque design, unexplainable process, and irrefutable results of AI algorithm evaluations are likely to raise questions about fairness. As a technological tool, it remains controversial as to whether it will enhance or undermine fairness in evaluation. In addition, little research exists on the topic of whether AI algorithmic evaluations are fairer than teacher evaluations, and empirical studies on the influencing mechanism and functional boundary behind it are also lacking.

## Perception of fairness in the field of organizations and education

Fairness is a goal that human society strives for diligently. The fairness perception, as an individual's subjective feeling and evaluation of justice, is an important indicator of the degree of fairness ([Lv and Liu, 2009](#)). The study of the perception of fairness began with Adams' equity theory, which pointed out that employees tended to care not

only about their absolute compensation but also about their relative compensation compared with others, i.e., perception of distributive justice ([Adams, 1965](#)). [Leventhal et al. \(1980\)](#) further argued from the perspective of the allocation process that the perception of fairness was closely related to the procedural operations that determined results, such as the development, implementation, and improvement of the allocation system, i.e., procedural justice. Based on previous theories, [Bies and Moag \(1986\)](#) focused on the influence of interpersonal interaction on the perception of fairness during the implementation and proposed the concept of "interactional justice."

Later, [Greenberg and Cropanzano \(1993\)](#) proposed a taxonomy that reclassified organizational justice into four categories by crosscutting two independent dimensions: (a) category of justice, namely procedural and distributive justice, and (b) focal determinants, including structural and social factors. The four categories are systemic justice, configure justice, informational justice, and interpersonal justice. This categorization further divides interactive justice of [Bies and Moag \(1986\)](#) into informational and interpersonal justice, with information justice being distinct from other types of justice according to Greenberg. Information justice refers to the provision of relevant information and explanations by leaders to their subordinates during the decision-making process, the timely sharing of information, the addressing of concerns related to relevant issues, and the promotion of open and transparent information dissemination ([Greenberg and Cropanzano, 1993](#)). Information justice is guided by two important principles or criteria: justification and truthfulness. Managers should provide sufficient justifications for their decision-making actions, and these explanations should be honest, open, and candid ([Greenberg and Cropanzano, 1993](#); [Bies, 2005](#)). Building upon this foundation, [Colquitt \(2001\)](#) conducted a series of studies that endorsed the distinction between informational and interpersonal fairness and formalized two approaches for measuring fairness: the direct method and the indirect method. These two approaches correspond to the measurement of subjective and objective aspects of fairness, respectively. The former captures people's feelings and perceptions of the fairness of an event, while the latter focuses on the inherent fairness of the event itself. This theoretical framework lays the groundwork for measuring perceptions of fairness. By considering the dimensions of justice and integrating the concepts of informational and interpersonal justice, researchers can effectively assess how individuals perceive fairness in various contexts and enhance our understanding of the multifaceted nature of fairness perception.

In the realm of higher education, colleges, and universities can be seen as microcosms of society, while individual classes in which students are enrolled can be likened to organizations. Within the context of educational evaluation, the relationship between the evaluator (whether it is a teacher or an AI algorithm) and the evaluatee (the student) can also be compared to the relationship between supervisors and employees. Just as employees place great importance on the fairness of their performance evaluations, students are equally concerned about the fairness with which their academic performance is assessed. Apart from the extensively discussed concepts of distributive and procedural fairness, the integration of AI algorithms into educational evaluation has also stirred concerns among students regarding information fairness. This dimension encompasses issues related to transparent communication, access to information, and the fairness of decision-making processes. Students want to ensure that their performance assessments are conducted in a manner where

information is adequately disclosed, unbiased, and consistent. Educational institutions need to address these concerns and create evaluation systems that not only adhere to principles of distributive and procedural fairness but also prioritize information fairness. By promoting transparency, ensuring equal access to information, and implementing fair decision-making practices, colleges and universities can foster an environment where students perceive their evaluations as fair and reliable.

Consequently, the appraisal processes undertaken for college students within the realm of higher education bear a resemblance to performance assessments observed in organizational behavior studies. This parallel provides a rationale for the application of organizational justice theory to the higher education context. Moreover, it is posited that information fairness plays a pivotal role in shaping college students' trust and perceptions of fairness (Sun and Huang, 2004). In essence, a comprehensive understanding of the fundamental tenets of information fairness theory as it pertains to higher education evaluation is instrumental in elucidating the determinants that sculpt college students' fairness perceptions during these evaluative interactions.

## The impact of AI algorithm on the perception of fairness in evaluation settings

Nowadays, the AI algorithm has been gradually applied to educational evaluation. In terms of theoretical research, some scholars emphasized that AI algorithm evaluation could achieve innovation in the content and methods of evaluation practice, which was conducive to promoting the development of educational evaluation theory in the context of digital intelligence (Mao et al., 2020; Yigitcanlar et al., 2020; Xu et al., 2022). In terms of empirical research, other scholars have analyzed the specific applications of AI algorithms in student learning behavior evaluation, classroom teaching quality evaluation, higher education enrollment, and so on. For example, AI algorithm models could be used to explain the major issues affecting learners' learning, predict the trend of changes in students' learning behavior, assess their learning behavior and status. Afterwards, students can receive effective adaptive feedback and interventions, and the quality of teaching can be improved (Martins and Von Wangenheim, 2022; Xiao et al., 2022).

Adams (1965) emphasized the subjective psychological perception of fairness in his theory of organizational fairness. Afterward, based on this theory, some scholars considered the perception of fairness in performance appraisal as the employees' overall perception of the fairness of the performance appraisal system in question within the appraisal cycle (Liu et al., 2003; Gu et al., 2020; Ryu and Hong, 2020). In addition, researchers also believed that the perception of fairness refers to an individual's judgment and emotional experience of the fairness or unfairness of a phenomenon (Lv, 2010; Barsky et al., 2011; Rupp, 2011; Engelmann and Tomasello, 2019). Thus, in the present study, we define "perception of fairness" as an individual's subjective psychological perception of the fairness or unfairness of the behaviors under specific circumstances.

In terms of the impact of AI algorithms on the perception of fairness, existing research on whether the impacts are positive is disputed. For one thing, given the technical manipulations, the uninterpretability of the AI algorithm as an evaluator was pointed out

as a risk (e.g., Adadi and Berrada, 2018; Leslie, 2019; Wang and Wang, 2020). The interpretability of an AI algorithm refers to its ability to be understood and explained, and it is claimed that its operational mechanism, process, and result can be clearly understood by users (Lundberg et al., 2020; Shin, 2021; Cong et al., 2022; Sun, 2022). However, for most users, the complex internal operations of AI algorithm are not easy to understand so there is a "black box" problem for it as an evaluator (Adadi and Berrada, 2018; Guidotti et al., 2018; Rudin, 2019; Luo et al., 2021; Zednik and Boelsen, 2022). As a result, compared to the human evaluation, which can be observed and understood more easily, the unobservable and unintelligible characteristics of AI algorithm evaluation were more likely to raise doubts about the fairness of the evaluation (Burrell, 2016). For another, some scholars argued that those "laymen" who do not know much about algorithms are more likely to have an attitude of "algorithm appreciation," and believed that AI algorithms had more positive characteristics (Logg et al., 2019; Araujo et al., 2020). That is to say, compared to human evaluation, the AI algorithm, as the evaluator, can collect, process and analyze all kinds of information and data with the help of technologies, and present the development and changes of students in thought, mentality, and behavior in a more scientific and fuller way, bringing a higher perception of fairness to the evaluation object (Wu and Chen, 2014; Williamson, 2017). In addition, the data information obtained in this way is more authentic and reliable, which can enrich the data sources and improve the objectivity of the evaluation (Zhang et al., 2022). In other words, the AI algorithm can eliminate the personal subjective bias of educational evaluators under the logic of digital technology, improving the evaluation method and process and enabling evaluators to obtain a relatively higher perception of fairness.

## Information transparency and the perception of fairness

According to the theory of informational justice, it is necessary to provide employees with relevant information timely based on the implementation of decisions (Greenberg and Cropanzano, 1993; Cropanzano and Greenberg, 1997). Only when the information is sufficiently transparent will employees demonstrate a high degree of trust in the team and thus generate a perception of fairness (Lind, 2001; Zheng and Liu, 2016). In the domain of higher education, transparency of the information is one of the key elements of modern university governance and decision-making, which guarantees students' rights to information, participation, and supervision, and improves their enthusiasm to participate in university affairs (Damme, 2001; Paloma Sánchez and Elena, 2006; Jihui and Zhilin, 2015; Jongbloed et al., 2018). Hence, information transparency is more than important for students to form a perception of fairness.

Despite that the concept of "transparency" has multiple meanings, the one more widely used is from the perspective of information itself, which defines transparency as a state of information sharing. It is the degree of information disclosure and public provision to external personnel, emphasizing the visibility and availability of information, i.e., the extent to which information is readily accessible, available, and applicable to those who need it or to all stakeholders (Florini, 2007; Tielenburg, 2018; Wang and Ren, 2022). Therefore, in conjunction

with informational justification, this study defines “information transparency” in the evaluation as the extent to which valid information, such as the evaluation criteria, methods, and results, is publicly presented and provided to evaluation objects.

In organizational studies on how information transparency affects the perception of fairness, it has been pointed out that differences in information fairness or low information transparency constructed information barriers that subsequently create suspicion and mistrust among these members (Carless, 2009; He et al., 2017). In other words, the level of presentation of information related to evaluation criteria, methods, and results by evaluators is one of the key factors affecting the evaluation objects’ perception of fairness (Sweeney and McFarlin, 1993; Dochy et al., 1999; Van Berkel et al., 2021). Therefore, information transparency is a vital link to promote the formation of the perception of fairness (Pralhad and Ramaswamy, 2002; Pei et al., 2021). To be more specific, when evaluation criteria are not made public in a way that is easy to understand, the perception of fairness of the person being evaluated will decline (Chan Kim and Mauborgne, 1998; Yuan et al., 2022). For instance, some scholars emphasized that employers had a responsibility to inform employees of the procedures and standards of their performance evaluation system and to disclose the relevant evaluation methods promptly to ensure the organization and evaluation were in a relatively transparent environment, which was a basic condition for employees to develop a perception of fairness (Belogolovsky and Bamberger, 2014; Schnackenberg and Tomlinson, 2016; Wei and Sun, 2022). Moreover, the transparency of the evaluation method is an effective basis for judging whether the evaluation is fair and reasonable. If the methods used by the evaluator are consistent, accurate, unbiased, reasonable, and transparent, then the evaluation object will be able to establish a high perception of fairness (Rasooli et al., 2019). For instance, the evaluation methodology employed by AI algorithms constitutes a comprehensive digital service based on algorithmic rules and programmatic design. In comparison to teacher evaluations, AI algorithmic assessments offer greater clarity and transparency, eliminating personal emotional factors that may bias evaluators and consequently yielding a heightened perception of fairness. To illustrate this, consider the issue of fairness in the university admissions process. When employing AI algorithms for student admissions evaluations, the algorithms assess and compare candidates based on a range of objective criteria and data indicators such as academic performance and standardized test scores. This evaluation approach remains unaffected by subjective biases, thereby avoiding deviations driven by evaluators’ personal preferences or emotions (e.g., Dennis, 2018; Marcinkowski et al., 2020; Ahmad et al., 2022). In contrast, teacher evaluations may be susceptible to subjective influences, such as individual preferences or emotional factors, which can potentially undermine fairness. Moreover, the utilization of AI algorithms in assessment processes enhances transparency. This approach affords students a lucid grasp of the evaluative criteria, methodologies, and their application to individual scenarios. For instance, certain academics have developed a framework for certifying learning outcomes grounded in blockchain technology, leveraging its inherent characteristics of public verifiability, authenticity of content, and resistance to alteration, thereby ensuring a true and precise documentation of learners’ educational trajectories (Wu et al., 2018). Evaluators can scientifically validate learners’ accomplishments by examining the recorded data, thereby furnishing a transparent, equitable, and objective certification of achievements. This level of clarity enables students to more

thoroughly understand the basis of their assessments and facilitates the pursuit of clarification or justification of the results. In juxtaposition, traditional instructor-led evaluations may suffer from a lack of such transparency, potentially leaving students in the dark regarding the exact standards and methodologies utilized in their assessment (Daud, 2022).

