ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2025-10-13

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Rasooli, A., Holden, M. & Sinval, J. (2025). Preservice teachers' assessment decisions: Exploring the role of fairness conceptions. British Educational Research Journal. 51 (5), 2450-2473

Further information on publisher's website:
10.1002/berj.4181

Publisher's copyright statement:
This is the peer reviewed version of the following article: Rasooli, A., Holden, M. & Sinval, J. (2025). Preservice teachers' assessment decisions: Exploring the role of fairness conceptions. British Educational Research Journal. 51 (5), 2450-2473, which has been published in final form at https://dx.doi.org/10.1002/berj.4181. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

Preservice Teachers' Assessment Decisions: Exploring the Role of Fairness Conceptions

Amirhossein Rasooli[1,2], Michael Holden3, & Jorge Sinval[4,5, 6,7,8]

[1]Learning Sciences and Assessment Department, National Institute of Education, Nanyang Technological University, Singapore, Singapore

[2]Centre for Assessment and Evaluation, National Institute of Education, Nanyang Technological University, Singapore,Singapore

[3]University of Winnipeg, Winnipeg,Manitoba, Canada

[4]Psychology and Child & HumanDevelopment, National Institute ofEducation, Nanyang Technological University, Singapore, Singapore

[5]Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, Brazil

[6]Faculty of Philosophy, Sciences and Languages of Ribeirão Preto, University of São Paulo, São Paulo, Brazil

[7]Business Research Unit (BRU- IUL), Instituto Universitário de Lisboa (Iscte- IUL), Lisbon, Portugal

[8]Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

This is the accepted version of the following article: Rasooli, A., Holden, M., & Sinval, J. (2025). Preservice teachers' assessment decisions: Exploring the role of fairness conceptions. *British Educational Research Journal*. https://doi.org/10.1002/berj.4181, which has been published in final form at the British Educational Research Journal. This article may be used for non-commercial purposes in accordance with the Wiley Self-Archiving Policy (http://www.wileyauthors.com/self-archiving).

Preservice Teachers' Assessment Decisions: Exploring the Role of Fairness Conceptions

Abstract

Teachers' conceptions of fairness influence their approaches to assessment. Perceptions of fairness underpin how teachers legitimate their values and provide reasons for teachers to defend assessment decisions and actions. This study therefore examined fairness conceptions at a critical stage in teachers' journey toward assessment capacity: their perceptions as teacher candidates. Specifically, we investigated 228 preservice teachers' conceptions of assessment fairness using the Classroom Assessment Fairness Inventory (CAFI). The results of an exploratory factor analysis (EFA) showed four factors: *assessment communication acts, grading decisions, responses to cheating,* and *firm assessment decisions*. These factors highlighted the domains and underpinning principles of fairness that participating preservice teachers used to evaluate issues of fairness in assessment. The results provide initial empirical foundations to promote nuanced understandings of fairness in assessment education.

***Keywords*:** Fairness, Classroom Assessment, Preservice Teachers, Conceptions

Key Insights

1) What is the main issue that the paper addresses?

This study explores how preservice teachers conceptualise fairness in classroom assessment using the Classroom Assessment Fairness Inventory (CAFI).

2) What are the main insights that the paper provides?

The study identifies four key dimensions in assessment fairness: communication acts, grading decisions, responses to cheating, and firm assessment decisions. Preservice teachers' conceptions

align with assessment theory but also diverge in key areas, such as grading and group work. Structured discussions in teacher education are needed to interrogate fairness conceptions.

Preservice Teachers' Assessment Decisions: Exploring the Role of Fairness Conceptions

Teachers' assessment practices are key to everyday decisions about teaching and learning (Brookhart, 2018). As McMillan (2013) argued over a decade ago, "classroom assessment is the most powerful type of measurement in education that influences student learning." (p. 4) However, assessment practices can also be biased and produce unintended consequences for student learning and social and emotional outcomes (Author et al., 2018, 2019). The benefits and consequences of assessment practices depend in no small part on how teachers approach assessment and why they take the decisions they do. A body of crucial research has therefore amassed attempting to shed light on teachers' assessment approaches. This literature particularly examined teachers' conceptions, beliefs, and knowledge about assessment (Barnes et al., 2014; Bonner, 2016). Researchers have used multiple frameworks to explore teachers' conceptions (Brookhart, 2004; Brown, 2004, 2006; DeLuca et al., 2016a). In addition, some researchers have begun using theories such as motivation (Brookhart, 2013; Daniels et al., 2021) and emotions (Vogl & Pekrun, 2016) to explain teachers' assessment practices.

While the extant literature has been successful in investigating various dimensions of teachers' conceptions of assessment, assessment fairness – a critical component of teachers' assessment literacy and daily practice – remains underexplored (Author et al., 2018; Xu & Brown, 2016). While a theory of fairness can provide valuable insights into teachers' conceptions of assessment, empirical research on this foundational aspect of classroom assessment, especially concerning teacher populations, remains limited and scattered (Author, 2024). Previous research has shown that teachers explicitly articulate concerns about fairness in relation to various assessment practices, including feedback (Lizzio & Wilson, 2008), grading (Brookhart et al., 2016), accommodation (Thurlow, 2014), peer assessment (Panadero, 2016), and group work (Forsell et al., 2020). These findings show that fairness values are pervasive in

teachers' mindsets about various assessment practices despite these fairness issues being identified inductively in previous studies. Since "classroom assessment professionals do need to have foundational competencies in validity and fairness" (Brookhart, 2024, p. 20), more systematic empirical inquiry via an explicit theory of fairness can support guiding teacher education on fairness in classroom assessment. The purpose of this study is thus to examine teachers' conceptions of assessment fairness by studying how preservice teachers at the beginning of their assessment journeys conceptualise issues of (un)fairness. To establish a more robust empirical foundation for fairness research in assessment, we applied a fairness theory – drawing explicitly on principles from social psychology theory including procedural justice, distributive justice, and interactional justice – via an existing empirical tool, the Classroom Assessment Fairness Inventory (CAFI). Directly studying preservice teacher conceptions in this way contributes to ongoing calls for heightened assessment literacy, especially in preservice teacher education, that attends to the social and ethical complexities of classroom assessment (DeLuca et al., 2024; Pastore & Andrade, 2019). To that end, this study is guided by the following research question: What factors underpin preservice teachers' conceptions of fairness in classroom assessment?

Literature Review

Assessment and Assessment Fairness

Assessment research in feedback, accommodation, grading, group work, and peer assessment practices has shown that teachers value fairness as a foundation for decisions on assessment processes (Brookhart et al., 2016; Lizzio & Wilson, 2008; Panadero, 2016; Thurlow, 2014). The inductive nature of most of these studies attests to the key role of fairness in assessment, even when data are not explicitly analysed via a fairness theory. Assessment practices are naturally imbued with fairness concerns because they determine how resources such

as grading, feedback, and recognition are distributed, what procedures for such distributions should be, and how students will be treated in their interactions with teachers, principals, parents, and their peers (Author, 2024).

Given the noted role of fairness in assessment decisions, multiple assessment surveys and inventories have included some fairness items as part of broader conceptualisations and measurement of teachers' approaches to assessment. For example, Brown's (2004, 2006) survey included an explicit fairness item: 'assessment is unfair to students.' Other items from the survey could also be discussed in relation to fairness theory; however, we limit such generalisations because Brown (2006) did not explicitly argue that fairness underpins such items. DeLuca et al. (2016b) identified assessment fairness as one of four dimensions constituting approaches to classroom assessment inventory. This inventory conceptualised assessment fairness as including three broad approaches: standardisation (i.e., treating all students the same), differentiation (i.e., accommodations and differential treatment only for students with exceptional needs), and equity (i.e., differential treatment for all students). Using this inventory, two separate studies showed that preservice teachers generally endorsed equitable ($\bar{x}$ = 4.69; *SD* = 0.63) and differentiated ($\bar{x}$ = 4.61, *SD* = 0.67) approaches more than standardisation ($\bar{x}$ = 3.96; *SD* = 0.79). The findings also showed that preservice teachers endorsed equitable approaches more strongly as they progressed in their teacher education programmes (Coombs et al. 2018; 2020).

