Fairness in Classroom Language Assessment from EFL Teachers' Perspective

Somaye Tofighi *   **Mohammad Ahmadi Safa****

Abstract

The significance of fairness in assessment is not a subject of doubt for educationalists even with fundamentally contrastive perspectives, but the study of methodological means for safeguarding this concept in educational classroom contexts has only received the scant attention of the researchers and educationalists. On this basis, this study intended to conceptualize foreign language (FL) classroom assessment fairness from EFL teachers' perspective. For this purpose and as a part of a larger enterprise, the researchers first strived for the creation and validation of a Classroom Assessment Fairness Scale (CAFS). This process was informed by the related literature and applied a critical incident technique. To verify the scale's validity and reliability, it was first expert viewed and modified accordingly; next pilot-tested on a group of 250 teachers, and finally, the amassed data were subjected to exploratory and confirmatory factorial analyses and internal consistency measures. The validated scale was taken by 120 Iranian EFL teachers. Descriptive analyses indicated that the EFL teachers were familiar with fairness principles in classroom assessment practices. They were aware of the significance of the opportunity to learn, the opportunity to demonstrate learning, the no-harm principle, the constructive classroom environment, and transparency as the constructive elements of classroom assessment practices' fairness. However, the same familiarity and awareness were not guaranteed for the score pollution component. The results further demonstrated that the teachers' gender, educational degree, teaching experience, and teaching context led to statistically significant differences among EFL teachers' attitudes to classroom assessment fairness.

Classroom assessment is an effective approach to gathering data about student learning, and a teacher's professional role in assessment is to employ high-quality and fair

How to cite this article:

Tofighi, S., & Ahmadi Safa, M. (2023). Fairness in Classroom Language Assessment from EFL Teachers' Perspective. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *42*(2), 81-110. Https://doi.org/10.22099/tesl.2023.46825.3173

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

82

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

assessment information to make judgments regarding their own instruction and the students' learning (Fan et al., 2020). On this basis, fairness of assessment practices in the classroom context has been recently viewed as a fundamental tenet of classroom assessment (Rasooli et al., 2019, Shah Ahmadi & Ketabi, 2019) partly on the grounds that social and educational initiatives in the 21st century toward equity, diversity, and educational inclusion have given new life to investigations into assessing fairness (Dorans & Cook, 2016; Herman & Cook, 2019; Kunnan, 2018; Tierney, 2016).

In a classroom context, the fair assessment makes the critical value of reverence to retain learners' dignity and welfare (Pettifor & Saklofske, 2012), improves students' academic achievement (Molinari et al., 2013), nullifies possible harm to students (Green & Johnson, 2010), and increases learning motivation (Chory-Assad, 2002). Among other probable reasons, such significant impacts are due to the fact that a fair assessment delivers valid information for a specific goal for all learners, independent of their background characteristics; it provides all learners an equal chance to demonstrate their knowledge, unhampered by factors unrelated to the construct being assessed (Herman & Cook, 2019). Unfair assessment, on the other hand, undermines students' trust in teachers and harms the relationship between students and teachers (Green et al., 2007), deteriorates pedagogical relationships, and impedes the potential for students' learning (Aitken, 2012).

Despite the attested significance of fairness in assessment, efforts to define and theorize fairness in educational contexts have been surprisingly limited (Nisbet, 2017; Tierney, 2013). Nisbet (2017, p. 6) concludes that a dearth of conceptual scrutiny of fairness prevails, and defining fairness is shunned by most authors. Motivated by the lack of definitive conceptualization of classroom assessment fairness, and given that the literature has mostly focused on advancing fairness theory and little research has focused on the teachers' and learners' attitudes towards fair classroom assessment (Rasooli et al., 2022) this study delved into Iranian EFL teachers' attitudes toward fairness in classroom assessment practices. As Hidri (2016) maintains, understanding more about teachers' attitudes helps to modify their practices, thereby improving student learning.

Literature Review

Fairness is considered the axiomatic feature of classroom assessment theory and quality assessment practices (Baniasadi et al., 2022; Liu et al., 2016; Rezai et al., 2021; Tierney, 2013; Xu & Brown, 2016) which has been constantly accentuated as a fundamental quality and an indispensable aspect of assessment literate teachers (DeLuca, 2012; DeLuca et al., 2016 b; Xu & Brown, 2016). A fair assessment is defined as "an assessment practice that is responsive to individual characteristics and testing contexts so that test scores will yield valid interpretations for intended uses" (AERA et al., 2014, p. 50). Consistently, Mislevy (2018) suggests fairness as a rational basis for adjusting to the test-takers' interests and prior knowledge. Mislevy argues that it is important to understand how work is performed in light of the person, background, tasks, settings, and circumstances.