In the context of higher education, students’ perception of fairness in evaluations is significantly shaped by the transparency of information provided (e.g., Awad and Krishnan, 2006; Kizilcec, 2016). To elaborate, when students are given comprehensive insights into the criteria and methodologies used in the evaluation process, it fosters a stronger sense of trust toward the evaluator. This, in turn, enhances their perception of fairness. However, if the evaluation process is veiled in ambiguity and lacks transparency, it leaves students grappling with uncertainty (Baartman et al., 2007; Lebovitz et al., 2022). This lack of clarity inevitably leads to a sense of perceived unfairness, especially when the results do not align with their expectations.

## Explanation and the perception of fairness

Organizational studies have shown that in addition to the basic information about the criteria, whether the supervisor justified the decision influenced the employee’s perception of information fairness as well (Bies and Shapiro, 1987; Dulebohn and Ferris, 1999; Hollensbe et al., 2008; Hewett and Leroy, 2019). If supervisors communicated with their employees promptly and provided them with specific reasons for the decision, employees not only perceived their leaders as sincere, transparent, and trustworthy, but also had positive psychological and emotional experiences, such as the perception of fairness (Chan Kim and Mauborgne, 1998; Gardner et al., 2005). In short, providing an explanation of the procedure or result is beneficial to strengthen trust and bring a higher perception of fairness to the evaluator (Shulner-Tal et al., 2022).

At present, the concept of “explanation” is not yet universally agreed upon academically. Some scholars believed that explanation was the logical and meaningful information provided by the decision maker about the result, which could explain the relationship between the input contents and output results (Goodman and Flaxman, 2017). It has also been claimed that explanation was the process of providing a multidimensional in-depth analysis and interpretation of the actual results, reasonably explaining the causes of the changes, the connections between things and the rules by which things develop, and making guiding suggestions (Wang, 2016). Thus, in essence, explanation is a kind of information transfer, whose basic elements include the interpreter, the receiver, and the explanatory information, and whose content often involves the comparison of target levels and the analysis of actual results. In summary, this study defines “explanation” as the process of delivering the evaluation result to the evaluation object in some form after comparing their current performance with the expected standard and explaining the specific reasons for that result to them.

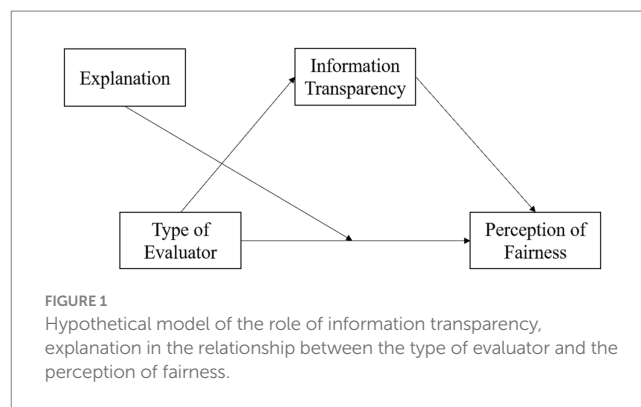
Most of the current research on explanation and the perception of fairness focused on organizational behaviors. Research in this area has found a strong link between the process of communicating performance evaluation results to the employees involved and their perception of fairness (e.g., Niehoff and Moorman, 1993; Kluger and

DeNisi, 1996; Lemons and Jones, 2001). Specifically, performance evaluations are of great importance to employees as they directly impact their interests. Not only do the evaluation results influence employees' perception of distributive fairness, but the communication, opportunities for appeals, and explanations regarding the evaluation outcomes also play a significant role in shaping their perception of procedural fairness (Magner et al., 1994; Fatt et al., 2010; Sudin, 2011; Alterman et al., 2021). When employees receive their manager's evaluation of their past work without understanding the rationale behind the results or the appropriate improvement measures based on that evaluation, they may inevitably question the fairness of the assessment (Gupta and Kumar, 2012; Gu et al., 2020). Therefore, if performance evaluations are not thoroughly explained and feedback is not provided to employees, it can lead to a certain level of job dissatisfaction and foster a perception of organizational unfairness. By contrast, adequate and effective explanations can make the overall performance evaluation fairer and increase the perception of fairness among employees, motivating them to work passionately and develop their capabilities (Xiong, 2011; De Clercq and Pereira, 2020). Similar to performance evaluation, educational evaluation is also a management tool to improve the quality of education. Evaluating and explaining the behavior of college students in education activities can not only diagnose the "shortcomings" and deficiencies in their studies but also provide important quality standards and directions for their future development.

Therefore, whether or not an explanation is given after the evaluation, and whether or not college students understand the reasons for this evaluation, will also affect their perception of fairness in educational evaluation. If these students can only see the evaluation results but do not have any access to an explanation of the results, then when there is a gap between the evaluation results and their psychological expectations, they will inevitably question the rationality of the evaluation results and the fairness of the evaluation process. Otherwise, if they are provided with accurate and comprehensive explanatory information, they will tend to feel fairer about the results, regardless of whether the evaluator is an AI or a teacher.

This study aims to investigate the impact of evaluator type (teachers vs. AI algorithms) on students' perceptions of fairness in higher education evaluation, as well as the role of information transparency and interpretation in this process. The research questions are: Do college students perceive a higher level of fairness when an AI algorithm is used as the evaluator compared to traditional teacher evaluation? How does the use of an AI algorithm affect the fairness perception of college students? What are the boundary conditions for this influence on fairness perception?

The hypotheses of the study are: (a) College students attribute higher perceptions of fairness to AI algorithms than to teachers as evaluators; (b) Information transparency serves as a mediating variable between different evaluators and the perception of fairness, such that college students perceive greater information transparency from an AI algorithm and therefore have higher perceptions of fairness; (c) The explanation of evaluation results moderates the influence of evaluators on fairness perceptions, meaning that when the reasons for evaluation decisions are explained to college students, the influence of evaluators on fairness perceptions diminishes. Based on these hypotheses, a hypothetical model of the role of information transparency and explanation in the relationship between the type of evaluator and the perception of fairness is proposed in this study (see Figure 1).



## Study 1: the impact of different evaluators on college students' perception of fairness

Based on three types of educational evaluation contexts (diagnostic, formative, and summative evaluations), Study 1 intends to investigate the effects of different evaluators (AI algorithms vs. teachers) on college students' perception of fairness and whether the effects differ.

## Methods

### Participants

The participants were randomly recruited on the online platform Credamo, and 7 were excluded for completing the questionnaire in too short a time or not meeting the response requirements. All participants must be current Chinese undergraduate students. We excluded individuals who had previously participated in similar decision-making or cognitive psychology studies to avoid any potential confounding effects of prior exposure to similar experimental paradigms. Specific demographic information of participants is as follows (see Table 1). The final sample included 172 participants ( $M_{age} = 21.22$ ,  $SD_{age} = 1.57$ ), of which 71 were male and 101 were female. Freshmen, sophomores, juniors, and seniors represented 8.2, 23.8, 39.6, and 28.4% of the sample, respectively. The participants were selected from a diverse range of majors, encompassing 10 different categories such as literature, engineering, law, medicine, and others. Among the subjects, 31.4% hailed from China's double-first-class universities, while the remaining 68.6% were enrolled in ordinary universities.

### Design

The current study employed the experimental vignette methodology, more specifically the paradigm of *paper people study*. This method involved presenting participants with meticulously designed and realistic scenarios related to educational evaluation, aiming to assess their perception of fairness. By employing this methodology, the study successfully enhanced the authenticity of the experiment while effectively manipulating and controlling independent variables, such as the evaluators in this study. Additionally, the experimental vignette methodology has demonstrated its significant value in investigating causal relationships between variables within sensitive contexts (Aguinis and Bradley,

TABLE 1 Demographic information of participants in Study 1.

Demographic characteristic		Frequency	Percentage	Demographic characteristic		Frequency	Percentage
Gender	Female	101	58.7%	Major	Law	8	4.8%
	Male	71	41.3%		Engineering	48	27.9%
Age	18	6	3.5%		Management	28	16.3%
	19	15	8.7%		Education	12	7.0%
	20	32	18.6%		Economics	17	9.9%
	21	54	31.4%		Science	16	9.3%
	22	39	22.7%		Literature	22	12.8%
	23	12	7.0%		Medicine	12	7.0%
	24	4	2.3%		Arts	6	3.5%
	25	10	5.8%		Philosophy	3	1.7%
Grade	Freshmen	41	8.2%	University type	Double-first-class	54	31.4%
	Sophomore	68	23.8%		Ordinary	118	68.6%
	Junior	49	39.6%				
	Senior	14	28.4%				

2014). It has gained considerable recognition and widespread application across diverse academic disciplines, particularly finding popularity in the realm of ethical decision-making (e.g., Finch, 1987; Hyman and Steiner, 1996; Clifford et al., 2015; Ekici et al., 2023). The study adopted a single-factor within-subject design. The independent variable was the evaluator (AI algorithm vs. teacher), and the dependent variable was students' self-reported perception of fairness for different evaluators. This design scrutinizes the variations in participants' perceptions of fairness among different evaluators for the same participant, thereby further quantifying the influence of the evaluator on students' perceptions of fairness.

## Materials and variable measurement

### Descriptive situational materials

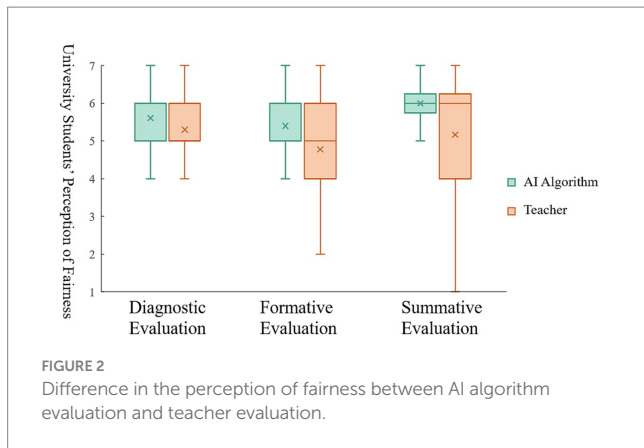
Referring to Bloom's classic classification of educational evaluation, this study classified evaluation into diagnostic evaluation, formative evaluation, and summative evaluation (Bloom, 1969, 1987), based on the timing and role of its implementation in teaching and learning. By these types mentioned above, students' learning starting point (current abilities), learning process and learning outcomes could be evaluated, respectively. Taking into account the reality of educational evaluation in China, descriptive situational materials were developed around each of the three types of evaluation. Two scenarios were created for each type of assessment to facilitate the selection of the most appropriate one through expert evaluation. The diagnostic evaluation comprised two contexts: students' oral proficiency test at the beginning of the school year and a pre-class quiz. The formative evaluation included a midterm examination and the assessment of regular performance. Lastly, the summative evaluation encompassed the evaluation of the final paper and the assessment of the final presentation. To ensure the quality of the materials, we enlisted the help of 10 master's degree students majoring in education to review the content and language of the texts. Additionally, three professors specializing in psychology were invited to evaluate all the materials. Finally, three kinds of descriptive situational materials were obtained

regarding the oral proficiency test, the scoring of regular grades, and the evaluation of the final paper.