Focused Research on Assessment Fairness

In addition to these general assessment studies which include fairness measures, recent empirical research also focuses directly on teachers' conceptions of fairness and ethics in assessment (Author et al., 2023b; Fan et al., 2019; Green et al., 2007; Murillo & Hidalgo, 2020; Tierney, 2014). Green et al. (2007) conceptualised ethics as following two principles: not causing harm and avoiding 'score pollution' (i.e., assessment results should only reflect

achievement factors, not other variables like student behaviour). Additional studies have investigated ethical values and showed their relevance in teachers' judgements of various assessment practices. For example, Fan et al. (2019) explored preservice teachers' perceptions of ethical issues in classroom assessment across the US and China. 16 preservice teachers from the US and 15 from China participated in the focus group interviews. Both American and Chinese participants agreed with including effort into grades. American preservice teachers were divided about using peer grading as it was a breach of confidentiality and Chinese participants considered it to be unethical. While American participants agreed with using multiple methods of assessment, Chinese participants disagreed, citing practical limitations of large classroom sizes. American participants agreed with informing students about test material; however, Chinese participants did not. Tierney (2014) also drew upon Aristotelian practical wisdom when speaking with teachers about their practical ethical experiences. Teachers in Tierney's study highlighted establishing clear and consistent assessment expectations, providing multiple assessment opportunities and types, and developing constructive classroom environments assessment as key values for their conceptions of fairness. Altogether, these studies leveraged qualitative methodologies to contribute to building an empirical foundation for a deeper understanding of fairness in classroom assessment.

Recently, Sonnleitner and Kovacs (2020) developed a quantitative 15-item fairness barometer to evaluate teacher and student fairness views in the context of two dimensions of social psychology theory: procedural and informational justice. These dimensions warrant a brief description of social psychology theory and the notion of distributive, procedural, and interactional justice. Distributive justice is evaluated based on whether outcomes are distributed equally, equitably, or consider needs and consequences (Adams, 1965; Deutsch, 1975). For example, teachers' distribution of attention and one-on-one classroom supports are instances for

perceptions of distributive justice. If a teacher values an equal distribution of resources (e.g., attention), they are more likely to structure assessment moments with this sense of distributional justice in mind. Procedural justice is evaluated based on whether assessment procedures are consistent, reasonable, bias-free, transparent, representative of student voice, correctable, and ethical (Leventhal, 1980; Thibaut & Walker, 1975). Students may perceive procedural injustice if they believe a teacher's assessment procedures violate these principles. Interactional justice is evaluated based on whether a teacher's assessment interactions are respectful, including whether a teacher provides adequate, truthful, justified, and timely information (Bies & Moag 1986). For example, if a teacher asks students to complete a rough draft but only provides feedback the day before the revisions is due, students may perceive that practice as unjustifiable and unfair because of underlying concerns about the (dis)respectful nature of providing last-minute feedback. Drawing on the procedural and informational justice dimensions, Sonnleitner and Kovacs (2020) developed the fairness barometer to investigate perceptions of fairness in grading and exams. They found that teachers and students diverged substantially across their judgments of fairness. For example, teachers tended to judge themselves more fairly than did students. The results point to examining the components and processes that drive divergent conceptions of fairness.

Author et al. (2023a) developed the Classroom Assessment Fairness Inventory (CAFI) to explore conceptions of fairness in classroom assessment using social psychology theory. Building on a systematic review of existing fairness instruments in education (Author et al., 2022), the inventory augmented Sonnleitner and Kovacs' (2020) instrument in two respects. First, the CAFI included broader domains of application for fairness in assessment via five scenarios: group work, exams, grading, cheating, and feedback. Further, the CAFI leveraged a scenario-based format to elicit participants' conceptions of fairness with hypothetical teachers

(i.e., not teachers that participants have existing beliefs about). This approach further supports leveraging the CAFI as an instrument for teacher education in assessment, as preservice teachers and other stakeholders can discuss diverse fairness issues across the scenarios included in the inventory.

The CAFI was first used to empirically investigate secondary students' conceptions of fairness in classroom assessment (Author et al., 2023a). The results showed a five-factor model including (a) *unfairness in group work*, (b) *fairness in cheating*, (c) *fairness in grading*, (d) *unfairness in feedback*, and (e) *fairness in feedback*. In this study, we leveraged the CAFI to add to empirical evidence supporting appreciation of preservice teachers' conceptions of fairness in assessment.

Methods

Procedure and Sample

Upon ethical clearance, all participants were preservice teachers recruited from one Bachelor of Education (BEd) programme in Ontario, Canada. The programme includes a required classroom assessment course that occurs halfway through a two-year programme immediately after a seven-week practicum placement where participants observe a partner teacher's assessment practice, teach, and in most cases, design and implement assessments in the classroom. All preservice teachers enrolled in the course (n=446) participated in course activities related to assessment fairness, including a lecture on fairness issues, class discussion, independent completion of the CAFI, as well as in-depth group discussions of assessment fairness scenarios. The preservice teachers independently completed the CAFI partway through the course, after the lecture on fairness issues and their earlier programme experiences (e.g., the first teaching practicum), but before the in-depth group discussions of the CAFI scenarios. Participants' CAFI responses therefore share some contextual details (e.g., direct instruction on

assessment fairness, enrolment in the same BEd programme) but differ in other respects (e.g., previous experiences of fairness and personal context).

All students enrolled in the course were invited to complete the CAFI either for their own professional learning or for inclusion in the research. Of the full cohort, 228 (51.1%) consented to have their responses included in the data corpus. Demographic data (i.e. age, gender, and grade level) for these participants appear in Table 1. The vast majority of participants (80.7%) were female, a somewhat higher proportion than the overall teacher population in Ontario (74.8%, Ontario College of Teachers, OCT, 2020). All participants were between 21 and 45 years old. Unsurprisingly, participating preservice teachers were predominantly 21-29 years old (90.4%), with a mean participant age of 24.0 and a mode of 22 (43.0%). A slight majority of participants (53.5%) were preparing to teach grades 7-12. It is difficult to compare this proportion to the distribution of Ontario teachers across grade levels, as the programme only admits students to teaching streams for (a) kindergarten to grade 6 and (b) grade 7 to 12, whereas the Government of Ontario (2025) reports data for teachers working in (a) kindergarten to grade 8 (66.5%) and (b) grade 9 to 12 (33.5%).

Table 1

Demographic statistics for CAFI participants

	n (228)	**%**
Gender		
Male	40	17.5
Female	184	80.7
Non-Binary	3	1.3
No Response	1	0.4
Age		
21-23	159	69.7
24-26	39	17.1
27-29	8	3.5
30-39	12	5.3
40-45	7	3.1
No Response	3	1.3
Grade Levels[1]		

Primary/Junior (K-6)	103	45.2
Intermediate/Senior (7-12)	122	53.5
No Response	3	1.3

[1] Ontario teachers are initially certified in two of four divisions (OCT, 2025): Primary (kindergarten to grade 3), Junior (grade 4-6), Intermediate (grade 7-10), and Senior (grade 11-12).

Measure

The CAFI was developed by Author (2021) to investigate fairness in classroom assessment, with particular attention to how participants respond to specific items related to perceived (un)fairness in assessment scenarios. The CAFI leverages a hypothetical scenario-based format that measures fairness in classroom assessment across five scenarios. After reading each scenario, respondents rate a series of Likert-type items (n = 40) focusing on (a) group work, (b) exams, (c) cheating, (d) grading, and (e) feedback. Items were rated on a Likert 7-point scale: (1) Highly unfair, (2) Unfair, (3) Somewhat unfair, (4) I am not sure, (5) Somewhat fair, (6) Fair, (7) Highly fair. The ratings close to 1.00 indicate that an item is perceived to be highly unfair and ratings close to 7.00 indicate an item to be perceived as highly fair.

Consistent with the role of distributive, procedural, and interactional (in)justice in assessment fairness, the 40 scenario items each represent an underpinning principle of social psychology theory. For example, item Q3_2 (*Ms. Johnston did not give the student an opportunity to explain the reasons for cheating before making her decision to give zero*) is underpinned by voice as an issue of procedural justice (Kazemi & Tornblom, 2008; Leventhal, 1980; Thibaut & Walker, 1975), since the decision to not allow students to voice their perspectives about assessment fairness may be perceived as a procedural injustice. Similarly, item Q5_6 (*Mr. Dembe did not treat students respectfully in his feedback procedure*) is underpinned by respect as an issue of interactional justice (Author, 2019; Bies & Moag, 1986; Greenberg, 1993), since respondents may differ in their conception of what demonstrations of

(dis)respect are acceptable in feedback practices in classroom assessment. The social psychology theory principles underpinning each item are detailed in Appendix 1.