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

83

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Researchers have tried to elucidate the fundamental characteristics of fair assessment practices. For instance, Peters et al. (2017) contemplate assessment practices to be fair if a) they are employed as a diagnostic tool rather than a mechanism for classification, b) to improve student learning rather than as an external tool to measure students' performance, and c) to even out the overall students' evaluation rather than as a punishment instrument for students who do not fulfill the intended requirements. Scores from a fair test represent the same construct and have the same meaning for all the learners for whom the test is developed and neither advantage nor disadvantage learners based on attributes that are unrelated to the construct being tested (Herman & Cook, 2019). Fairness in classroom assessment entails avoiding bias and ensuring equity, holding clear learning expectations, using diverse techniques to elicit learning, welcoming individual needs, assisting learners in acquiring how to accomplish the assessment task and presenting the learners with detailed and constructive feedback (Camili, 2006; Cowie, 2015; Tierney, 2014; Zhang & Burry-Stock, 2003).

Studies have also confirmed that fair assessment leads to EFL teachers' credibility. Chory (2007), for instance, concluded that instructors who were considered to be fair had credibility. Similarly, students of fair instructors are shown to demonstrate extreme motivation and effective learning (Chory-Assad, 2002) and engagement (Berti et al., 2010). On the other hand, the perception of unfairness has been linked to absenteeism (Ishak & Fin 2013), anger and violence (Chory-Assad & Paulsel 2004), and cheating (Murdock et al., 2007).

Despite the briefly reviewed merits of fair assessment, as Nisbet (2017) maintained, the conceptual definition of fairness is yet in need of empirical clarification. Focusing on classroom assessment fairness studies, Rasooli et al. (2018) reviewed theoretical and empirical literature and expressed surprise that only 8 of the 50 reviewed studies specifically characterized fairness. They concluded that concepts such as justice, equity, equality, equitability, ethics, and nondiscriminatory practices are used in conjunction with or interchangeably with fairness in these definitions. They also confirmed Tierney's (2013) conclusion that there is an absence of a transparent definition of the classroom assessment fairness concept in the literature. Tierney (2013) indicates that 'fuzziness' and 'circularity' encircle fairness in the classroom assessment literature as scholars resort to various terminologies like ethics, bias, justice, and objectivity.

To partially address the underscored ambiguity of the concept, the present study aimed at the Iranian EFL teachers' attitudes towards fairness in classroom assessment and in an attempt to operationally distance themselves from the perplexity of the fuzzy concept of classroom assessment fairness, the researchers regarded fairness as the opportunity to learn, transparency, the opportunity to illustrate learning, no harm principle, and avoiding score pollution (Downing & Haladyna, 2004; Tierney, 2013, 2014, 2016).

The opportunity to learn generally considers fairness ahead of assessment and is intimately associated with education (Rasooli et al., 2018). It simply encompasses being exposed to test content or refers more widely to the congruence between curriculum and

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

84

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

assessment (Tierney, 2016). Moreover, it entails various learning opportunities catering to students' unique learning styles, competencies, and exceptionalities (Lantolf & Poehner, 2013). An opportunity to demonstrate learning is the provision of multiple, different, fair, and purposeful opportunities for learners to exhibit their learning (Tierney, 2016). Supporting the opportunity to demonstrate the learning principle of classroom assessment fairness, Mauldin (2009) empirically verified that assessment is fair when multiple assessment opportunities are presented.

Transparency refers to the idea that learners should know how their work will be judged before an assessment starts (Tierney, 2016). Tierney recommends that by clearly articulating learning and assessment requirements, transparency contributes to learners' access to opportunities to learn and opportunities to exhibit learning. Do no harm principle is intimately associated with the constructive classroom environment dimension (Tierney, 2013, 2014, 2016). Tierney (2014, 2016) highlights the role of power dynamics and courteous relationships in creating a constructive classroom environment.