In this study, the independent variable was managed by changing the evaluators in three descriptive situational materials. The first half of each material was the same, e.g., "Before taking a public English class, your university will organize a formal speaking test. The level of your test score will determine whether you are placed in a higher-level class or a lower-level class." Next, each scenario will include both an AI algorithm evaluation and a teacher evaluation. The AI algorithm evaluation scenario was described as "The university has applied an AI evaluation system and the speaking test is in the form of a human-computer dialog, after which the AI algorithm scores you based on your pronunciation, expression, fluency, etc." while the teacher evaluation scenario was described as "It's test is a face-to-face oral test and at the end of the test, the teacher scores you based on your pronunciation, expression, fluency, etc." The AI algorithm was assigned as 1 and the teacher was assigned as 2.

### Student's perception of fairness

There are two types of measuring methods for assessing the perception of fairness: "indirect measurement" and "direct measurement." The former focuses on evaluating the fairness of the event itself. For instance, in a competition for outstanding students, indirect measures would involve questions such as "Did the evaluation committee treat each applicant equally?," "Was the committee's decision based on accurate and sufficient information?." The latter, on the other hand, places greater emphasis on individuals' subjective feelings. For instance, in a competition for outstanding students, direct measures would inquire about individuals' perceptions through questions such as "How fair do you feel you were treated in the selection process?," and "How fair do you think the committee was in the way it made its decision?." (Colquitt and Shaw, 2013). Given that this study focused more on the subjective perceptions of students as the object of educational evaluation, the direct measurement was adopted and only one item was used, i.e., after reading the material, the participants were asked to answer how fair he or she thought the



teacher/AI algorithm was as the evaluator in this context. The perception of fairness was scored on a 7-point scale from 1 to 7, ranging from “very unfair” to “very fair.” The Cronbach’s  $\alpha$  coefficient for the scale in this study was 0.67.

## Procedures

All the participants were asked to read three descriptive situational materials in turn, and intuitively rated the fairness of the AI algorithm as the evaluator and the teacher as the evaluator, respectively. Finally, they filled in the demographic information.

## Results

With the evaluator as the independent variable and the perception of fairness of the three evaluation types as the dependent variable, a single-factor analysis of variance (ANOVA) was conducted for the samples. The results showed (see Figure 2) that in diagnostic evaluation, participants’ perception of fairness for AI algorithm evaluation was significantly higher than that for teacher evaluations ( $M_{AI}=5.60$ ,  $SD=1.12$ ,  $M_{Teacher}=5.30$ ,  $SD=0.96$ ;  $F(1, 342)=7.24$ ,  $p<0.01$ ). In formative evaluation, participants’ perception of fairness for AI algorithm evaluation was significantly higher than that for teacher evaluations ( $M_{AI}=5.40$ ,  $SD=1.21$ ,  $M_{Teacher}=4.78$ ,  $SD=1.15$ ;  $F(1, 342)=23.87$ ,  $p<0.001$ ). In summative evaluation, participants’ perception of fairness for AI algorithm evaluation was significantly higher than that for teacher evaluations as well ( $M_{AI}=5.48$ ,  $SD=1.17$ ,  $M_{Teacher}=5.20$ ,  $SD=1.22$ ;  $F(1, 349)=4.26$ ,  $p<0.05$ ). Among the three types of evaluation, the effect-size coefficients for the evaluator were  $\eta^2_{diagnostic\ evaluation}=0.03$ ,  $\eta^2_{formative\ evaluation}=0.08$ ,  $\eta^2_{summative\ evaluation}=0.02$ .<sup>1</sup> These coefficients suggest that AI algorithm evaluation consistently led to higher perceived fairness for students compared to traditional teacher evaluation, regardless of the specific type of evaluation being conducted. This implies that students perceive AI algorithm evaluation as fairer and more unbiased in comparison to evaluations conducted by human teachers, irrespective of whether the evaluation is diagnostic, formative, or summative in nature.

1 Eta-squared ( $\eta^2$ ) is a measure of effect size for use in ANOVA (Analysis of variance). It is analogous to  $R^2$  from multiple linear regression. The eta-square effect sizes are interpreted usually as follows:  $\eta^2=0.01$  indicates a small effect.  $\eta^2=0.06$  indicates a medium effect.  $\eta^2=0.14$  indicates a large effect.

Study 1 verifies the impact of the evaluator on the perception of fairness, i.e., college students’ perception of fairness brought by the AI algorithm as an evaluator is significantly higher than that of teachers. Next, Study 2 and Study 3 will further reveal the mechanism of the evaluator’s influence on the perception of fairness, examine the role of information transparency, and analyze the boundary conditions under which the evaluator’s influence on the perception of fairness may diminish.

## Study 2: the mediating role of information transparency on the influence of different evaluators on college students’ perception of fairness

Study 2 followed the design of Study 1 but additionally included a new variable information transparency. It intended to further examine the mediating role of information transparency and to investigate the main effect of different evaluators (AI algorithm/teacher) on college students’ perception of fairness and the mediating effect of information fairness.

## Methods

### Participants

The participants were randomly recruited on the online platform Credamo, and 4 were excluded for completing the questionnaire in too short a time or not meeting the response requirements. All participants must be current Chinese undergraduate students. We excluded individuals who had previously participated in similar decision-making or cognitive psychology studies to avoid any potential confounding effects of prior exposure to similar experimental paradigms. Specific demographic information of participants is as follows (see Table 2). The final sample included 149 participants ( $M_{age}=21.29$ ,  $SD_{age}=1.57$ ), of which 58 were male and 91 were female. Freshmen, sophomores, juniors, and seniors constituted 6.0, 23.4, 40.4, and 30.2% of the participants, respectively. The participants represented a diverse range of majors, including literature, engineering, law, medicine, and others, encompassing a total of 10 different categories. Within the sample, 30.2% of the subjects were enrolled in China’s double-first-class universities, while the remaining 69.8% attended ordinary universities.

### Design

This study employed the experimental vignette methodology as study 1. The independent variables were the evaluator (AI algorithm vs. teacher) and the information transparency (high transparency vs. low transparency), and the dependent variable was students’ self-reported perception of fairness for different evaluators.

### Materials and variable measurement

Descriptive situational material and manipulation of evaluators and measurement upon fairness perceptions were kept consistent with Study 1, but the measurement of information transparency was additionally added as a mediator variable. The Information Transparency Scale developed by Liu et al. was selected as the measurement tool and revised to fit the experimental context, resulting in the retention of six items,

TABLE 2 Demographic information of participants in Study 2.

Demographic characteristic		Frequency	Percentage	Demographic characteristic		Frequency	Percentage
Gender	Female	91	61.1%	Major	Law	8	5.4%
	Male	58	38.9%		Engineering	41	27.5%
Age	18	5	3.4%		Management	20	13.4%
	19	10	6.7%		Education	12	8.1%
	20	27	18.1%		Economics	14	9.4%
	21	48	32.2%		Science	13	8.7%
	22	36	24.2%		Literature	21	14.1%
	23	10	6.7%		Medicine	11	7.4%
	24	4	2.7%		Arts	6	4.0%
	25	9	6.1%		Philosophy	3	2.0%
Grade	Freshmen	9	6.0%	University type	Double-first-class	54	30.2%
	Sophomore	35	23.5%		Ordinary	118	69.8%
	Junior	60	40.3%				
	Senior	45	30.2%				

TABLE 3 Descriptive statistics of variables and correlation coefficient matrix in Study 2.

Variable	$M \pm SD$	1	2	3	4	5
1. Type of evaluator	1.50 ± 0.50					
2. Information transparency	3.99 ± 1.23	-0.66**				
3. Perception of fairness in diagnostic evaluation	5.46 ± 1.10	-0.17**	0.36**			
4. Perception of fairness in formative evaluation	5.10 ± 1.23	-0.28**	0.42**	0.33**		
5. Perception of fairness in summative evaluation	5.31 ± 1.16	-0.12*	0.37**	0.36**	0.49**	

Coding of evaluator types: AI evaluation = 1, Teacher = 2.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

including two reverse scoring items (Liu et al., 2015). The questions included “The teacher/AI algorithm relies on more useful information,” “The information relied on by the teacher/AI algorithm is easier to obtain,” “The information relied on by the teacher/AI algorithm is not sufficient” etc. The scale was scored on a 7-point scale from 1 to 7, ranging from “strongly disagree” to “strongly agree.” The Cronbach’s  $\alpha$  coefficient for the scale in this study was 0.80.

## Procedures

All participants were asked to read three descriptive situational materials in turn, and intuitively rated the fairness of the AI algorithm as the evaluator and the teacher as the evaluator, then completed the scale of information transparency. Finally, they filled in the demographic information.

## Results

### Descriptive statistics and correlation analysis

The correlation analysis of the main variables of Study 2 was performed and the results are shown in Table 3. The results showed

that there was a significant negative correlation between evaluator and information transparency and the perception of fairness of the three types of evaluation ( $r = -0.66$ ,  $p < 0.01$ ;  $r = -0.17$ ,  $p < 0.01$ ;  $r = -0.28$ ,  $p < 0.01$ ;  $r = 0.12$ ,  $p < 0.5$ ), which means that information transparency and the perception of fairness is higher in the AI algorithm evaluation condition. Besides, there was a significant positive correlation between information transparency and the perception of fairness of the three types of evaluation ( $r = 0.36$ ,  $p < 0.01$ ;  $r = 0.42$ ,  $p < 0.01$ ;  $r = 0.37$ ,  $p < 0.01$ ). The above analysis results are in line with the theoretical expectations of this study and provide preliminary support for the research hypotheses.

### Mediating effect test

This study used PROCESS developed by Preacher and Hayes to test the mediating effect of information transparency (Preacher and Hayes, 2004). With the evaluator as the independent variable and the perception of fairness of the three types of evaluation as the dependent variable, the general mediating effect was tested, and the results were shown in Table 4.

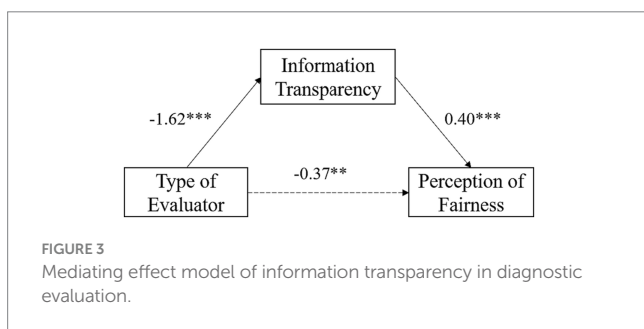
In diagnostic evaluation, equation 1 showed that the evaluator significantly predicted the perception of fairness ( $\beta = -0.37$ ,  $SE = 0.13$ ,



TABLE 4 Mediating effect test of information transparency.