The CAFI has been previously validated using expert panel reviews as well as confirmatory factory analysis (Author, 2021, Author et al., 2023a). The review for content validity evidence by 10 international experts showed moderate agreement with the initial conceptualised model (Appendix 1) across the scenarios (scenario 1= 0.62, $p < 0.01$; scenario 2 = 0.45, $p < 0.06$; scenario 3 = 0.61, $p < 0.01$; scenario 4 = 0.45, $p < 0.07$, and scenario 5 = 0.51, $p < 0.03$). The results of the confirmatory factor with former secondary students showed a five-factor model including (a) *unfairness in group work (α = 0.79)*, (b) *fairness in cheating (α = 0.82)*, (c) *fairness in grading (α = 0.85)*, (d) *unfairness in feedback (α = 0.84)*, and (e) *fairness in feedback (α = 0.87)*. The overall goodness-of-fit indices for the five-factor model showed RMSEA = 0.06 (90% CI [0.05– 0.07]), CFI = 0.91 and TLI = 0.89. This is the first study to use the CAFI to explore preservice teacher conceptions of assessment fairness.

Data Analysis

All statistical analyses were conducted using *R* (R Core Team, 2024) and *RStudio* (Posit Team, 2024). Descriptive statistics were obtained via the *skimr* package (McNamara et al., 2021) and the standard error of the mean (*SEM*) was estimated with *plotrix* (Lemon, 2006).

The distributional properties of the items reveal a range of means (*M*) from 1.41 to 6.69, indicating varying levels of response tendencies. Standard deviations (*SD*) range from 0.65 to 1.89, showing differences in response variability. Minimum (*Min*) and maximum (*Max*) values span the full response scale (1 to 7), suggesting the presence of the complete range of possible responses for the majority of the CAFI's items. Histograms show diverse patterns of responses across items. Standard errors of the mean (*SEM*) are generally low, indicating precise mean estimates. See Table 2 for additional information about descriptive statistics.

Table 2

Items' Distributional Properties

	Item	*M*	*SD*	*Min*	*Max*	Histogram	*SEM*
Q1_1	Mr. Chu selected group members based on mixed ability.	4.98	1.47	1	7		0.10
Q1_2	Mr. Chu did not provide students a choice in selecting their group members.	4.50	1.43	1	7		0.09
Q1_3	Mr. Chu allowed students' complaints over group dynamics.	5.08	1.72	1	7		0.11
Q1_4	Mr. Chu was not detailed in communicating how he will assess students' groupwork.	1.69	1.02	1	7		0.07
Q1_5	Mr. Chu gave the same grades to all group members.	2.64	1.41	1	7		0.09
Q1_6	Mr. Chu did not give individual grades for each group member based on their contributions and learning.	2.34	1.27	1	7		0.08
Q1_7	Mr. Chu did not justify his grades to students who appealed.	1.50	0.97	1	7		0.06
Q2_1	Mr. Ahmed held firm on the exam date.	3.84	1.59	1	7		0.11
Q2_2	Mr. Ahmed did not explicitly state what would be on the exam.	2.68	1.47	1	7		0.10
Q2_3	Mr. Ahmed included in the exam both easy to difficult questions.	5.93	0.87	2	7		0.06
Q2_4	Students with disabilities and English language learners received accommodations for the exam (e.g., more writing time).	6.54	0.75	1	7		0.05
Q2_5	Mr. Ahmed graded his students more leniently than other teachers.	4.30	1.33	1	7		0.09
Q2_6	Mr. Ahmed did not remove the two questions on the content that were not taught before.	1.92	1.10	1	6		0.07
Q2_7	Mr. Ahmed did not respond to students' complaints with a respectful tone.	1.41	0.80	1	6		0.05
Q2_8	Mr. Ahmed gave another exam opportunity to students who missed the exam.	4.64	1.89	1	7		0.12
Q3_1	Ms. Johnston was not transparent about her cheating policy at the beginning of the course.	1.98	1.07	1	7		0.07
Q3_2	Ms. Johnston did not give the student an opportunity to explain the reasons for cheating before making her decision to give zero.	2.32	1.30	1	7		0.09
Q3_3	Ms. Johnston gave a zero because cheating is unfair to other students' efforts.	4.27	1.58	1	7		0.10
Q3_4	Ms. Johnston gave a zero because cheating is ethically wrong.	4.47	1.52	1	7		0.10
Q3_5	Ms. Johnston gave a zero to signal that anyone who cheats should be punished for this action.	3.34	1.56	1	7		0.10
Q3_6	Ms. Johnston asked the student to leave the classroom in front of other students.	2.18	1.37	1	7		0.09
Q3_7	Ms. Johnston explained her cheating decision to the student.	5.39	1.26	1	7		0.08

Q3_8	Ms. Johnston did not forgive the student's cheating this time.	3.28	1.46	1	7		0.10
Q4_1	Ms. Mendes treated students respectfully during classroom assessment.	6.46	1.05	1	7		0.07
Q4_2	Ms. Mendes largely considered student achievement of learning objectives in her grading.	5.92	1.34	1	7		0.09
Q4_3	Ms. Mendes detailed her grading criteria, with test scores making up 70% of a student's grade.	4.98	1.67	1	7		0.11
Q4_4	Ms. Mendes communicated test results in one week after handing the test in.	6.06	1.06	2	7		0.07
Q4_5	Ms. Mendes would sometimes allow students to discuss their grades if there was enough time in class.	5.30	1.40	1	7		0.09
Q4_6	Ms. Mendes gave adequate justification for students who spoke up for their grades.	5.89	1.21	1	7		0.08
Q4_7	Ms. Mendes adjusted the grades of failing students with at-risk backgrounds.	4.47	1.64	1	7		0.11
Q4_8	Ms. Mendes considered students' future university admissions to adjust grades.	3.71	1.72	1	7		0.11
Q4_9	Ms. Mendes considered student misbehaviour (e.g., disruptions) in her grading.	2.37	1.46	1	7		0.10
Q5_1	Mr. Dembe provided a clear rubric for assessing students' essays.	6.69	0.65	2	7		0.04
Q5_2	Mr. Dembe provided feedback after four days of essay submissions.	6.56	0.83	1	7		0.05
Q5_3	Mr. Dembe gave students a chance to further discuss his feedback.	6.64	0.70	1	7		0.05
Q5_4	Mr. Dembe provided feedback based on the amount of effort each student has put in the work.	5.29	1.73	1	7		0.11
Q5_5	Mr. Dembe did not provide more feedback to students who had weaker performance.	5.27	0.92	1	6		0.06
Q5_6	Mr. Dembe did not treat students respectfully in his feedback procedure.	6.61	0.85	1	7		0.06
Q5_7	Mr. Dembe gave more feedback to his favourite students.	5.70	0.77	1	6		0.05
Q5_8	Mr. Dembe explained that his feedback procedure was based on the amount of effort each student has put in.	5.98	1.33	1	7		0.09

Confirmatory Factor Analysis (CFA) was used to evaluate the fit of the conceptual model to the data based on the five-factor model according to the five scenarios (see Appendix 1). The *lavaan* package (Rosseel, 2012) was used to estimate the CFA using the WLSMV estimator. The selected goodness-of-fit indices were: *NFI*, *TLI*, *CFI*, *RMSEA*, χ^2, and *SRMR*. Values of *CFI*, *NFI*, and *TLI* $\geq .95$, *SRMR* and *RMSEA* $\leq .05$, were considered as indicative of good fit (Hu & Bentler, 1999). The CFA did not present an acceptable fit of the conceptual

model to the data ($\chi^2_{(730)}$ = 1711.407; p < .001; n = 212; *CFI* = .547; *NFI* = .389; *TLI* = .516; *SRMR* = .108; *RMSEA* = .080; $p_{(RMSEA \leq .05)}$ < .001; 90% CI (.075; .084)). After removing several items (11 items) with very low factor loadings, the data did not meet the fit threshold ($\chi^2_{(367)}$ =854.980; p < .001; n = 212; *CFI* = .669; *NFI* = .545; *TLI* = .634; *SRMR* = .106; *RMSEA* = .079; *P*(*RMSEA* ≤ .05) < .001; 90% CI (.073; .086)).