Avoid score pollution underscores the conviction that learners' scores should exclude construct-irrelevant (Green et al., 2007) and construct-underrepresented factors (Rasooli et al., 2018). It is suggested that any technique that enhances performance without simultaneously improving learners' mastery of the content gives rise to score pollution. Students find assessments unfair when instructors and test-makers base their judgments on information that is not appropriate, sufficient, and reliable (Scott et al., 2014; Alm & Colnerud, 2015). Teachers do not generally agree on the factors polluting learners' scores; hence, the provision of a set of standards and guidelines to recognize and address the factors that lead to score pollution and, ultimately, unfair classroom assessment practices is still a real challenge (Pope et al., 2009).

Against the backlog of the briefly reviewed theoretical underpinnings of educational assessment fairness, a growing body of literature is developing theoretical conceptualizations of fairness based on empirical data that are drawn from instructors' and students' perceptions of fairness in classroom evaluation (Azizi, 2022; Darabi Bazvand & Rasooli, 2022; Fan et al., 2019; Fan, Liu & Johnson, 2020; Murillo & Hidalgo, 2020, Rasooli et al., 2018, 2022; Torkey & Sayed Haider, 2017), a brief description of the findings of only some of the most recent and relevant studies is in order in the following paragraphs.

In the context of Iranian higher education, Darabi Bazvand and Rasooli (2022) looked at the experiences of fairness in summative assessments among postgraduate university students and concluded that two themes of "equity and interactional fairness" were quite essential determiners of assessment fairness or unfairness. In a rather different context, Azizi (2022) examined the fairness of assessment practices in online learning contexts and named three overarching categories of distributive justice, procedural justice, and interactional justice as the main factors contributing to assessment fairness in online learning contexts. Still from another perspective, Rasooli et al. (2022) investigated the driving forces behind teachers' ideas of fairness and interviewed 27 experienced high school teachers. The findings showed that three themes of individual mechanisms, social

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

85

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

mechanisms, and dialectical interactions between individual and social mechanisms- had an impact on teachers' perceptions of fairness in classroom assessment. The findings highlighted the interaction between the personal philosophies and experiences of teachers and their interactions with the social contexts of society, schools, and classrooms to shape teachers' conceptions and practices of fairness in classroom assessments.

On the other hand and focusing on the students' perspectives rather than the teachers', Rasooli et al. (2019) explored university students' critical incidents of fairness and unfairness and their affective and behavioral responses to such events. The findings revealed that distributive, procedural, and interactional justice principles influenced students' views of classroom assessment fairness as well. In their conceptualization of fairness, the students looked at the distribution of results, the methods for outcome distributions, interpersonal connections, and communication protocols. When discussing fair incidents, students also expressed good emotions such as contentment, satisfaction, feeling appreciated, and hopefulness, however when describing unfair incidents, they tended to express negative emotions such as wrath, distress, disappointment, and humiliation.

From another perspective, the degree of agreement between different stakeholders' perspectives on classroom assessment fairness and ethicality and those of classroom assessment experts was the subject of investigation. For example, Liu et al. (2016) evaluated pre-service teachers' perspectives on the ethics of classroom assessment techniques in the United States and China. They revealed that pre-service teachers' judgments of the assessment ethics did not align with the principles put forward in the classroom assessment literature. In another study, Fan et al. (2020) looked at the classroom assessment ethical judgments of Chinese university instructors. A poll of 555 professors was administered with fifteen scenarios describing professors' opinions on ethics in assessment practices. The results confirmed that the professors shared similar perspectives with experts in providing manifold assessment opportunities. Nevertheless, they appeared to be in low harmony with assessment professionals regarding grading communication, confidentiality, and grading activities. Fan et al. (2020) also studied Chinese university students' opinions on the ethicality of classroom assessment procedures. This rather large-scale project explored the opinions of 2711 college students from 177 institutions around China and reported that college students demonstrated diverse degrees of agreement with professionals in various assessment situations. Further, it seemed extremely challenging for them to distinguish unethical assessment techniques from ethical assessment practices.

The brief literature reviewed above confirms Murillo and Hidalgo (2020) and Rasooli et al. (2022) who contended that there is a dearth of empirical research on perceptions of fair assessment with different stakeholders in general and the teachers, in particular. On this basis and given that the attitudinal perspectives of the teachers are the front-line experts to honor the primary obligation for ensuring inclusive teaching and assessments practices (Lupart & Webber, 1996), this study looked at how EFL teachers' attitudes towards fairness in the classroom assessment were affected by factors including

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**86** **Somaye Tofighi**

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

their educational background, years of teaching experience, and even gender. For this purpose, the following research questions were raised:

1. What are Iranian EFL teachers' attitudes toward fairness in classroom assessment practices?
2. Are there statistically significant differences among Iranian EFL teachers' attitudes toward fairness in classroom assessment practices in terms of their gender, educational degree, years of teaching experience, and teaching context?