Type	Equation	Dependent variable	Independent variable	R <sup>2</sup>	F	β	SE	95%CI
Diagnostic evaluation	1	Perception of fairness	Type of evaluator	0.01	8.69**	-0.37**	0.13	[-0.62, -0.12]
	2	Information transparency	Type of evaluator	0.44	227.88***	-1.62***	0.11	[-1.83, -1.41]
	3	Perception of fairness	Information transparency	0.14	24.12***	0.40***	0.06	[0.27, 0.52]
			Type of evaluator			0.27		0.16
Formative evaluation	4	Perception of fairness	Type of evaluator	0.07	24.84***	-0.69***	0.22	[-0.96, -0.41]
	5	Information transparency	Type of evaluator	0.44	227.88***	-1.62***	0.11	[-1.83, -1.41]
	6	Perception of fairness	Information transparency	0.17	30.98***	0.41***	0.07	[0.28, 0.56]
			Type of evaluator			-0.02		0.17
Summative evaluation	7	Perception of fairness	Type of evaluator	0.01	4.26*	-0.28**	0.13	[-0.54, -0.01]
	8	Information transparency	Type of evaluator	0.44	227.88***	-1.62***	0.11	[-1.83, -1.41]
	9	Perception of fairness	Information transparency	0.16	27.53***	0.48***	0.07	[0.34, 0.61]
			Type of evaluator			-0.49**		0.16

Coding of evaluator types: AI evaluation = 1, Teacher = 2.  
 \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001.



*p* < 0.01). According to equation 2, the evaluator significantly predicted information transparency ( $\beta = -1.62, SE = 0.11, p < 0.001$ ). And following equation 3, when both the evaluator and information transparency entered the regression equation, the evaluator could not significantly predict the perception of fairness ( $p > 0.05$ ), whereas information transparency significantly predicted the perception of fairness ( $\beta = 0.40, SE = 0.06, p < 0.001$ ). Therefore, the bias-corrected and accelerated bootstrap indicated that information transparency played a full mediating effect between the evaluator and the perception of fairness, with the mediating effect accounting for 63.65% of the total effect (See Figure 3).

In formative evaluation, equation 1 showed that the evaluator significantly predicted the perception of fairness ( $\beta = -0.69, SE = 0.22, p < 0.001$ ). According to equation 2, the evaluator significantly predicted information transparency ( $\beta = -1.62, SE = 0.11, p < 0.001$ ). And following equation 3, when both the evaluator and information transparency entered the regression equation, the evaluator could not significantly predict the perception of fairness ( $p > 0.05$ ), whereas information transparency significantly predicted the perception of

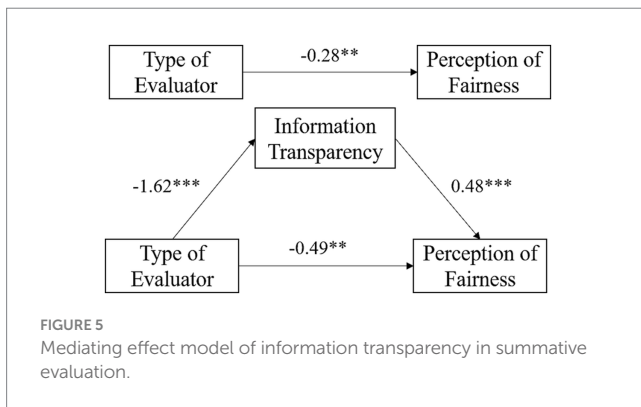
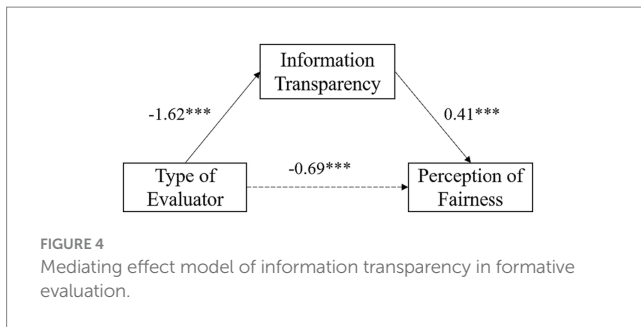
fairness ( $\beta = 0.41, SE = 0.07, p < 0.001$ ). Therefore, information transparency played a full mediating effect between the evaluator and the perception of fairness, with the mediating effect accounting for 49.05% of the total effect (See Figure 4).

In summative evaluation, equation 1 showed that the evaluator significantly predicted the perception of fairness ( $\beta = -0.28, SE = 0.13, p < 0.01$ ). According to equation 2 a, the evaluator significantly predicted information transparency ( $\beta = -1.62, SE = 0.11, p < 0.001$ ). And following equation 3, when both the evaluator and information transparency entered the regression equation, the evaluator was still able to significantly predict the perception of fairness ( $\beta = -0.49, SE = 0.16, p < 0.01$ ), and information transparency also significantly predicted the perception of fairness ( $\beta = 0.48, SE = 0.07, p < 0.001$ ). Therefore, the bias-corrected and accelerated bootstrap indicated that information transparency played a partial mediating effect between the evaluator and the perception of fairness, with the mediating effect accounting for 73.52% of the total effect (See Figure 5).

Hence, in general, college students perceive that the AI algorithm, as the evaluator, has higher information transparency, and thus have higher perceptions of fairness.

### Study 3: the moderating effect of explanation on the influence of different evaluators on college students' perception of fairness

Study 3 followed the design of Study 1 and 2, intended to further investigate the main effect of different evaluators (AI algorithm vs.



teacher) on college students' perception of fairness and examine the moderating effect of explanation.

## Methods

### Participants

Taking college students as the basic criteria, the participants were randomly recruited on the online platform Credamo, and 5 were excluded for completing the questionnaire in too short a time or not meeting the response requirements. Specific demographic information of participants is as follows (see Table 5). The final sample included 145 participants ( $M_{age} = 21.31$ ,  $SD_{age} = 1.67$ ), of which 56 were male and 89 were female. Freshmen, sophomores, juniors, and seniors comprised 5.5, 26.9, 41.4, and 26.2% of the participant pool, respectively. These participants represented a diverse range of majors, including literature, engineering, law, medicine, and others, spanning 10 different categories. Of the subjects, 33.8% were hailed in China's double-first-class universities, while the remaining 66.2% were from ordinary universities.

### Design

The study adopted a mixed design of 2 (evaluator: AI algorithm/teacher)  $\times$  2 (with explanation/without explanation). The independent variable was the evaluator (within-subject design), and the dependent variable was students' self-reported perception of fairness. More specifically, in this study, each participant was asked to read materials about the AI algorithm and the teacher's evaluation, but one group received a situation with an explanation while the other group received a situation without an explanation. By comparing the differences between these two groups, the moderating effect of explanation could be examined.

## Materials and variable measurement

The descriptive situational material and the manipulation of evaluators and measurement employed in Study 1 were maintained for consistency in this study. Additionally, an explanatory variable was introduced into the material. To illustrate with formative evaluation as an example, it is important to note that in China, university courses typically involve two grades. The first grade is the final exam grade, which reflects the comprehensive learning outcomes of the student for the entire semester and can be considered a summative evaluation. The second part is the regular grade, which takes into account classroom performance, coursework, and other factors, and captures the student's performance during the learning process. The first half of the reading material was the same for both groups: "All optional courses' scores at your university consist of a regular grade and a final exam grade. The final exams are all objective questions with standard answers; the regular grades are given entirely by the teacher of the course." "All optional courses' scores at your university consist of a regular grade and a final exam grade. The final exams are all objective questions with standard answers; the regular grades are given entirely by the AI evaluation system applied by the university." And then, the materials from the group of participants for whom explanations were provided were "After the course is over, the teacher scores your class performance and provides a note to explain the basis for giving that grade." and "After the course is over, the AI evaluation system scores your class performance and provides a note to explain the basis for giving that grade." However, the materials from the group without explanation were "After the course was over, the teacher scores your class performance, but does not provide any explanation of the basis for the grade." and "After the course was over, the AI evaluation system scores your class performance, but does not provide any explanation of the basis for the grade."

### Procedures

The participants in each group were asked to read three descriptive situational materials in turn, and intuitively rated the fairness of the AI algorithm as the evaluator and the teacher as the evaluator, then completed the scale of information transparency. Finally, they filled in the demographic information.

## Results

### Descriptive statistics and correlation analysis

The correlation analysis of the main variables of Study 3 was performed and the results are shown in Table 6. The results showed that there was a significant negative correlation between the evaluator and the perception of fairness of the three types of evaluation ( $r = -0.18$ ,  $p < 0.01$ ;  $r = -0.14$ ,  $p < 0.01$ ;  $r = -0.10$ ,  $p < 0.5$ ). Besides, there was a significant positive correlation between the explanation and the perception of fairness of the three types of evaluation ( $r = 0.61$ ,  $p < 0.01$ ;  $r = 0.58$ ,  $p < 0.01$ ;  $r = 0.59$ ,  $p < 0.01$ ). The above analysis results are in line with the theoretical expectations of this study that the presence or absence of explanation will influence the evaluators' perception of fairness.

### Test for moderating effect

With the evaluator as the independent variable, the perception of fairness of the three types of evaluation as the dependent variable, and

TABLE 5 Demographic information of participants in Study 3.

Demographic characteristic		Frequency	Percentage	Demographic characteristic		Frequency	Percentage
Gender	Female	56	38.6%	Major	Law	6	4.1%
	Male	89	61.4%		Engineering	41	28.3%
Age	18	4	2.8%		Management	23	15.9%
	19	12	8.3%		Education	9	6.2%
	20	31	21.4%		Economics	12	8.3%
	21	37	25.5%		Science	26	17.9%
	22	31	21.4%		History	3	2.1%
	23	20	13.8%		Literature	15	10.3%
	24	4	2.8%		Medicine	6	4.1%
	25	4	2.8%		Arts	4	2.8%
	26	2	1.4%				
Grade	Freshmen	8	5.5%	University type	Double-first-class	49	33.8%
	Sophomore	39	26.9%		Ordinary	96	66.2%
	Junior	60	41.4%				
	Senior	38	26.2%				

TABLE 6 Descriptive statistics of variables and correlation coefficient matrix in Study 3.

Variable	$M \pm SD$	1	2	3	4	5
1. Type of evaluator	1.50 ± 0.50					
2. Explanation	1.50 ± 0.50	0.00				
3. Perception of fairness in diagnostic evaluation	4.94 ± 1.51	-0.18**	0.61**			
4. Perception of fairness in formative evaluation	4.98 ± 1.50	-0.14**	0.58**	0.74**		
5. Perception of fairness in summative evaluation	5.06 ± 1.41	-0.10*	0.59**	0.74**	0.76**	

Coding of evaluator types: AI evaluation = 1, Teacher = 2; coding of explanation: Without explanation = 1, With explanation = 2.  
 \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

the explanation as the moderator variable, Model 1 in the PROCESS program was adopted to test for moderating effects after all continuous variables were centralized.