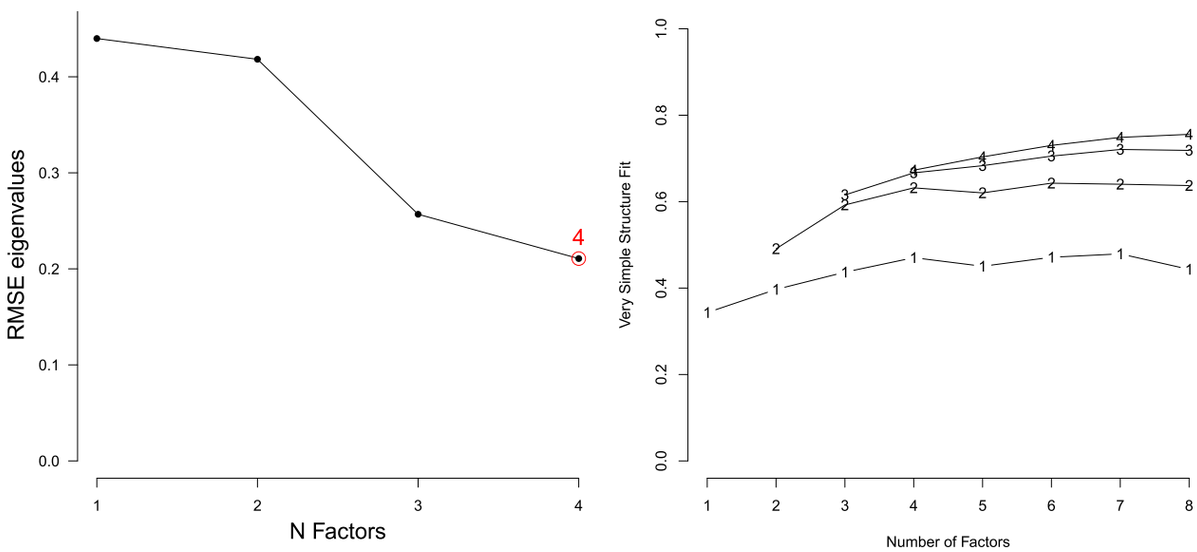
As the original conceptualised model did not present a good fit to the data, an Exploratory Factor Analysis (EFA) was conducted. EFA was performed to explore alternative dimensionality for the CAFI. The Kaiser–Meyer–Olkin (KMO) coefficient was used as a measure of sampling adequacy and Bartlett's test of sphericity was used to test the null hypothesis that the variables are uncorrelated in the population. KMO ≥ .8 and a Bartlett test with a statistically significant *p*-value indicate adequate sampling. The selection of the number of factors was based on the comparison data (Ruscio & Roche, 2012) and very simple structure methods (VSS; Revelle & Rocklin, 1979). EFA results were based on the polychoric correlations (ρ_{PC}), which are more adequate for ordinal manifest variables and when indicators have asymmetrical distributions (Fabrigar et al., 1999). The factors from EFA were extracted using weighted least squares and oblimin rotation, with a loading cutoff of .30. The *psych* package (Revelle, 2020) was used for estimating the KMO, Bartlett's test, VSS, ρ_{PC}, and EFA factor extraction. The comparison data analysis was conducted using the *EFAtools* package (Steiner & Grieder, 2020). The reliability was estimated using the ω internal consistency estimator through the *psych* package (McNeish, 2018).

Results

Sample adequacy was obtained (*KMO* = .754; $\chi^2_{(780)}$ = 3,282.65; p < .001) and the comparison data and the VSS approach suggested retaining four factors (see Figure 1).

Figure 1

Comparison Data and Very Simple Structure Plots



An EFA solution using the WLS factoring method identified four factors (Table 3), which explained 39.9% of the variance with the following distribution per factor: *assessment communication acts* = 12.5%, *grading decisions* = 10.4%, *responses to cheating* = 9.0%, and *firm assessment decisions* = 8.0%. The reliability in terms of internal consistency (i.e., ω) ranged from acceptable (.75) to very good (.90). The factor analysis revealed a distinct pattern of item loadings across four factors. Items Q5_3, Q5_2, Q4_6, Q5_1, Q4_1, Q4_4, Q4_2, Q2_3, Q2_4, Q4_5, and Q4_3 demonstrated strong loadings on the first factor (*assessment communication acts*), with loadings ranging from .383 to .756. The items in this factor centred on how teachers used individual assessment moments – such as introducing a rubric (Q5_1) or sharing assessment results with students and families (Q4_2) – as communicative acts during the assessment process. Related principles included voice, timeliness, and respect in provision of academic feedback as well as transparency, justification, and reasonableness in the grading procedure. Preservice teachers perceived all the items in this factor as contributing to fairer assessments.

The second factor (*grading decisions*) was characterized by relevant loadings from items Q4_9, Q4_8, and Q4_7, with loadings of .606, .578, and .481, respectively. All three items that loaded to this factor related to decisions teachers made when entering or adjusting student grades. While preservice teachers perceived grade adjustments based on future university admission and student behaviour to be unfair, they tended to be more favourable to consider grade adjustments for at-risk students as fair.

The third factor (*Responses to cheating*) included items Q3_4, Q3_3, Q3_5, Q3_2, Q3_8, Q3_1, Q3_7, and Q3_6, with loadings from .420 to .751. Notably, all the items in this factor loading were drawn from the same CAFI scenario, which involved a teacher's responses to student cheating. These items incorporated issues of ethicality, caring, justification, voice, transparency, and respectfulness in addressing cheating in assessments. Preservice teachers tended to judge giving students a zero for cheating as fair for equity and ethical reasons (i.e., due to violation of others' efforts and cheating as an intrinsically wrong action). However, they perceived giving a zero as somewhat unfair if the purpose of the zero was to signal the consequences of cheating as a lesson for others. Participants also considered explaining and justifying cheating decision to students as fair. On the contrary, they perceived untransparent cheating policy, lack of student voice for explaining cheating behaviours, disrespectful treatment of students who cheated, and not forgiving student cheating to be unfair.

The fourth factor (*Firm assessment decisions*) was represented by items Q1_6, Q1_7, Q2_6, Q1_5, Q2_1, Q1_2, Q2_2, Q1_4, and Q2_7, with loadings ranging from .616 to .392, respectively. The items that loaded to this factor described a teacher holding firm on an existing assessment decision, such as not assigning individual grades after students protested the use of group grading (Q1_6), or not adjusting the date of an exam in response to student requests (Q2_1). Preservice teachers perceived failure to assign individual grades for groupwork and lack

of transparency and justification in assessing groupwork to be unfair. Participants also considered untransparent exam content, lack of respect in justifying exam decisions, and inflexibility in listening to student voice in changing the exam date as unfair. However, they perceived the formation of groups without student choice as fair.

Table 3

Extracted Exploratory Factor Analysis Solution

Label	Item	Assessment Communication Acts	Grading Decisions	Responses to Cheating	Firm Assessment Decisions	*M*
Q5_3	Mr. Dembe gave students a chance to further discuss his feedback.	**.756**				6.64
Q5_2	Mr. Dembe provided feedback after four days of essay submissions.	**.714**				5.29
Q4_6	Ms. Mendes gave adequate justification for students who spoke up for their grades.	**.693**				5.89
Q5_1	Mr. Dembe provided a clear rubric for assessing students' essays.	**.682**				6.69
Q4_1	Ms. Mendes treated students respectfully during classroom assessment.	**.648**				6.46
Q4_4	Ms. Mendes communicated test results in one week after handing the test in.	**.635**				6.06
Q4_2	Ms. Mendes largely considered student achievement of learning objectives in her grading.	**.614**				5.92
Q2_3	Mr. Ahmed included in the exam both easy to difficult questions.	**.415**			.413	5.93
Q2_4	Students with disabilities and English language learners received accommodations for the exam (e.g., more writing time).	**.412**				6.54
Q4_5	Ms. Mendes would sometimes allow students to discuss their grades if there was enough time in class.	**.399**	.354			5.30
Q4_3	Ms. Mendes detailed her grading criteria, with test scores making up 70% of a student's grade.	**.383**				4.98
Q4_9	Ms. Mendes considered student misbehaviour (e.g., disruptions) in her grading.		**.606**			2.37
Q4_8	Ms. Mendes considered students' future university admissions to adjust grades.		**.578**			3.71