Method

Participants

Two groups of participants took part in this study. The first group comprised of one hundred and twenty Iranian EFL teachers from Kermanshah, and Fars provinces who were sampled conveniently. Their age range was between 22- 42 years. In addition, the second group of participants was 250 EFL teachers from all over the country who partook in the pilot phase of the study with their own consent. They were within the age range of 23-47. Announcements were made in Iranian EFL teachers' WhatsApp groups, and the study's aims were elucidated. The Internet link to a questionnaire was sent to those who consented to partake in the pilot phase of the study. The demographic information of the participants for both the main and pilot phases is displayed in Table 1 below.

Table 1
Demographic Information of the Participants

		Gender		Years of teaching experience			Educational degree			Teaching context		
		Male	Female	1-5	6-10	More than 10	B.A	M.A	Ph.D.	Public school	Language institute	University
Main study	N	51	69	54	45	21	39	67	14	62	45	13
	%	42	58	45	37.5	17.5	32.5	55.8	11.7	51.7	37.5	10.8
Pilot study	N	88	162	90	115	45	88	119	43	135	89	26
	%	35.2	64.8	36	46	18	35.2	47.6	17.2	54	35.6	10.4

Instruments

For data collection purposes, the researchers made and validated a classroom assessment fairness scale and applied critical incident techniques for this purpose. The description of the design and validation steps of the scale, followed by a brief account of the critical incident technique are following.

Critical Incident Technique: Applying this means, the researchers called for Iranian EFL teachers' fairness incidents in language classroom assessment practices. The call was sent to 150 EFL teachers via WhatsApp. They were requested to recall a time when they did something they assumed as un/fair assessment practice. Colnerud (1997) believes that an efficient method to investigate ethical and teaching dilemmas is to

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

87
Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

examine the ethical challenges that instructors encounter in their professional relationships with others.

Classroom Assessment Fairness Scale (CAFS): To assess teachers' perceptions of fairness in classroom assessment practices, the researchers developed and validated the CAFS (Appendix A). The Questionnaire comprised 35 items aligned with five dimensions of fairness in classroom assessment. The development of the scale was informed by the related literature and the EFL teachers' critical incident reports. To begin with, the researcher developed an item pool for fair classroom assessment practices based on the related literature. Simultaneously, a critical incident technique was applied to address the problem of lack of unanimity on what teachers consider fair assessment practices across contexts (Liu et al., 2016). The researchers asked 150 EFL teachers to explain a time when they said or did something they perceived as fair/unfair. The data from the critical incident technique were thematically analyzed to identify the Iranian EFL teachers' classroom assessment fairness principles. To analyze the teachers' critical incidents, a deductive approach (Berg, 2001) was applied in that the initial item pool was used as the benchmark to analyze teachers' accounts. The researchers read and reread the incident accounts and coded the resulting extracted principles of classroom assessment fairness. The incidents that seemed similar to each other were grouped under the same principle. The codes for each fairness principle were tallied to calculate their frequency, and the ones with the most frequencies were selected as the questionnaire items (Weber, 1990).

In sum, out of 150 EFL teachers who replied to critical incident calls, 66 teachers described a fair incident and 84 described as an unfair classroom assessment experience. Five selected dimensions of fairness, including the opportunity to learn, the opportunity to demonstrate learning, transparency, no harm principle, constructive learning environment, and avoiding score pollution (Downing & Haladyna, 2004; Tierney, 2013, 2014, 2016), were applied as the criterion framework to code teachers' responses, and each incident was located under each respective distinctive dimension. The content analyses resulted in that 17 incidents were relevant to the opportunity to learn factor, 13 incidents to the opportunity to demonstrate learning factor, 15 to transparency, 54 incidents were categorized under the no harm principle and constructive classroom environment, and finally, 51 incidents were related to avoiding score pollution.

Based on Dornyei's (2003) demarcation on allocating four items to each scale component, at least four items were devoted to each sub-scale in the first pool. For the first dimension which was "opportunity to learn", 7 items were included. The second dimension, "opportunity to demonstrate learning", was represented through 6 items. The third dimension was "transparency" and it was measured through 5 items. The fourth dimension was the "no harm principle and constructive classroom environment" which comprised 9 items. The last dimension was "avoid score pollution" and included 12 items. Items were on a 5-point Likert scale ranging from completely unfair to completely fair.