The results showed that the interaction between the evaluator and the explanation could have a significant positive effect on the perception of fairness in diagnostic evaluation ( $\beta = 0.43$ ,  $SE = 0.19$ ,  $p < 0.05$ ), and the specific moderating effect was shown in Figure 6. Moreover, according to the results, when there was no explanation, there was a significant difference between the student's perception of fairness brought by the AI algorithm and that brought by the teacher ( $M_{AI} = 4.40$ ,  $M_{Teacher} = 3.63$ ,  $t(145) = -5.60$ ,  $p < 0.001$ ). However, when there was an explanation, the influence of the evaluator on the perception of fairness diminished, and the difference between the perception of fairness under the two evaluators was no longer significant ( $M_{AI} = 6.03$ ,  $M_{Teacher} = 5.70$ ,  $t(145) = 2.42$ ,  $p > 0.05$ ).

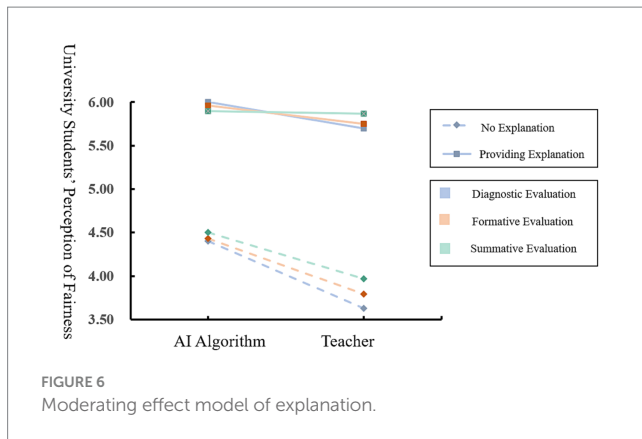
In formative evaluation, the interaction between the evaluator and the explanation also had a significant positive effect on the perception of fairness ( $\beta = 0.43$ ,  $SE = 0.20$ ,  $p < 0.05$ ). When there was no explanation, the student's perception of fairness brought by the AI algorithm was significantly higher than that brought by the teacher ( $M_{AI} = 4.43$ ,  $M_{Teacher} = 3.79$ ,  $t(145) = -4.57$ ,  $p < 0.001$ ). However, when there was an explanation, the difference between the perception of fairness under the two evaluators was no longer significant ( $M_{AI} = 5.96$ ,  $M_{Teacher} = 5.75$ ,  $t(145) = -1.47$ ,  $p > 0.05$ ).

In summative evaluation, the interaction between the evaluator and the explanation still had a significant positive effect on the perception of fairness ( $\beta = 0.50$ ,  $SE = 0.19$ ,  $p < 0.01$ ). When there was no explanation, the student's perception of fairness brought by the AI algorithm was significantly higher than that brought by the teacher ( $M_{AI} = 4.50$ ,  $M_{Teacher} = 3.97$ ,  $t(145) = -4.01$ ,  $p < 0.001$ ). However, when there was an explanation, the difference was no longer significant ( $M_{AI} = 5.90$ ,  $M_{Teacher} = 5.87$ ,  $t(145) = -1.47$ ,  $p > 0.05$ ).

The results of Study 3 indicated that the presence or absence of an explanation moderated the effect of the evaluator on college students' perception of fairness, regardless of the type of evaluation. To be more specific, the effect of the evaluator on college students' perception of fairness was insignificant when an explanation is present and produced a significant difference in the effect of the evaluator on college students' perception of fairness when no explanation was present.

## Discussion

At the theoretical level, this study has made substantial contributions to the research on the fairness of AI algorithmic evaluation. Firstly, Study 1 findings augment our understanding of how evaluation subjects perceive fairness in the context of digital education. While most existing studies have approached educational



evaluation fairness from a social justice perspective or focused on designing pedagogically fair assessments, limited research has examined the influence of evaluation subjects on fairness perception through an organizational justice lens (Thomas and Madison, 2010; McArthur, 2016; Zhu, 2018; Li and Xin, 2019; Roshid et al., 2022). This study extends the application of organizational justice theory to the domain of educational evaluation and sheds light on the informational fairness concerns inherent to diverse evaluators, that is, college students tend to attribute higher perceptions of fairness to AI algorithms than to teachers as evaluators. Secondly, Study 2 results hold significant implications for comprehending the process through which students develop perceptions of fairness in a digital educational environment. This study underscores the mediating role of information transparency. The evaluation of Artificial Intelligence engenders heightened perceptions of fairness due to students' perception of it as an algorithmic program characterized by greater information transparency. Recent scholarship has increasingly focused on the "black box" problem in AI, highlighting the opacity surrounding AI algorithms and their inner workings (Orlikowski and Scott, 2014; Lee et al., 2015). However, this study offers a contrasting perspective, suggesting that students' perceptions of the information transparency of AI algorithm evaluations tend to be in the opposite direction and are not or only slightly influenced by their concerns about the algorithms' transparency. Evidently, under the guidance of AI algorithms, educational evaluation becomes more transparent in students' views owing to explicit assessment criteria, well-defined processes, and precise outcomes. Lastly, Study 3 findings reaffirm previous research by substantiating the positive impact of explanations on fostering perceptions of fairness, thereby providing valuable guidance for the effective implementation of educational evaluation within the digital education milieu. This study significantly advances our understanding of fairness perception in AI algorithmic evaluation, expanding theoretical frameworks, uncovering informational fairness issues, and offering practical guidelines and the knowledge base for conducting fair educational evaluations in the era of digitization.

Beyond the theoretical implications of this research, further discourse is warranted on potential areas of disagreement within the article. Initially, this investigation applies organizational justice theory to the realm of educational justice, a method not entirely innovative. Numerous preceding studies have employed organizational justice theory to scrutinize equity concerns in education. For instance, Chory-Assad's (2002) research on classroom justice conceptualized the class as an organization, effectively integrating organizational justice theory into the classroom context. Similarly, we have

transposed the concept of performance appraisal into the educational sphere, despite their distinct natures. Performance appraisal of employees constitutes a summative evaluation based on work output, whereas academic performance assessment can be categorized into diagnostic, formative, and summative evaluations (Bloom, 1969, 1987), contingent upon its timing and role in teaching and learning processes. The primary distinction between these two resides in the evaluation type or context. However, they share considerable similarities. Both are evaluative measures of individuals' achievements, typically quantified. Furthermore, the social relationships underpinning these evaluations are akin. Performance appraisal hinges on superior-subordinate dynamics, paralleling the teacher-student relationship in Chinese academic appraisal. Lastly, both employees and students exhibit significant concern regarding evaluation fairness. It is therefore justified to transpose organizational justice theory to educational justice and recast performance appraisal as academic assessment.

The second salient issue necessitates acknowledging that the evaluation type may influence students' fairness perceptions. However, we did not consider it a variable but rather a research context. This decision was predicated on several key reasons. Primarily, our focus was on students' fairness perceptions, not the evaluations themselves. Consequently, we sought to measure students' fairness perceptions across different contexts to isolate the impact of assessment contexts on fairness perceptions. Our findings suggest that students' fairness judgments of AI algorithmic evaluations and teacher evaluations exhibit similar trends across all three assessment contexts, indicating that the evaluation type does not significantly influence students' fairness perceptions. Secondly, we aimed to underscore the subjective nature of assessment contexts rather than differences in assessment types. Despite variations in timing and roles of diagnostic, formative, and summative evaluations, all three essentially constitute subjective assessments. The contrast between human evaluators (teachers) and machine evaluators (AI algorithms) introduces a significant difference in subjective evaluation, likely contributing to variations in students' fairness perceptions. It is this evaluative discrepancy, rather than the specific assessment context, that accounts for differences in students' fairness perceptions. Consequently, we opted to use different evaluation types as contexts in our study, rather than as variables.

Furthermore, in the mediator analysis of Study 2, the correlation between "type of evaluator" and "perception of fairness" exhibits a significantly higher magnitude ( $-0.49$ ) compared to the direct correlation between the two variables ( $-0.28$ ) in the absence of a mediator variable. To gain a more comprehensive understanding of this unexpected effect, future research should investigate the precise mechanisms by which the mediator variable influences the relationship between "type of rater" and "perception of fairness." This exploration could involve examining potential mediators such as trust, communication, or perceived competence of the rater. By delving into these variables, we can enhance our comprehension of their contributions to the perception of fairness. In summary, the counterintuitive effect observed in the mediator analysis can be elucidated by considering the mediator variable's role as either a confounding or intervening variable. Further research is necessary to delve into the specific mechanisms underlying this relationship and identify potential mediators that may help clarify the observed results.

At the practical level, this study provides valuable insights into potential issues in current higher educational evaluation practices. Study 1 reveals that students tend to perceive AI algorithms as fairer

when it comes to subjective evaluations. However, it is important to acknowledge that although the integration of new technologies like AI algorithms into the evaluation system is an important trend in educational evaluation reform, AI algorithmic evaluation has yet to achieve widespread adoption, particularly in China. Despite this, Chinese students still express approval of AI algorithmic evaluation, which highlights the significant challenges faced by teachers as the evaluation subjects in terms of fairness. Taking formative assessment as a case in point, it is intended to facilitate direct communication between students and educators. However, in practice, this interaction often resembles a dialog between students and authority figures, rather than a genuine exchange. This relational perception inadvertently hinders students' active engagement in the assessment and feedback process, subsequently diminishing their perceptions of fairness and trust in teacher evaluations. The findings from Studies 2 and 3 indicate that increasing transparency in information and providing comprehensive explanations of the evaluation process can partially improve students' perceptions of fairness. It is crucial to recognize that questioning the fairness of teacher evaluation is not a flaw inherent to the evaluators themselves, but rather a problem within the evaluation process. Procedural justice and information justice perspectives suggest that enhancing students' sense of fairness does not lie solely in replacing the evaluation subjects, but rather in clarifying the roles and responsibilities of different evaluation participants, while simultaneously improving the scientific rigor, transparency, and interpretability of the entire evaluation process. By addressing these aspects, the source of students' perceived unfairness can be mitigated.