Q4_7	Ms. Mendes adjusted the grades of failing students with at-risk backgrounds.	.386	**.481**			4.47
Q3_4	Ms. Johnston gave a zero because cheating is ethically wrong.			**.751**		4.47
Q3_3	Ms. Johnston gave a zero because cheating is unfair to other students' efforts.			**.75**		4.27
Q3_5	Ms. Johnston gave a zero to signal that anyone who cheats should be punished for this action.			**.652**		3.34
Q3_2	Ms. Johnston did not give the student an opportunity to explain the reasons for cheating before making her decision to give zero.			**.606**		2.32
Q3_8	Ms. Johnston did not forgive the student's cheating this time.			**.603**		3.28
Q3_1	Ms. Johnston was not transparent about her cheating policy at the beginning of the course.			**.498**		1.98
Q3_7	Ms. Johnston explained her cheating decision to the student.	.38		**.446**		5.39
Q3_6	Ms. Johnston asked the student to leave the classroom in front of other students.			**.42**		2.18
Q1_6	Mr. Chu did not give individual grades for each group member based on their contributions and learning.				**.616**	2.34
Q1_7	Mr. Chu did not justify his grades to students who appealed.				**.587**	1.50
Q2_6	Mr. Ahmed did not remove the two questions on the content that were not taught before.				**.573**	1.92
Q1_5	Mr. Chu gave the same grades to all group members.				**.543**	2.64
Q2_1	Mr. Ahmed held firm on the exam date.				**.468**	3.84
Q1_2	Mr. Chu did not provide students a choice in selecting their group members.				**.432**	4.50
Q2_2	Mr. Ahmed did not explicitly state what would be on the exam.				**.419**	2.68
Q1_4	Mr. Chu was not detailed in communicating how he will assess students' groupwork.		.397		**.345**	1.69
Q2_7	Mr. Ahmed did not respond to students' complaints with a respectful tone.		.381		**.392**	1.41
Q1_3	Mr. Chu allowed students' complaints over group dynamics.					
Q2_8	Mr. Ahmed gave another exam opportunity to students who missed the exam.					
Q2_5	Mr. Ahmed graded his students more leniently than other teachers.					

Q5_4	Mr. Dembe provided feedback based on the amount of effort each student has put in the work.				
Q5_7	Mr. Dembe gave more feedback to his favourite students				
Q5_5	Mr. Dembe did not provide more feedback to students who had weaker performance.				
Q5_8	Mr. Dembe explained that his feedback procedure was based on the amount of effort each student has put in.				
Q5_6	Mr. Dembe did not treat students respectfully in his feedback procedure.				
Q1_1	Mr. Chu selected group members based on mixed ability.				
Omega reliability ω		**.90**	**.75**	**.89**	**.90**

Note. Factor loadings with absolute values < .32 are omitted from the table. Factor loadings in bold indicate the factor to which each item is attributed in the proposed solution.

Having described the item loadings for the four factors, we now turn to the means ($\bar{x}$), standard deviations (s), and correlations between the latent scores, presented in Table 4. *Assessment communication acts* had the highest mean ($\bar{x}$ = 6.09, s = 0.55) and *Firm assessment decisions* had the lowest ($\bar{x}$ = 2.50, s = 0.69). *Assessment communication acts* had negative associations with *Grading decisions* (r = -.18; p < .05) and *Firm assessment decisions* (r = -.14; p < .05). *Grading decisions* showed moderate associations with *Firm assessment decisions* (r = .44; p < .001) and *Responses to cheating* (r = .25; p < .001). Finally, *Firm assessment decisions* displayed statistically significant correlations with *Responses to cheating* (r = .34; p < .001).

Table 4

Mean and Standard Deviation of the Raw Scores, and Correlations of the Suggested Factors

Factors	$\bar{x}$	s	**1**	**2**	**3**	**4**
Assessment communication acts (1)	6.09	0.55				
Grading decisions (2)	3.51	1.27	−.18*			
Responses to cheating (3)	3.41	0.89	−.04	.25***		
Firm assessment decisions (4)	2.50	0.69	−.14*	.44***	.34***	

Note. * ≤ .05; *** < .001.

Discussion

In the last two decades, continued research has investigated teachers' conceptions of assessment as a basis to guide teacher education in assessment (Brookhart, 2004; Brown, 2004,

2006; DeLuca et al., 2016a). To add to this literature, this study explicitly leveraged a fairness theory to provide insights into preservice teachers' conceptions of assessment. Accordingly, this study offers initial empirical findings into the role of fairness theory in guiding preservice teachers' conceptions of assessment.

The results support using social psychology theory as operationalized in the CAFI to account for preservice teachers' conceptions of assessment. As "dimensionality depends on the population of test-takers" (Mellenbergh, 2011, p. 214), the findings indicated a four-factor model whereby preservice teachers in this study conceived of fairness in assessment. The factors include (a) *assessment communication acts*, (b) *grading decisions*, (c) *responses to cheating*, and (d) *firm assessment decisions*. These factors help to explain how preservice teachers viewed fairness issues in the context of assessment scenarios about feedback, grading, group work, cheating, and exams. These domains have been argued and shown empirically to constitute conceptions of fairness in assessment (Author et al., 2018; Author, 2021).

Preservice teachers' views in this study aligned with predictions from social psychology theory across the four factors. Social psychology theory contends that individuals would perceive fairness if resource distributions, procedures for these distributions, and interactions and information involved in these distributions and procedures follow principles of justice (e.g., equity, transparency, respect, justification; Author et al., 2019; Kazemi & Törnblom, 2008). Concurrently, social psychology theory suggests that individuals would perceive unfairness if these principles were violated (Adams, 1965; Bies & Moag, 1986; Deutsch, 1975; Leventhal, 1980; Thibaut & Walker, 1975). Across the four factors, preservice teachers tended to report a practice as fair if the assessment treatment followed principles of ethicality, equity, justification, respect, voice, transparency, reasonableness, and timeliness during grading, feedback, cheating, and exam interactions. Likewise, participants tended to rate a practice as unfair if they perceived

the assessment outcomes, procedures, interactions, and information to be inequitable, untransparent, disrespectful, unjustified, or devoid of opportunities for student voice.

Despite these concurrences, there were some cases where preservice teachers diverged from predictions of social psychology theory in their conceptions of fairness. For example, while social psychology theory suggests that providing for student voice contributes to perceptions of fairness based on social psychology theory, and while students perceive voice as key to fairness in forming groups (Forsell et al., 2020), preservice teachers viewed group formation without providing for student voice as fairer. Preservice teachers might be concerned with disadvantaging low-readiness students if groups are formed by students' choices (Authors, 2021). There were also a few cases where preservice teachers' conceptions of fairness diverged from established beliefs in the assessment literature. For example, although preservice teachers highly endorsed the fairness of achievement-based grading, they tended to rate grade adjustments for at-risk students as fair, a practice that is frowned upon in standards-based grading (Brookhart et al., 2016). This variation highlights that grading is often seen as a 'third rail' in teachers' practice, creating a great deal of tension and disagreement among teachers in terms of which approaches are (in)defensible (Feldman, 2019).

These findings have implications for teacher education on assessment fairness and assessment literacy, and for strengthening research in these areas. Chiefly, the data suggest that participants conceptualised fairness in diverse ways, sometimes conflicting with each other and existing research. To support teacher education on assessment fairness, more structured opportunities are needed to develop competencies in interrogating what preservice teachers mean by (un)fairness in assessment (Brookhart, 2024; Pastore & Andrade, 2019). This may involve specific discussions of assessment communication acts, grading decisions, responses to cheating, and firm assessment decisions, as these CAFI participants experienced, or it may relate more

broadly to social psychology theory and applying the principles of distributive justice, procedural justice, and interactional justice to teachers' specific classroom contexts. For example, teacher educators can build on the firm assessment decisions factor to interrogate preservice teachers' fairness values in inclusion of group work skills (e.g., respectful relationships, adequate communication) in student grades. While academic enablers such as effort and participation are typically excluded from achievement-based grading (Brookhart et al., 2016), it is worth investigating whether preservice teachers align with this view. Such research will support re-examining achievement-based grading values to respond to 21st century competencies emphasising teamwork learning (Maxwell, 2023). If both teamwork skills and curriculum outcomes are valued in grading, can teachers support recognising epistemic values like collaboration skills (with less weightage) in grading holistic learning outcomes (Wijsen et al., 2022)?

Future research needs to continue refining the CAFI as a tool for examining a phenomenon as complex as conceptions of fairness in assessment. The findings shed light on several possible paths forward. First, the CAFI contains items with fairness and unfairness wording within each scenario. Psychometrics research has consistently showed that the item wording effect (i.e., use of positively and negatively worded items) can create differential response patterns as well as biased parameters and reliability estimates (Wang et al., 2015). Consequently, the items with different wording effects can merge to shape separate factors. In this study and in prior research with former secondary students (Author et al., 2023a), factors in the CAFI could be interpreted to be a product of either adhering to principles of justice (i.e., positively worded) or violating them (i.e., negatively worded). For example, an item (i.e., *Q4_6 Ms. Mendes gave adequate justification for students who spoke up for their grades*) loaded onto the first factor and another item (i.e., *Q1_7 Mr. Chu did not justify his grades to students who*

appealed) loaded onto the fourth factor. Looking closely across these two factors shows that wording has effect on the shared loadings within each factor. Although psychometric research considers item wording effect to be a matter of bias, the use of factor analysis, as in the case of this study, includes item wording as part of the construct of fairness.