To ensure the content validity of the scale, three assessment and evaluation experts were requested to review the instrument, give feedback concerning its precision and

alignment with the identified fairness classroom assessment themes, and suggest revisions if necessary. Next, the modified scale containing 39 items was pilot administered to a group of 250 EFL teachers comparable to the study's target population to assess the construct validity and internal consistency of the scale.

To establish the factor structure of the scale principal components factor analysis (PCA) was applied to the obtained data from the pilot administration of the 39-item CAFS. Prior to PCA, the suitability of data for factor analysis was examined. Inspection of the correlational matrix manifested in the existence of numerous coefficients of .3 and above. The results of the sampling adequacy test (KMO = .84 > .6) revealed the adequacy of the size of the sample as it exceeded the criterion level of 0.6 (Pallant, 2013), and Bartlett's test of Sphericity was significant at $p < .01$, indicating that there was a strong interrelationship among items reinforcing the factorability of the correlation matrix (Table 2).

Table 2
KMO and Bartlett's Tests

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	Bartlett's Test of Sphericity		
	Approx. Chi-Square	Df	Sig.
.840	11877.863	741	.000

PCA results confirmed six components with eigenvalues above 1, explaining 25.46, 19.71, 10.84, 9.54, 8.82, and 4.95 percent of the variance, respectively (Table 3).

Table 3
EFA: Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.930	25.463	25.463	9.586	24.579	24.579	9.117
2	7.689	19.716	45.179	4.428	11.354	35.934	7.590
3	4.231	10.849	56.028	7.106	18.220	54.153	5.934
4	3.721	9.542	65.570	3.209	8.227	62.380	4.227
5	3.442	8.825	74.395	2.745	7.038	69.418	4.416
6	1.931	4.952	79.347				
…	.	.	.				
39	.026	.067	100.000				

Extraction Method: Maximum Likelihood.
a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

However, the inspection of the screeplot (Figure 1) revealed a clear break after the fifth component.
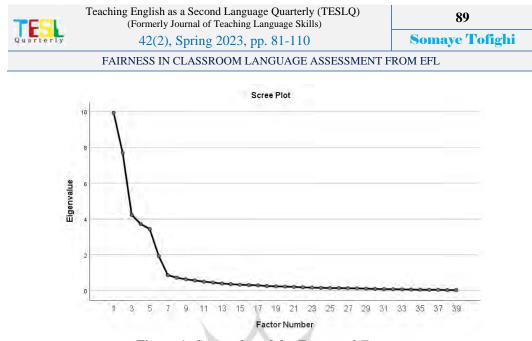
	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**89** **Somaye Tofighi**

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Figure 1. *Scree plot of the Extracted Factors*

Following Catell's (1966) scree test, five components were preserved for additional analysis. Next, the researchers re-ran the analysis with five fixed factors. The five-component solution explained a total of 69.42 percent of the variance, with components 1 to 5 explaining 25.46, 19.71, 10.84, 9.54, and 8.82 percent of the total variance, respectively (Table 4).

Table 4
EFA: Factor Correlation Matrix

Factor	1	2	3	4	5
1	1.000	.106	.342	.270	-.368
2	.106	1.000	.035	.058	-.092
3	.342	.035	1.000	.198	-.314
4	.270	.058	.198	1.000	-.234
5	-.368	-.092	-.314	-.234	1.000

Since there was a correlation between the factors, Promax rotation with Kaiser Normalization was performed to minimize the number of variables with high loadings on each factor. Table 5 displays the loadings of the 39 variables on the five factors extracted. The results indicated that the five components exhibited several strong loadings, and each variable loaded substantially on only one component. The factor loadings demonstrated that seven items were loaded on component 1, six items were subsumed under component 2, component 3 had a subset of five items, nine items were loaded on component 4, and component 5 contained twelve items (Table 5).