In this regard, several schools have implemented noteworthy initiatives in practice. For instance, a senior high school in Hefei, China, has adopted an artificial intelligence system to capture data generated by students during physical education tests and perform detailed analyses to support teachers in delivering accurate instruction. As an illustration, in the case of long jump assessments, the system assesses whether a student passes or fails while simultaneously capturing screenshots of key movements and recording videos of the complete test. Furthermore, the system provides personalized comments and suggestions for improvement alongside the test results, aiding students in their endeavors to enhance performance. Such comprehensive, transparent, and interpretable evaluation outcomes naturally foster higher perceptions of fairness among students. Furthermore, to enhance the fairness of technology-assisted educational evaluation, it is essential to remain vigilant regarding the potential risks associated with its implementation. On one hand, excessive reliance on technology for evaluation can lead to the phenomenon known as "algorithmic hegemony," which diminishes the interpretability of evaluation results (Zhang and Qi, 2021). AI algorithms, in particular, involve intricate data collection and processing procedures that are challenging for individuals to comprehend within limited time frames. With a monopoly on technical information, the evaluation results wield an unquestionable authority over individuals, essentially establishing algorithmic hegemony (O'Neil, 2016). For example, in the context of US college admissions, the use of AI algorithms can more accurately predict students' abilities and facilitate the efficient allocation of financial aid. A case study by Othot (2018) reported that their analysis assisted the New Jersey Institute of Technology (NJIT) in enrolling 173 students while staying within budgetary constraints, resulting in improved recruitment efficiency and cost savings. However, when utilized in educational admissions, these algorithms often operate

under the premise of a single theoretical construct. This means they are engineered to focus on one specific aspect or objective of education. In such cases, the algorithm might be programmed to prioritize the "strength of candidate" as its primary metric. This could encompass evaluating factors such as academic accomplishments, standardized test scores, or extracurricular involvement. As a result, the algorithm would rank applicants based on these parameters, potentially offering more admission opportunities to those who excel in these areas. Nonetheless, this singular focus can lead to a narrow interpretation of educational goals. While emphasizing the strength of a candidate, the algorithm might overlook other vital facets of education, such as access and equity. Educational access signifies the opportunity for all individuals, regardless of their background or circumstances, to pursue higher education. Equity, conversely, involves ensuring fairness in treatment, equality of opportunity, and equitable access to information and resources. By concentrating exclusively on the strength of a candidate, the AI algorithm may unintentionally exacerbate disparities in educational access and equity. For example, it might favor applicants from privileged backgrounds who have had more opportunities to develop strong academic or extracurricular profiles, while disadvantaging those from less privileged backgrounds. Therefore, while AI algorithms can streamline the admissions process and identify high-achieving candidates, it's crucial to consider how their design and implementation might influence broader educational objectives. Future versions of these algorithms should aim to balance multiple educational goals, fostering not only excellence but also access and equity (Liu et al., 2023).

On the other hand, ethical considerations surrounding the use of new technologies must be addressed to ensure the acceptability of the evaluation process. In the case of AI algorithms, evaluations rely on extensive data collection, storage, and analysis, often involving technologies such as facial recognition and body recognition. While these technologies provide robust support for more accurate and scientific evaluations, they also raise concerns about information leakage and privacy infringement (Lu and Gao, 2023). Therefore, before the integration of new technologies into the evaluation system, privacy regulations should be established to safeguard the security and confidentiality of users' information.

This study extends organizational justice theory to educational evaluation, finding that college students perceive AI algorithms as fairer than teachers. The study highlights the importance of information transparency in shaping students' perceptions of fairness and contradicts concerns about the "black box" problem in AI. It also confirms that explanations positively impact fairness perceptions. While acknowledging practical challenges in implementing AI in educational evaluation, the study suggests that transparency and comprehensive explanations can enhance fairness perceptions.

## Conclusion

This study contributes to the understanding of fairness perception in AI algorithmic evaluation in the digital context. The findings demonstrate that the evaluator, whether it is an AI algorithm or a teacher, significantly influences college students' perceptions of fairness in diagnostic, formative, and summative evaluations. Specifically, students perceive AI algorithms as fairer evaluators compared to teachers, with formative evaluation showing the strongest

effect. This perception is mediated by the perceived information transparency of AI algorithms, which is perceived to be higher than that of teachers. Additionally, the provision of explanations for the evaluation process moderates the impact of the evaluator on fairness perception. Without explanations, students are more likely to perceive AI evaluations as fairer than teacher evaluations. However, when clear explanations are provided, the influence of the evaluator on fairness perception weakens, leading to no significant difference in perceived fairness between AI algorithms and teachers among college students.

These findings have significant practical implications for educational institutions and policymakers. Recognizing the impact of the evaluator on fairness perception, it is imperative to ensure that both AI algorithms and teachers adhere to the principles of fairness in educational evaluation. This necessitates fostering information transparency in AI algorithmic evaluations and providing comprehensive explanations for the evaluation process. By doing so, educational institutions can enhance students' perceptions of fairness and promote trust in the evaluation system. From an institutional standpoint, to ensure fair educational assessment, relevant policies, and guidelines should be formulated and implemented. These policies and guidelines should explicitly outline the fairness principles that teachers and AI algorithms must adhere to during the evaluation process. For instance, educational institutions can require teachers to openly disclose grading criteria and evaluation procedures to increase transparency (Carless, 2009; Mhlanga, 2023). Simultaneously, institutions can support teachers utilizing AI algorithms for evaluation by providing technical training and support to ensure they can correctly utilize and explain algorithmic results. Additionally, educational institutions can establish a feedback mechanism allowing students to provide input on the evaluation process, facilitating prompt correction of any potential biases or unfair practices. For policymakers, it is crucial to address fairness concerns in educational assessment and enact corresponding policies and regulations to safeguard student rights. Policymakers can promote standardization and regulation of AI algorithmic assessments, ensuring that the evaluation process possesses verifiability and reproducibility. Furthermore, policymakers should develop measures for privacy protection to prevent personal data disclosure and misuse. In the EU's Ethics Guidelines for Trustworthy AI, the responsible development of AI is addressed, highlighting its potential to benefit individuals, helping people track their personal data and increase access to education, thus supporting their right to education (European Commission, 2021). Encouraging collaboration between educational institutions and AI algorithm developers, policymakers can work together to establish guidelines and best practices to ensure the fairness and transparency of the evaluation system. These measures will contribute to strengthening students' perceptions of fairness in educational assessment and building trust in the evaluation system. Moreover, they will provide an effective means to monitor and improve the quality and fairness of educational assessment. By ensuring that the evaluation process aligns with principles of fairness, educational institutions, and policymakers can provide students with a just, objective, and reliable learning environment, fostering their academic achievement and development.

While this study has yielded promising findings, it is important to acknowledge the inherent limitations that may have influenced the results. One limitation arises from the data collection method chosen

due to the constraints imposed by the COVID-19 pandemic. The use of an online platform for data collection introduces potential sources of bias, such as sample selection bias and self-selection bias. Additionally, relying on online surveys may introduce technological limitations and digital divide issues, as not all individuals may have had equal access or willingness to participate. Another limitation lies in the scope of variables considered in exploring the relationship between evaluation subjects and perceptions of fairness. Due to feasibility constraints, only two variables were examined in this study. However, it is crucial to recognize that the phenomenon of AI algorithmic evaluation and its impact on perceptions of fairness is multifaceted and complex, warranting further investigation into additional influencing factors.

As we look toward the future, it is important for further research to address the limitations identified in our study and advance our understanding of the topic. One potential avenue for exploration is to expand the sample size to include college students from diverse national and cultural backgrounds. By doing so, researchers can gain valuable insights into the cross-cultural and emotional variations in how AI algorithms and teacher evaluations are perceived in terms of fairness. Including participants from different countries and cultural contexts would allow for a more comprehensive examination of how cultural factors may influence individuals' perceptions of fairness. It would provide an opportunity to explore whether there are cultural differences in expectations, values, or norms that shape how individuals evaluate the fairness of AI algorithms and teacher assessments. Additionally, considering emotional variations across cultures could shed light on how emotions impact fairness perceptions in educational evaluation settings. For instance, empirical research has demonstrated that students who appraise environmental cues positively (such as the professor's demeanor and examination settings) are more inclined to conceptualize examinations as opportunities rather than adversities. This positive appraisal extends to their perceptions of other individuals involved in the assessment process (for example, examiners), culminating in an enhanced sense of fairness. Conversely, the presence of negative affective states, such as disappointment or frustration, is associated with a diminished perception of fairness (Ashton-James and Ashkanasy, 2005; Butucescu and Iliescu, 2022). By incorporating a diverse range of participants, future studies can contribute to a more nuanced understanding of the complex interplay between culture, emotions, and fairness perceptions. This expanded scope would enhance the generalizability of findings and help inform the development of more culturally sensitive and equitable educational evaluation practices. Overall, by expanding the sample to encompass college students from diverse national and cultural backgrounds, future research has the potential to deepen our understanding of cross-cultural and emotional variations in the perceived fairness of AI algorithms and teacher evaluations, ultimately contributing to more inclusive and effective educational assessment approaches (Shepherd and Willis-Esqueda, 2018; Ho et al., 2022). Moreover, incorporating more variables into the research design would enable a comprehensive exploration of the mechanisms underlying the effects of AI algorithmic evaluation on fairness perception. These variables could include factors like the transparency and comprehensibility of AI algorithms, the timing and methods of explanations, and individual characteristics like attribution style and

trust in technology. In this study, the concept of information transparency in artificial intelligence (AI) algorithms is delineated as the degree to which details pertaining to the evaluation process are disclosed and made accessible to users, drawing on the definitions provided by Florini (2007), Tielenburg (2018), and Wang and Ren (2022). Nevertheless, it is imperative to recognize that AI algorithms are intended to fulfill the objectives of transparency in a context that transcends merely being available and accessible. Various scholars have also highlighted that transparency encompasses the extent to which the internal workings of a system, including the capacity to grasp and articulate the logic underpinning its decisions, are observable and intelligible to human users (Zerilli et al., 2019; De Freitas et al., 2023). Informed by these insights, our subsequent research endeavors will aim to refine the definition of transparency by incorporating additional pertinent dimensions such as explainability and comprehensibility into the AI decision-making framework. This augmented conceptual framework aspires to capture the intricate essence of transparency and provide a more nuanced understanding of the term. Furthermore, future studies can also explore the practical implications of AI algorithmic evaluations and fairness perceptions on students' academic performance, motivation, and attitudes toward learning. These investigations can shed light on the potential educational implications and inform the development of effective practices and policies in the context of technology-assisted educational evaluation.

This study contributes to the growing body of literature on fairness perception in AI algorithmic evaluation. By examining the impact of the evaluator, information transparency, and explanations, the study provides valuable insights into the factors influencing college students' perceptions of fairness. These findings have theoretical and practical implications for creating a fairer educational evaluation environment, emphasizing the importance of transparency and explanation provision in promoting fairness of AI algorithmic evaluations. Moreover, this study highlights the need for further research on the long-term effects of AI algorithmic evaluation in the digital education context. Future studies could examine the implications of AI algorithms for student motivation, engagement, and academic performance. Additionally, considering the ethical concerns raised by the use of AI algorithms in educational evaluation, further research should explore strategies to address issues such as algorithmic bias, privacy protection, and accountability. By addressing these issues, we can ensure that educational evaluation remains fair, transparent, and effective in promoting student learning and development in the digital education era.