Research in organisational psychology, where assessment fairness research draws on (Author et al., 2022; Author, 2024), has showed that perceptions of justice provoke different responses than perceptions of injustice (Colquitt et al., 2015). Viewed from a nomological validity lens (Cronbach & Meehl, 1955), this differential predictive validity evidence for (in)justice perceptions provide evidence for different underlying dimensions (i.e., justice vs. injustice). Colquitt et al. (2015) also showed that a survey that includes both fairness *and* unfairness items explain more variance in outcome variables than surveys with exclusively fairness *or* unfairness items. These results beg the question whether unfairness and fairness are different constructs based on predictive validity evidence as well as the factor analytic reflective models (Borsboom, 2005) or are one construct with biased measures of item wording effects. The argument-based validity frameworks (e.g., Kane, 2006) do not provide straightforward answers to these questions, and we do not have direct access to the mental structure of constructs. Thus, future research into the validity of the CAFI could be strengthened by the relevant neurological and philosophical discussions to establish whether conceptions of fairness vs. unfairness are distinct dimensions or byproduct of an item wording effect. The conclusion of those investigations could provide a guidance to revise the inventory.

Overall, the EFA results largely conform to the principles of simple structure (Thurstone, 1947), characterized by high variable loadings on a single factor (typically > .40 or .50), minimal cross-loadings, and clear factor assignments. The four identified factors correspond to the provided assessment scenarios, with most items exhibiting high loadings on

their respective factors, indicating a coherent and interpretable structure. This suggests that the factors represent meaningful constructs related to fairness in assessment practices.

The cumulative variance explained (40%) in this study falls within acceptable ranges for social sciences and humanities research, where constructs such as fairness in assessment are inherently complex and multidimensional (Tinsley & Tinsley, 1987). Hair et al. (2019) emphasized that factor solutions explaining 60% or less of variance can still be satisfactory, particularly when factor loadings align well with theoretical expectations. Additionally, decisions about the number of factors should be guided by theoretical considerations and scree plots, as extracting additional factors may introduce noise and reduce interpretability (Nunnally & Bernstein, 1994).

Deviations from simple structure were observed in some items, including cross-loadings that suggest overlap between constructs, particularly for items reflecting behaviours relevant to multiple aspects of fairness. Additionally, certain items displayed insufficient loadings on any factor, indicating potential misalignment with the measured constructs or limited contribution to the factor solution (Nunnally & Bernstein, 1994). We had nine items that did not load onto any of the factors. One identified reason was that items such as Q1_3, Q2_8, and Q2_5 showed high uniqueness and low communality, implying that they are less well-explained by any of the four factors and may require further investigation or consideration of additional factors. This result signals that further attention may be needed to adjust the scenario items given that the CAFI is a new tool that needs continued investigation of its structural and substantive aspects. Overall, the CAFI provides initial empirical foundations for interpreting preservice teachers' conceptions of fairness in classroom assessment. As fairness is a complex construct and there exists limited research in this area, continued validity investigations for the CAFI will provide a sound evidentiary foundation for theory, policy, and practice in educational assessment.

Practical Implications for Using CAFI in Teacher Education in Assessment

The primary focus of fairness research in assessment has largely been on ensuring fairness in educational testing (Author, 2024). However, with the growing emphasis on classroom assessment and the recognition of teachers as key figures in students' daily assessments (Brookhart, 2018), the theoretical understanding of fairness—particularly regarding teachers' conceptions of fairness—remains notably underexplored in teacher education. Fairness is a critical component of teachers' judgments regarding assessment processes and outcomes; however, there remains a lack of conceptual and methodological resources to facilitate meaningful discussions on fairness in assessments among teachers and educators.

Relying on social psychology theory, this study introduces a novel instrument designed to engage teachers in considering various scenarios of fairness in assessments. Teacher educators can use these scenarios in initial teacher education to encourage preservice teachers to reflect on their values regarding fairness in assessment. For example, authors have developed a US-based course module for the National Council on Measurement in Education (Author, 2023) based on the scenarios freely accessible to support teacher education on fairness in assessment. Teacher educators can use the CAFI scenarios in courses on classroom assessment (in a programme where that course is mandatory) and in pedagogy courses (in a programme where there is no assessment course) to challenge unspoken assumptions about what (un)fairness is and whether all teachers conceptualise fairness in the same way.

These reflections offer a valuable opportunity for both educators and preservice teachers to critically examine, and potentially reconsider, their approaches to assessment through a fairness lens. For example, while the assessment acts described in the first factor occur throughout the assessment process (e.g., during assessment design, initial instruction, assessment of learning, and feedback discussions), they are unified as assessment communication acts –

moments when teachers and students use assessment to talk about assessment and learning in explicit and implicit ways. The first factor, then, may serve as a vehicle for discussing the (un)fairness of what assessment communicates, which we have elsewhere called "assessment talk" (Author et al., 2024). Likewise, while the CAFI was developed before the onset of generative artificial intelligence, the third factor and scenario three both provide avenues to discuss issues of cheating, fairness, and their relationship to comprehensive academic integrity (Eaton, 2024).

Each scenario provides a rich context for preservice teachers to reflect on their views of fairness. For instance, the group work scenario raises key issues such as decisions regarding the group formation, the allocation of individual versus group grades, and the relational fairness of group work. Teacher educators can use this scenario to expand discussions on the fairness of treatment for students from minority backgrounds and those with special needs in group work assessments. The perspectives expressed by preservice teachers can be contrasted with those of students, fostering deeper reflection. For example, if some preservice teachers support mixed-ability groupings, as indicated by the descriptive statistics in this study, teacher educators can invite them to consider their choices from a student's perspective, recalling their own experiences in K-12 education. This discussion can then be further informed by empirical data identified in assessment fairness literature. Ultimately, such conversations across diverse scenarios can contribute to more focused and effective teacher education on fairness in assessment.

More broadly, we contend that the CAFI and other explicit discussions of assessment fairness stand to benefit practicing teachers and teacher educators. Just as preservice teachers may benefit from discussing their conceptions of assessment fairness, teachers and faculties of education may benefit from explicitly examining how (un)fairness manifests in their classrooms

and whether schools and faculties are operating under consistent models of assessment fairness (Feldman, 2019). We are encouraged by the growing recognition of the importance of fairness to classroom assessment (e.g., Brookhart, 2024; DeLuca et al., 2024; Pastore & Andrade, 2019), and suggest that more explicit engagement with a theory of assessment fairness may help to address longstanding inconsistencies in the assessment cultures surrounding students and teachers (Ylonen et al., 2018). Closely examining fairness conceptions may be especially important for faculties of education, since they have the dual responsibility of both fairly assessing preservice teachers and developing those preservice teachers' ability to support assessment fairness in their own classrooms moving forward.

Limitations

The limitations of this research study include the small sample size and the focus on only one teacher education programme in Canada. The sample size represented a modest subject-to-item ratio (5.7:1, see Costello & Osborne, 2005). Whereas the strength of the factor loadings warranted pursuing EFA analysis, future research requires greater sample size to examine the validity of the CAFI. Because teachers' assessment practices are socially and culturally situated (Shepard, 2019), conceptions of assessment fairness may vary across social and cultural contexts. Likewise, teachers operating in different policy contexts may respond to specific CAFI items differently. For example, a teacher operating in a jurisdiction that forbids teachers from assigning zeroes may respond differently to scenario three's questions about the fairness of zeroes. All the participants in this sample — preservice teachers in Ontario, Canada — operate in a context where providing accommodations to language learners and students with identified needs is an explicit mandate. To that end, we were unsurprised to see the item on accommodations (Q2_4) had one of the lowest standard deviations and one of the highest fairness endorsements among these participants (see Table 2). Similar issues toward specific

assessment acts may appear differently in different social, cultural, and policy contexts. Further research with the CAFI is needed to explore the conceptions of preservice teachers from various educational programmes and jurisdictions, including what conditions may contribute to changes in conceptions of assessment fairness.