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**90** **Somaye Tofighi**

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Table 5

EFA: Pattern Matrix

	Factor				
	1	**2**	**3**	**4**	**5**
Q01					.453
Q02					.467
Q03					.717
Q04					.700
Q05					.607
Q06					.831
Q07					.419
Q08				.807	
Q09				.748	
Q10				.871	
Q11				.687	
Q12				.595	
Q13				.788	
Q14			.966		
Q15			.999		
Q16			.982		
Q17			.997		
Q18			.974		
Q19	.955				
Q20	.932				
Q21	.952				
Q22	.948				
Q23	.988				
Q24	.956				
Q25	.911				
Q26	.985				
Q27	.929				
Q28		.847			
Q29		.717			
Q30		.821			
Q31		.831			
Q32		.755			
Q33		.869			
Q34		.831			
Q35		.770			
Q36		.738			
Q37		.868			
Q38		.750			
Q39		.638			

Extraction Method: Maximum Likelihood.

Rotation Method: Promax with Kaiser Normalization.

Rotation converged in 6 iterations.

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

91

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

In order to assess and confirm the construct validity of the CAFS, CFA was conducted using IBM AMOS (version 24) statistical package. First, items with non-significant loadings in unstandardized estimation were to be discarded. Although the findings revealed that none of the items had such a poor loading on the factors, the standardized estimate for four items (items 4 and 6 in component 1 and 11 and 12 in component 2) was below the cut-off point of .5 (Hair et al., 2010). Hence, these items were removed from the scale, reducing it to 35 items.

Next, the goodness of fit indices of the model was assessed. Hu and Bentler (1999) assert that a number of requirements must be satisfied for a model to have an acceptable goodness of fit. These criteria, alongside the values obtained from the data, are displayed in Table 6. The results demonstrated acceptable to excellent goodness of fit.

Table 6
Goodness of Fit Indices

Criteria		Threshold			Evaluation
		Terrible	Acceptable	Excellent	
CMIN	1503.884				
Df	535				
CMIN/df	2.811	> 5	> 3	> 1	Excellent
RMSEA	.059	> 0.08	< 0.08	< 0.06	Excellent
CFI	.914	< 0.9	> 0.9	> 0.95	Acceptable
TLI	.904	< 0.9	> 0.9	> 0.95	Acceptable
SRMR	.062	> 0.1	> 0.08	< 0.08	Excellent

Next, each factor's composite reliability (CR) and convergent/discriminant validity were evaluated (Table 7). As mentioned above, all variables had composite reliability values above 0.7, indicating an acceptable reliability level (Hair et al., 2014). For all factors, the average variance explained (AVE) was safely above 0.5, which in turn confirms the convergent validity (Hair et al., 2014) of the scale. Moreover, the maximum shared variance (MSV) for each respective factor was below AVE, which also confirms the convergent validity (Hair et al., 2010). Finally, the square root of AVE (bold values under the Fornell–Larcker Criterion) for each factor was above its inter-correlations (not bold values under Fornell–Larcker Criterion) with other factors, indicating discriminant validity (Fornell & Larcker, 1981).

Table 7
Composite Reliability and Validity of the Factors

				Fornell – Larcker Criterion				
	CR	AVE	MSV	OL	ODL	Trans.	NHPCCE	SP
OL	0.833	0.521	0.036	0.722				
ODL	0.890	0.680	0.129	0.179	0.825			
Trans.	0.975	0.888	0.056	0.180	0.223	0.942		

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills)	**92**
TESL Quarterly	42(2), Spring 2023, pp. 81-110	**Somaye Tofighi**

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

				Fornell – Larcker Criterion				
	CR	AVE	MSV	OL	ODL	Trans.	NHPCCE	SP
NHPCCE	0.979	0.835	0.129	0.190	0.359	0.236	0.914	
SP	0.948	0.602	0.007	0.085	0.068	-0.037	0.058	0.776

Finally, based on the verified modified measurement model (Figure 2), Classroom Assessment Fairness Scale was confirmed to tap into five componental factors: 1. The opportunity to learn, 2. The opportunity to demonstrate learning, 3. Transparency, 4. No-harm principle and constructive learning environment, and 5. Avoiding score pollution. Figure 2 below represents the final verified model with the five components and their respective questionnaire items.
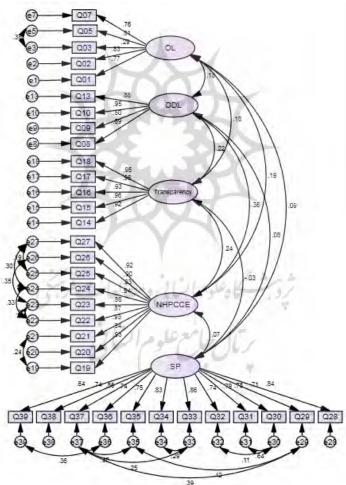