## References

- Adadi, A., and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6, 52138–52160. doi: 10.1109/ACCESS.2018.2870052
- Adams, J. S. (1965). Inequity in social exchange. *Adv. Exp. Soc. Psychol.* 2, 267–299. doi: 10.1016/S0065-2601(08)60108-2
- Aguinis, H., and Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organ. Res. Methods* 17, 351–371. doi: 10.1177/1094428114547952
- Ahmad, S. F., Alam, M. M., Rahmat, M. K., Mubarik, M. S., and Hyder, S. I. (2022). Academic and administrative role of artificial intelligence in education. *Sustainability* 14:1101. doi: 10.3390/su14031101
- Alterman, V., Bamberger, P. A., Wang, M., Koopmann, J., Belogolovsky, E., and Shi, J. (2021). Best not to know: pay secrecy, employee voluntary turnover, and the conditioning effect of distributive justice. *Acad. Manag. J.* 64, 482–508. doi: 10.5465/amj.2019.0231
- Araujo, T., Helberger, N., Kruikemeier, S., and De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Soc.* 35, 611–623. doi: 10.1007/s00146-019-00931-w
- Ashton-James, C. E., and Ashkanasy, N. M. (2005). What lies beneath? A process analysis of affective events theory. *Effect Affect Organ. Sett.* 1, 23–46. doi: 10.1016/S1746-9791(05)01102-8
- Awad, N. F., and Krishnan, M. S. (2006). The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Q.* 30, 13–28. doi: 10.2307/25148715

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

FC, JM, JZ, and TH contributed to conception and design of the study. JM organized the database. JM and YW performed the statistical analysis. FC and JM wrote the first draft of the manuscript. YW, JZ, and TH wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## Funding

This work was Supported by the Fundamental Research Funds for the Central Universities (item number 2020JJ031).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Baartman, L. K., Bastiaens, T. J., Kirschner, P. A., and Van der Vleuten, C. P. (2007). Teachers' opinions on quality criteria for competency assessment programs. *Teach. Educ.* 23, 857–867. doi: 10.1016/j.tate.2006.04.043
- Barsky, A., Kaplan, S. A., and Beal, D. J. (2011). Just feelings? The role of affect in the formation of organizational fairness judgments. *J. Manag.* 37, 248–279. doi: 10.1177/0149206310376325
- Belogolovsky, E., and Bamberger, P. A. (2014). Signaling in secret: pay for performance and the incentive and sorting effects of pay secrecy. *Acad. Manag. J.* 57, 1706–1733. doi: 10.5465/amj.2012.0937
- Bies, R. J. (2005). "Are procedural and interactional justice conceptually distinct?" in *The handbook of organizational justice*. eds. J. Greenberg and J. A. Colquitt (Mahwah, NJ: Erlbaum)
- Bies, R. J., and Moag, J. F. (1986). Interactional justice: communication criteria of fairness. *Res. Negotiation Organ.* 1, 43–55.
- Bies, R. J., and Shapiro, D. L. (1987). Interactional fairness judgments: the influence of causal accounts. *Soc. Justice Res* 1, 199–218. doi: 10.1007/BF01048016
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. *Teach. Coll. Rec.* 70, 26–50. doi: 10.1177/016146816907001003
- Bloom, B. S. (1987). *Education evaluation* (Qiu Yuan & Wang Gang, Trans.). Shanghai: East China Normal University Press.
- Burrell, J. (2016). How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data Soc.* 3, 251–212. doi: 10.1177/2053951715622512
- Butucescu, A., and Iliescu, D. (2022). Fairness perceptions in educational assessment: the role of positive and negative affect. *Educ. Stud.* 48, 310–322. doi: 10.1080/03055698.2020.1753176
- Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assess. Eval. High. Educ.* 34, 79–89. doi: 10.1080/02602930801895786
- Chan Kim, W., and Mauborgne, R. (1998). Procedural justice, strategic decision making, and the knowledge economy. *Strateg. Manag. J.* 19, 323–338. doi: 10.1002/(SIC)11097-0266(199804)19:4<323::AID-SMJ976>3.0.CO;2-F
- Chory-Assad, R. M. (2002). Classroom justice: perceptions of fairness as a predictor of student motivation, learning, and aggression. *Commun. Q.* 50, 58–77. doi: 10.1080/01463370209385646
- Clifford, S., Iyengar, V., Cabeza, R., and Sinnott-Armstrong, W. (2015). Moral foundations vignettes: a standardized stimulus database of scenarios based on moral foundations theory. *Behav. Res. Methods* 47, 1178–1198. doi: 10.3758/s13428-014-0551-2
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: a construct validation of a measure. *J. Appl. Psychol.* 86, 386–400. doi: 10.1037/0021-9010.86.3.386
- Colquitt, J. A., and Shaw, J. C. (2013). "How should organizational justice be measured?" in *Handbook of organizational justice*. eds. J. Greenberg and J. A. Colquitt (London: Psychology Press)
- Cong, Y. N., Wang, Z. Y., and Jingqin, Z. (2022). Insights into dataset and algorithm related problems in artificial intelligence for law. *Comp. Sci.* 49, 74–79.
- Cropanzano, R., and Greenberg, J. (1997). Progress in organizational justice: tunneling through the maze. *Int. Rev. Ind. Organ. Psychol.* 12, 317–372.
- Damme, D. V. (2001). Quality issues in the internationalization of higher education. *High. Educ.* 41, 415–441. doi: 10.1023/A:1017598422297
- Daud, M. (2022). Artificial intelligence in the Malaysian legal system: issues, challenges and way forward. *INSAF J. Malays. Bar* 39, 1–24.
- De Clercq, D., and Pereira, R. (2020). Knowledge-sharing efforts and employee creative behavior: the invigorating roles of passion for work, time sufficiency and procedural justice. *J. Knowl. Manag.* 24, 1131–1155. doi: 10.1108/JKM-06-2019-0274
- De Freitas, J., Agarwal, S., Schmitt, B., and Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. *Nat. Hum. Behav.* 7, 1845–1854. doi: 10.1038/s41562-023-01734-2
- Dennis, M. J. (2018). Artificial intelligence and recruitment, admission, progression, and retention. *Enroll. Manag. Rep.* 22, 1–3. doi: 10.1002/emt.30479
- Dochy, F. J. R. C., Segers, M., and Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Stud. High. Educ.* 24, 331–350. doi: 10.1080/03075079912331379935
- Dulebohn, J. H., and Ferris, G. R. (1999). The role of influence tactics in perceptions of performance evaluations' fairness. *Acad. Manag. J.* 42, 288–303. doi: 10.2307/256920
- Ekici, H., Yücel, E., and Cesur, S. (2023). Deciding between moral priorities and COVID-19 avoiding behaviors: a moral foundations vignette study. *Curr. Psychol.* 42, 5922–5938. doi: 10.1007/s12144-021-01941-y
- Engelmann, J. M., and Tomasello, M. (2019). Children's sense of fairness as equal respect. *Trends Cogn. Sci.* 23, 454–463. doi: 10.1016/j.tics.2019.03.001
- European Commission. (2021). Requirements of Trustworthy AI. Available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.html>.
- Fatt, C. K., Khin, E. W. S., and Heng, T. N. (2010). The impact of organizational justice on employee's job satisfaction: the Malaysian companies' perspectives. *Am. J. Econ. Bus. Admin.* 2, 56–63. doi: 10.3844/ajebasp.2010.56.63
- Finch, J. (1987). The vignette technique in survey research. *Sociology* 21, 105–114. doi: 10.1177/0038038587021001008
- Florini, A. (2007). *The right to know: transparency for an open world*. New York: Columbia University Press.
- Gardner, W. L., Avolio, B. J., Luthans, F., May, D. R., and Walumbwa, F. (2005). "Can you see the real me?" A self-based model of authentic leader and follower development. *Leadersh Q.* 16, 343–372.
- Goodman, B., and Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Mag.* 38, 50–57. doi: 10.1609/aimag.v38i3.2741
- Greenberg, J., and Cropanzano, R. (1993). "The social side of fairness: Interpersonal and informational classes of organizational justice," in *Justice in the workplace: approaching fairness in human resource management*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gu, F., Nolan, J., and Rowley, C. (2020). Organizational justice in Chinese banks: understanding the variable influence of guanxi on perceptions of fairness in performance appraisal. *Asia Pac. Bus. Rev.* 26, 169–189. doi: 10.1080/13602381.2019.1700622
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Comp. Surv.* 51, 1–42. doi: 10.1145/3236009
- Gupta, V., and Kumar, S. (2012). Impact of performance appraisal justice on employee engagement: a study of Indian professionals. *Empl. Relat.* 35, 61–78. doi: 10.1108/01425451311279410
- He, W., Fehr, R., Yam, K. C., Long, L. R., and Hao, P. (2017). Interactional justice, leader-member exchange, and employee performance: examining the moderating role of justice differentiation. *J. Organ. Behav.* 38, 537–557. doi: 10.1002/job.2133
- Hewett, R., and Leroy, H. (2019). Well it's only fair: how perceptions of manager discretion in bonus allocation affect intrinsic motivation. *J. Manag. Stud.* 56, 1105–1137. doi: 10.1111/joms.12445
- Ho, M. T., Mantello, P., Ghotbi, N., Nguyen, M. H., Nguyen, H. K. T., and Vuong, Q. H. (2022). Rethinking technological acceptance in the age of emotional AI: surveying gen Z (Zoomer) attitudes toward non-conscious data collection. *Technol. Soc.* 70:102011. doi: 10.1016/j.techsoc.2022.102011
- Hollensbe, E. C., Khazanchi, S., and Masterson, S. S. (2008). How do I assess if my supervisor and organization are fair? Identifying the rules underlying entity-based justice perceptions. *Acad. Manag. J.* 51, 1099–1116. doi: 10.5465/amj.2008.35732600
- Hyman, M. R., and Steiner, S. D. (1996). The vignette method in business ethics research: current uses, limitations, and recommendations. *Studies* 20, 74–100.
- James, A., Kashyap, M., Chua, Y. H. V., Maszczyk, T., Núñez, A. M., Bull, R., et al. (2018). Inferring the climate in classrooms from audio and video recordings: a machine learning approach. In 2018 IEEE international conference on teaching, assessment, and learning for engineering (TALE) (983–988). IEEE.
- Jihui, D., and Zhilin, H. (2015). Research on the role of information disclosure in promoting the rule of law in universities. *China High. Educ.* 51, 42–44.
- Jongbloed, B., Vossensteyn, H., van Vught, F., and Westerheijden, D. F. (2018). Transparency in higher education: the emergence of a new perspective on higher education governance. *Euro. High. Educ. Area*, 441–454. doi: 10.1007/978-3-319-77407-7\_27
- Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2390–2395).
- Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119, 254–284. doi: 10.1037/0033-2909.119.2.254
- Lebovitz, S., Lifshitz-Assaf, H., and Levina, N. (2022). To engage or not to engage with AI for critical judgments: how professionals deal with opacity when using AI for medical diagnosis. *Organ. Sci.* 33, 126–148. doi: 10.1287/orsc.2021.1549
- Lee, M. K., Kusbit, D., Metsky, E., and Dabbish, L. (2015). Working with machines: the impact of algorithmic and data-driven management on human workers. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (1603–1612).
- Lemons, M. A., and Jones, C. A. (2001). Procedural justice in promotion decisions: using perceptions of fairness to build employee commitment. *J. Manag. Psychol.* 16, 268–281. doi: 10.1108/02683940110391517
- Leslie, D. (2019). Understanding artificial intelligence ethics and safety. arXiv:1906.05684 [Preprint].
- Leventhal, G. S., Karuza, J., and Fry, W. R. (1980). Beyond fairness: a theory of allocation preferences. *Just. Soc. Interact.* 3, 167–218.
- Li, H. X., and Xin, T. (2019). Analysis on the evaluation path of the new college entrance examination target. *J. Chin. Soc. Educ.* 40, 67–71.
- Lind, E. A. (2001). Fairness heuristic theory: justice judgments as pivotal cognitions in organizational relations. *Adv. Organ. Just.* 56, 56–88.
- Liu, Y., Eisingerich, A. B., Auh, S., Merlo, O., and Chun, H. E. H. (2015). Service firm performance transparency: how, when, and why does it pay off? *J. Serv. Res.* 18, 451–467. doi: 10.1177/1094670515584331