Conclusion

This study leveraged the CAFI to investigate preservice teachers' conceptions of fairness in assessment. Despite the crucial role of fairness to teachers' conceptions and practices of assessment (Tierney, 2013, 2014), limited attention has been focused on examining preservice teachers' conceptions of assessment fairness (Author, 2024). The results in this study showed that preservice teachers perceived assessment fairness in relation to four factors. These factors highlighted the domains and underpinning principles of fairness that participating preservice teachers used to evaluate issues of fairness in assessment. The results provide initial empirical foundation to promote assessment education in fairness in relation to critical population in educational assessment: preservice teachers.

References

Adams J. S. (1965). Inequity in social exchange. In Berkowitz L. (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). Academic Press.

Barnes N., Fives H., & Dacey C. M. (2014). Teachers' beliefs about assessment. In Fives H. & Gill M. G. (Eds.), *International handbook of research on teacher beliefs* (pp. 284–300). Routledge.

Bies R. J., & Moag J. S. (1986). Interactional justice: Communication criteria of fairness. In Lewicki R. J., Sheppard B. H., & Bazerman, M. H. (Eds.), *Research on negotiation in organizations* (Vol. 1, pp. 43–55). JAI Press.

Bonner, S. M. (2016). Teachers' perceptions about assessment: Competing narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 21–39). Routledge

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. New York, NY: Cambridge University Press.

Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record, 106*(3), 429–458.

Brookhart, S. M. (2013). Classroom assessment in the context of motivation theory and research. In J. H. McMillan (Ed.), *Sage Handbook of research on classroom assessment* (pp. 35–54). Sage.

Brookhart, S. M. (2018). Learning is the primary source of coherence in assessment. *Educational Measurement: Issues and Practice*, 37, 35–38. https://doi.org/10.1111/emip.12190

Brookhart, S. M. (2024). Commentary: Where does classroom assessment fit in educational measurement?. *Educational Measurement: Issues and Practice*, *43*(3), 18-22. https://doi.org/10.1111/emip.12626

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research 86,* 803–848. https://doi.org/10.3102/0034654316672069

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. Assessment in Education: Principles, Policy & Practice, 11(3), 301-318. https://doi.org/10.1080/0969594042000304609

Brown, G. T. L. (2006). Teachers' conceptions of assessment: validation of an abridged instrument. *Psychological Reports 99*, 166–170. Doi:10.2466/Pr0.99.1.166-170

Colquitt, J. A., Long, D. M., Rodell, J. B., & Halvorsen-Ganepola, M. D. (2015). Adding the "in" to justice: A qualitative and quantitative investigation of the differential effects of justice rule adherence and violation. *Journal of Applied Psychology, 100*(2), 278.

Coombs, A., DeLuca, C., LaPointe McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education, 71*, 134–144. https://doi.org/10.1016/j.tate.2017.12.010

Coombs, A., DeLuca, C., & MacGregor, S. (2020). A person-centered analysis of teacher candidates' approaches to assessment. *Teaching and Teacher Education, 87*, 102952. https://doi.org/ 10.1016/j.tate.2019.102952

Costello, A., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(7), 1-9.

Cronbach, L. J., & P. E. Meehl. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52, 281–302. doi:10.1037/h0040957

Daniels, L. M., Pelletier, G., Radil, A. I., & Goegan, L. D. (2021). Motivating assessment: How to leverage summative assessments for the good of intrinsic motivation. In Nichols, S. L, & Varier, D. (Eds.), *Teaching on assessment* (pp. 107–128). Information Age Publishing: Charlotte, NC.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability,* 1–22. https://doi.org/10.1007/s11092-015-9233-6

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, *21*(4), 248-266. https://doi.org/10.1080/10627197.2016.1236677

DeLuca, C., Willis, J., Cowie, B., Harrison, C., & Coombs, A. (2024). Learning to assess: Cultivating assessment capacity in teacher education. Springer Nature. https://doi.org/10.1007/978-981-99-6199-3

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

Eaton, S. E. (2024). Comprehensive academic integrity (CAI): an ethical framework for educational contexts. In S. E. Eaton (Ed.), *Second handbook of academic integrity* (pp. 1-14). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54144-5_194

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, *4*(3), 272.

Fan, X., Johnson, R., Liu, J., Zhang, X., Liu, X., & Zhang, T. (2019). A comparative study of pre-service teachers' views on ethical issues in classroom assessment in China and the United States. *Frontiers of Education in China, 14*(2), 309–332.

Feldman, J. (2019). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin.

Forsell, J., Forslund Frykedal, K., & Hammar Chiriac, E. (2020). Group work assessment: Assessing social skills at group level. *Small Group Research, 51*, 87-124. doi:10.1177/1046496419878269

Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. In R. Cropanzano (Ed.), *Justice in the workplace: Approaching fairness in human resource management* (pp. 79–103). Mahwah, NJ: Erlbaum.

Government of Ontario. (2025). Facts about elementary and secondary education. https://www.ontario.ca/page/facts-about-elementary-and-secondary-education

Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education, 23*, 999-1011. doi:10.1016/j.tate.2006.04.042

Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning. https://cengage.com.au/product/title/multivariate-data-analysis/isbn/9781473756540

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. S*tructural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10. 1080/10705519909540118

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4 ed., Vol. 29, pp. 17–64). Westport, CT: American Council on Education and Praeger.

Kazemi, A., & Törnblom, K. (2008). Social psychology of justice: Origins, central issues, recentdevelopments, and future directions. *Nordic Psychology*, 60, 209–234. https://doi.org/10.1027/1901-2276.60.3.209

Lemon, J. (2006). Plotrix: A package in the red light district of R. *R-News*, *6*(4), 8–12. https://www.r-project.org/doc/ Rnews/Rnews_2006-4.pdf

Leventhal, G. (1980). What should be done with equity theory? New approaches to the study of justice in social relationships. In K. Gergen, M. Greenberg, & R. Willis (Eds.), *Social exchange: Advances in theory and research* (Vol. 9, pp. 27–55). New York: Plenum Press.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education, 33*, 263-275. doi:10.1080/02602930701292548

Maxwell, G. S. (2023). Assessment and accountability: Global trends and future directions in 21st-century competencies. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International encyclopedia of education* (4th ed., pp. 313–320). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.09069-2

McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *Sage Handbook of research on classroom assessment* (pp. 3–16). Sage.

McNamara, A., Arino de la Rubia, E., Zhu, H., Ellis, S., & Quinn, M. (2021). skimr: Compact and flexible summaries of data (R package version 2.1.3) [Computer software]. https://cran.r-project.org/web/packages/skimr/index.html

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

Mellenbergh, G. J. (2011). *A conceptual introduction to psychometrics*. Den Haag, The Netherlands: Eleven International.

Murillo, F. J., & Hidalgo, N. (2020). Fair student assessment: A phenomenographic study on teachers' conceptions. *Studies in Educational Evaluation, 65*, 1-10. doi:10.1016/j.stueduc.2020.100860

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill

Ontario College of Teachers (OCT). (2020). Membership demographics. https://reports.oct.ca/en/2020/Statistics/Membership-Demographics

Ontario College of Teachers (OCT). (2025). Teaching divisions. https://www.oct.ca/public/professional-standards/how-teachers-are-certified/initial-teacher-education/teaching-divisions

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: a review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 247–266). New York: Routledge.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. https://doi.org/10.1016/j.tate.2019.05.003

R Core Team. (2024). R: A language and environment for statistical computing (version 4.4.0) [Computer software] (4.4.0). R Foundation for Statistical Computing. https://www.r-project.org/

Revelle, W. (2020). How to: Use the psych package for factor analysis and data reduction. *Northwestern University, Department of Psychology: Evanston, IL, USA*.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate behavioral Research*, *14*(4), 403-414.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–21. 10.18637/jss.v048.i02

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282.

Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *Annals of the American Academy of Political and Social Science, 683*(1), 183–200. https://doi.org/10.1177/0002716219843818

Sonnleitner, P., & Kovacs, C. (2020). Differences between students' and teachers' fairness perceptions: Exploring the potential of a self-administered questionnaire to improve teachers' assessment practices. *Frontiers in Education, 5*, 1-17. doi:10.3389/feduc.2020.00017

Steiner, M. D., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, *5*(53), 2521.

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: L. Erlbaum Associates.

Thurlow, M. L. (2014). Instructional and assessment accommodations in the 21st century. In L. Florian (Ed.), *The Sage handbook of special education* (Vol. 2, pp. 597–612).

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vectors of the mind*. Chicago, IL: University of Chicago.

Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), SAGE Handbook of Research on Classroom Assessment (pp. 125-144). SAGE Publications.