- Liu, Y., Long, L. R., and Ye, L. (2003). The unique effects on organizational justice on key outcomes. *J. Manag. World* 19, 126–132.
- Liu, L. T., Wang, S., Britton, T., and Abebe, R. (2023). Reimagining the machine learning life cycle to improve educational outcomes of students. *Proc. Natl. Acad. Sci.* 120:e2204781120. doi: 10.1073/pnas.2204781120
- Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm appreciation: people prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* 151, 90–103. doi: 10.1016/j.obhdp.2018.12.005
- Lu, X. G., and Gao, X. W. (2023). On the AI enabling education reform: development trend, risk inspection and countermeasures. *J. Chin. Soc. Educ.* 2, 48–54.
- Lundberg, S. M., Erion, G. G., Chen, H., DeGrave, A. J., Prutkin, J. M., Nair, B. G., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 56–67. doi: 10.1038/s42256-019-0138-9
- Luo, J., Bai, T., and Zhao, J. (2021). Generating Adversarial yet Inconspicuous Patches with a Single Image (Student Abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35, 15837–15838.
- Lv, X. J. (2010). A review of research on the psychological mechanisms underlying the formation of the sense of social justice. *J. Henan Norm. Univ.* 37, 27–31.
- Lv, X. J., and Liu, B. C. (2009). Research on the fairness psychology and behavior of college students—based on the perspective of sense of fairness in education. *J. Shanghai Jiao Tong Univ.* 72–78.
- Magner, N., Johnson, G. G., and Elfrink, J. (1994). Evidence on the relationship between procedural and distributive justice in performance appraisal and accounting faculty attitudes and performance. *J. Account. Educ.* 12, 325–341. doi: 10.1016/0748-5751(94)90025-6
- Mao, G., Zhou, Y. L., and He, W. T. (2020). Development trend of teaching evaluation theory under background of educational big data. *e-Educ. Res.* 41, 22–28.
- Marcinkowski, F., Kieslich, K., and Starke, C., & Lünich, M. (2020). Implications of AI (un-)fairness in higher education admissions: the effects of perceived AI (un-)fairness on exit, voice and organizational reputation. Proceedings of the 2020 FAT\* Conference on Fairness, Accountability, and Transparency, 122–130.
- Martins, R. M., and Von Wangenheim, G. C. (2022). Findings on teaching machine learning in high school: a ten-year systematic literature review. *Inform. Educ.* 22, 421–440. doi: 10.15388/infedu.2023.18
- McArthur, J. (2016). Assessment for social justice: the role of assessment in achieving social justice. *Assess. Eval. High. Educ.* 41, 967–981. doi: 10.1080/02602938.2015.1053429
- Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning (February 11, 2023)
- Ministry of Education of The people's Republic of China. (2023). Call for Global Partnership on Digital Education Development. Available at: [http://en.moe.gov.cn/news/press\\_releases/202302/t20230214\\_1044475.html](http://en.moe.gov.cn/news/press_releases/202302/t20230214_1044475.html)
- Niehoff, B. P., and Moorman, R. H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Acad. Manag. J.* 36, 527–556. doi: 10.2307/256591
- O'neil, C. (2016). Weapons of math destruction: how big data increases inequality and threatens democracy. *Sci. Am.* 315, 74–74.
- Orlikowski, W. J., and Scott, S. V. (2014). What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector. *Organ. Sci.* 25, 868–891. doi: 10.1287/orsc.2013.0877
- Othot. (2018). Using advanced analytics to plan, budget & allocate financial aid. Available at: <https://www.othot.com/case-studies/new-jersey-institute-of-technology> (accessed March 30, 2023).
- Paloma Sánchez, M., and Elena, S. (2006). Intellectual capital in universities: improving transparency and internal management. *J. Intellect. Cap.* 7, 529–548. doi: 10.1108/14691930610709158
- Pei, J. L., Liu, S. S., Zhong, C. Y., and Yifan, C. (2021). Can AI algorithmic decision-making improve employees' perception of procedural fairness? *For. Econ. Manag.* 43, 41–55.
- Pralhad, C. K., and Ramaswamy, V. (2002). The co-creation connection. *Strategy Business* 50–61.
- Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instrum. Comput.* 36, 717–731. doi: 10.3758/BF03206553
- Rasooli, A., Zandi, H., and DeLuca, C. (2019). Conceptualizing fairness in classroom assessment: exploring the value of organizational justice theory. *Assess. Educ.* 26, 584–611.
- Roshid, M. M., Sultana, S., Kabir, M. M. N., Jahan, A., Khan, R., and Haider, M. Z. (2022). Equity, fairness, and social justice in teaching and learning in higher education during the COVID-19 pandemic. *Asia Pac. J. Educ.* 1–22. doi: 10.1080/02188791.2022.2122024
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215. doi: 10.1038/s42256-019-0048-x
- Ruiji, R., Dong, W., Wang Shijin, H., and Guoping, L. T. (2018). Elegant sentence recognition for automated essay scoring. *J. Chin. Inform. Process.* 32, 88–97.
- Rupp, E. (2011). An employee-centered model of organizational justice and social responsibility. *Organ. Psychol. Rev.* 1, 72–94. doi: 10.1177/2041386610376255
- Ryu, G., and Hong, S. W. (2020). The mediating effect of trust in supervisors in the relationship between constructive performance feedback and perceived fairness of performance appraisal. *Public Perform. Manag. Rev.* 43, 871–888. doi: 10.1080/15309576.2019.1676274
- Schnackenberg, A. K., and Tomlinson, E. C. (2016). Organizational transparency: a new perspective on managing trust in organization-stakeholder relationships. *J. Manag.* 42, 1784–1810. doi: 10.1177/0149206314525202
- Shepherd, S. M., and Willis-Esqueda, C. (2018). Indigenous perspectives on violence risk assessment: a thematic analysis. *Punishment Soc.* 20, 599–627. doi: 10.1177/1462474517721485
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *Int. J. Hum. Comp. Stud.* 146:102551. doi: 10.1016/j.ijhcs.2020.102551
- Shulner-Tal, A., Kuflik, T., and Klinger, D. (2022). Fairness, explainability and in-between: understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics Inf. Technol.* 24:2. doi: 10.1007/s10676-022-09623-4
- Sudin, S. (2011). Fairness of and satisfaction with performance appraisal process. *J. Glob. Manag.* 2, 66–83.
- Sun, B. (2022). Explainable AI: the key to unlocking the “black box” of future digital education. *Chin. J. ICT Educ.* 28, 3–4.
- Sun, W., and Huang, P. L. (2004). A review of equity theory research. *Sci. Technol. Manag. Res.* 24, 102–104.
- Sweeney, P. D., and McFarlin, D. B. (1993). Workers' evaluations of the “ends” and the “means”: an examination of four models of distributive and procedural justice. *Organ. Behav. Hum. Decis. Process.* 55, 23–40. doi: 10.1006/obhd.1993.1022
- Thomas, V. G., and Madison, A. (2010). Integration of social justice into the teaching of evaluation. *Am. J. Eval.* 31, 570–583. doi: 10.1177/1098214010368426
- Tielenburg, D. S. (2018). The dark sides of transparency: rethinking information disclosure as a social praxis (Master's thesis).
- United Nations. (2022). Assuring and improving quality public digital learning for all. Available at: <https://www.un.org/en/transforming-education-summit/digital-learning-all>.
- Van Berkel, N., Goncalves, J., Russo, D., Hosio, S., and Skov, M. B. (2021). Effect of information presentation on fairness perceptions of machine learning predictors. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (1–13).
- Wang, W. (2016). Interpretation of school evaluation results: a study based on hermeneutics. *J. Chin. Soc. Educ.* 273, 54–58.
- Wang, S. Y., and Ren, Y. D. (2022). The ethical design of transparency for artificial intelligence autonomous decision system. *Sci. Econ. Soc.* 40, 47–58.
- Wang, L., and Wang, K. (2020). Artificial intelligence in education: application, risk and governance. *Heilongjiang Res. High. Educ.* 38, 45–49.
- Wei, R. Z., and Sun, Y. P. (2022). The dilemma and governance path of the “drifters in government agencies” from the perspective of organizational equity theory. *Leadersh. Sci.* 812, 93–96.
- Williamson, B. (2017). Big data in education: the digital future of learning, policy and practice. *Big Data Educ.* 1–256. doi: 10.4135/9781529714920
- Wu, X. W., and Chen, X. Y. (2014). The education application of “big data” and China's education reform: opportunity recognition from data classification to evidence transformation. *Inner Mongol. Soc. Sci.* 35, 168–171.
- Wu, W., Mayo, G., McCuen, T. L., Issa, R. R., and Smith, D. K. (2018). Building information modeling body of knowledge. I: Background, framework, and initial development. *J. Constr. Eng. Manag.* 144:04018065.
- Xiao, L. P., Tian, Z. J., and Tian, J. W. (2022). Research on teaching quality wvaluation based on hidden Markov algorithm. *Heilongjiang Res. High. Educ.* 40, 151–155.
- Xiong, Y. (2011). Employee's perception of fairness in performance management in state-owned enterprises: the current situation and coping strategies. *Statis. Decis.* 345, 179–182.
- Xu, X., Yu, S. J., Pang, N., Dou, C. X., and Dan, L. (2022). Review on a big data-based innovative knowledge teaching evaluation system in universities. *J. Innov. Knowl.* 7:100197. doi: 10.1016/j.jik.2022.100197
- Yang, X. (2022). Trace on calculation method of evaluation reform of education. *J. East China Norm. Univ.* 40, 19–29.
- Yigitcanlar, T., Desouza, K. C., Butler, L., and Roozkhosh, F. (2020). Contributions and risks of artificial intelligence (AI) in building smarter cities: insights from a systematic review of the literature. *Energies* 13:1473. doi: 10.3390/en13061473
- Yuan, W., Liu, P., and Neubig, G. (2022). Can we automate scientific reviewing? *J. Artif. Intell. Res.* 75, 171–212. doi: 10.1613/jair.1.12862

Zednik, C., and Boelsen, H. (2022). Scientific exploration and explainable artificial intelligence. *Mind. Mach.* 32, 219–239. doi: 10.1007/s11023-021-09583-6

Zerilli, J., Knott, A., Maclaurin, J., and Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: is there a double standard? *Philos. Technol.* 32, 661–683. doi: 10.1007/s13347-018-0330-6

Zhang, Z. Z., and Qi, W. X. (2021). Information technology in educational assessment: empowerment, challenges and solutions. *Chinese journal of. Dist. Educ.* 41:1-11+76.

Zhang, Z. H., Wang, L., and Ji, K. (2022). Transformation of educational in new era of big data empowerment: technical logic, realistic dilemma and realization path. *e-Educ. Res.* 43, 33–39.

Zheng, X. M., and Liu, X. (2016). The effect of interactional justice on employee well-being: the mediating role of psychological empowerment and the moderating role of power distance. *Acta Psychol. Sin.* 48, 693–709. doi: 10.3724/SP.J.1041.2016.00693

Zhu, L. (2018). From “selection first” to “quality first”: reform of teaching evaluation over the last 40 years in China. *Glob. Educ.* 47, 37–47.