Tierney, R. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation, 43*, 55-69. doi:10.1016/j.stueduc.2013.12.003

Tinsley, H. E. A., & Tinsley, D. J. (1987). Use of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, 34, 414-424. https://doi.org/10.1037/0022-0167.34.4.414

Wang, W. C., Chen, H. F., & Jin, K. Y. (2015). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement, 75*(1), 157-178.

Wijsen, L. D., Borsboom, D., & Alexandrova, A. (2022). Values in psychometrics. *Perspectives on Psychological Science, 17*(3), 788–804. https://doi.org/10.1177/17456916211014183

Vogl, E., & Pekrun, R. (2016). Emotions that matter to achievement: Student feelings about assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 111–128). New York: Routledge.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher education*, *58*, 149-162. https://doi.org/10.1016/j.tate.2016.05.010

Ylonen, A., Gillespie, H., & Green, A. (2018). Disciplinary differences and other variations in assessment cultures in higher education: exploring variability and inconsistencies in one university in England. *Assessment & Evaluation in Higher Education, 43*(6), 1009-1017. https://doi.org/10.1080/02602938.2018.1425369

Appendix 1: Classroom Assessment Fairness Inventory

Scenario 1: Groupwork Mr. Chu highly values student groupwork. Based on his initial assessments, Mr. Chu formed each group with three students from different ability levels: struggling, average, and high performing students. He believed that high performing students contribute to the learning of other group members. Each group worked on their projects and prepared a final presentation. Mr. Chu has left it to groups to discuss how to distribute workload and allowed students to discuss with him if they had issues over group dynamics. As a response to students' questions about assessment, Mr. Chu provided an overview of the project to students but not a rubric showing how he will assess students' groupwork. Mr. Chu encouraged group members to work hard as all group members will receive the same grades as a reflection of group performance and cooperation. Several who were not satisfied with their grades appealed, but Mr. Chu did not accept their complaints.	
To what extent do you consider each of these actions to be fair based on the above scenario?	
Actions	**Relevant underlying principle**
1. Mr. Chu selected group members based on mixed ability.	Consequence
2. Mr. Chu did not provide students a choice in selecting their group members.	Voice
3. Mr. Chu allowed students' complaints over group dynamics.	Voice
4. Mr. Chu was not detailed in communicating how he will assess students' groupwork.	Adequate communication
5. Mr. Chu gave the same grades to all group members.	Equality
6. Mr. Chu did not give individual grades for each group member based on their contributions and learning.	Equity
7. Mr. Chu did not justify his grades to students who appealed.	Justification

Scenario 2. Exam

Mr. Ahmed announced that the class would have an exam the day before winter break (in 5 days). Students preferred moving the exam date because they had many assignments for other subjects that were also due on the same date. Even so, Mr. Ahmed was firm on his decision as moving the exam date back would create more intensive workload later in the year. Mr. Ahmed did not explicitly state what would be on the exam. However, he did include a mix of easy and difficult questions to give all students an opportunity to show their learning. He also provided accommodations (e.g., more time) to students with disabilities and English language learners. In general, Mr. Ahmed is a lenient teacher in grading compared with other teachers in the school who teach the same subject. On the exam, all students complained about two questions that were not covered during the course. Mr. Ahmed *harshly responded* that students should be able to answer the two questions from what had been taught. For a few students who missed the exam date, Mr. Ahmed decided to give another exam opportunity after winter break.

To what extent do you consider each of these actions to be fair based on the above scenario?

Actions	**Relevant underlying principle**
1. Mr. Ahmed held firm on the exam date.	Reasonableness
2. Mr. Ahmed did not explicitly state what would be on the exam.	Adequate communication
3. Mr. Ahmed included in the exam both easy to difficult questions.	Equity
4. Students with disabilities and English language learners received accommodations for the exam (e.g., more writing time).	Need
5. Mr. Ahmed graded his students more leniently than other teachers.	Consistency
6. Mr. Ahmed did not remove the two questions on the content that were not taught before.	Correctability
7. Mr. Ahmed did not respond to students' complaints with a respectful tone.	Respect
8. Mr. Ahmed gave another exam opportunity to students who missed the exam.	Need (Equality, Consistency)

Scenario 3: Cheating

Ms. Johnston is very strict when she catches a student cheating. However, she did not tell students her policy on cheating at the beginning of the year. One student was caught cheating on an exam and Ms. Johnston decided to give the student a grade of zero. Ms. Johnston did not give the student an opportunity to explain the reasons for cheating before making her decisions. She explained to the class that cheating is unfair to other students and asked the student to leave the classroom. The exam constituted 20% of students' final grade. After the exam, she met with the student and explained that cheating is ethically wrong, is unfair in relation to the classmates, and she would punish anyone who cheats. The student provided a reason for their behaviour and apologized.

To what extent do you consider each of these actions to be fair based on the above scenario?

Actions	**Relevant underlying principle**
1. Ms. Johnston was not transparent about her cheating policy at the beginning of the course.	Transparency
2. Ms. Johnston did not give the student an opportunity to explain the reasons for cheating before making her decision to give zero.	Voice
3. Ms. Johnston gave a zero because cheating is unfair to other students' efforts.	Equity
4. Ms. Johnston gave a zero because cheating is ethically wrong.	Ethicality
5. Ms. Johnston gave a zero to signal that anyone who cheats should be punished for this action.	Consequence
6. Ms. Johnston asked the student to leave the classroom in front of other students.	Respect
7. Ms. Johnston explained her cheating decision to the student.	Justification
8. Ms. Johnston did not forgive the student's cheating this time.	Need

Scenario 4: Grading

Ms. Mendes had students from diverse backgrounds in her classroom. She treated all her students respectfully during classroom teaching, assessment, and interactions. Ms. Mendes informed students that she would give grades based on student achievement. 70% of students' grades were from multiple tests during the course plus 30% for students' individual essays. Ms. Mendes communicated test results in one week after handing the test in. Due to busy schedule, she would sometimes allow students to appeal their grades if there was enough time in class. She would fully explain her grading for students who spoke up looking for their grade adjustments. At the end of the course, Ms. Mendes adjusted the grades of failing students with at-risk backgrounds to support their success. She also increased marks for a few students to ensure admission into their desired universities. However, she lowered the grades of a few disruptive students who interrupted the classroom learning.

To what extent do you consider each of these actions to be fair based on the above scenario?

Actions	**Relevant underlying principle**
1. Ms. Mendes treated students respectfully during classroom assessment.	Respect
2. Ms. Mendes largely considered student achievement of learning objectives in her grading.	Equality
3. Ms. Mendes detailed her grading criteria, with test scores making up 70% of a student's grade.	Transparency
4. Ms. Mendes communicated test results in one week after handing the test in.	Timeliness
5. Ms. Mendes would sometimes allow students to discuss their grades if there was enough time in class.	Voice
6. Ms. Mendes gave adequate justification for students who spoke up for their grades.	Justification
7. Ms. Mendes adjusted the grades of failing students with at-risk backgrounds.	Need
8. Ms. Mendes considered students' future university admissions to adjust grades.	Consequence
9. Ms. Mendes considered student misbehaviour (e.g., disruptions) in her grading.	Bias suppression

Scenario 5: Feedback

Mr. Dembe has asked students to write an essay about their science lab project. He has shared with students a clear rubric that he will use to assess students' essays. Four days after the essay submission deadline, Mr. Dembe got back to students with his feedback and expressed that students can contact him for further discussion of his feedback. Students noticed that Mr. Dembe gave more feedback to students with good quality essays as well as his favourite students than students who had handed in essays that were of low quality. Mr. Dembe *harshly* explained that he had given variable feedback on the essays based on the amount of effort he deemed each student had put in completing essays.

To what extent do you consider each of these actions to be fair based on the above scenario?

Actions	**Relevant underlying principle**
1. Mr. Dembe provided a clear rubric for assessing students' essays.	Transparency
2. Mr. Dembe provided feedback after four days of essay submissions.	Timeliness
3. Mr. Dembe gave students a chance to further discuss his feedback.	Voice
4. Mr. Dembe provided feedback based on the amount of effort each student has put in the work.	Equity
5. Mr. Dembe did not provide more feedback to students who had weaker performance.	Need
6. Mr. Dembe did not treat students respectfully in his feedback procedure.	Respect
7. Mr. Dembe gave more feedback to his favourite students.	Bias suppression
8. Mr. Dembe explained that his feedback procedure was based on the amount of effort each student has put in.	Justification