Figure 2. *The Final Modified CFA Model with Standardized Estimates*

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

93

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Data Collection Procedure

First, as is described above, a valid scale for measuring teachers' attitudes to classroom assessment fairness was designed and made. Next, an online version of the validated scale was made and posted to Iranian EFL teachers in several virtual groups in What's App. One hundred and twenty EFL teachers took the online version of the scale and data from these completed questionnaires were subjected to descriptive and inferential statistical analyses to answer the research questions.
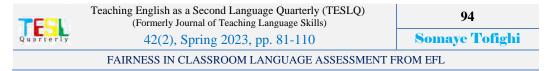
Results

Descriptive statistic analyses were applied to address the first research question exploring Iranian EFL teachers' attitudes toward the fairness of classroom assessment practices. The results of analyses for the opportunity to learn component indicated that participants' mean score of sub-items ranged from 3.24 to 4.22 (Table 8). The second item that tapped into postponing the final exam until the material was comprehensively covered had the highest mean of 4.22 and was strongly endorsed as fair and completely fair by most teachers (88.3 %). In addition, most teachers (86.6 %) assigned high fairness value to item one, which called for the similarity between classroom activities and test items. Additionally, item three, which called for including a few surprise items in the assessment, had the lowest mean of 3.24 and was confirmed unfair and completely unfair by just 32.5 percent of teachers.

Table 8
Descriptive Statistics for Opportunity to Learn Component

Item	N	Completely unfair	Unfair	Undecided	Fair	Completely fair	Mean
1. Similarity between classroom activities and test items	120	--	1.7	11.7	60.8	25.8	4.10
2. Postponing final exam till all the material is covered	120	--	.8	10.8	53.3	35.0	4.22
3. Including surprise items in the final exam	120	5.0	27.5	20.8	31.7	15.0	3.24
4. Administer a parallel form of the test	120	.8	2.5	21.7	54.2	20.8	3.91
5. Using methods that students have regularly encountered	120	.8	1.7	15.0	47.5	35.0	4.14

The analysis of the teachers' responses to the items subsumed under the opportunity to demonstrate the learning component demonstrated that items 6 and 9 which referred to drawing on many types of assessment forms and inclusion of a variety of activities in assessment, had the highest mean of 3.99 and 3.96 and the majority of teachers (78.4 % & 81.6 %) admitted the fairness and completely fairness of these practices respectively.

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**94** **Somaye Tofighi**

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

While item 7, designating observation as the sole method to assess, had the lowest mean of 3.63 and was regarded as unfair and completely unfair by only 22.5 percent of teachers (Table 9).

Table 9
Descriptive statistics for Opportunity to Demonstrate Learning Component

Item	N	Completely unfair	Unfair	Undecided	Fair	Completely fair	Mean
				Likert Scale			
6. Including many types of assessment forms	120	--	8.3	13.3	49.2	29.2	3.99
7. Using observation as the sole method to assess	120	4.2	18.3	15.0	35.0	27.5	3.63
8. Assessing oral proficiency through different activities	120	.8	9.2	18.3	52.5	19.2	3.80
9. Giving students a variety of activities	120	.8	6.7	10.8	58.3	23.3	3.96

As for the transparency component, Table 10 shows that item 13, addressing clarifying the policy, procedures, and decisions, had the highest mean of 4.12 and was approved fair and completely fair by 85.9 percent of teachers. In addition, 85.8 percent of the teachers viewed illuminating the policy of class attendance (item 14), and 82.5 percent considered a priori sharing of the rubrics for each task (item 11) as fair and completely fair assessment practices. Item 10, stating how a task would be graded, was approved fair and completely fair by 75.9 percent of the teachers. However, item 12, which addressed keeping the details of the student's performance assessment rubrics confidential, had the lowest mean of 3.32, and only 37.5 percent of the teachers were certain about its unfairness and complete unfairness.

Table 10
Descriptive Statistics for Transparency Component

Item	N	Completely unfair	Unfair	Undecided	Fair	Completely fair	Mean
				Likert Scale			
10. Stating how a task would be graded	120	--	2.5	21.7	56.7	19.2	3.92
11. A priori sharing of the rubrics of each task with students	120	.8	3.3	13.3	59.2	23.3	4.00
12. Keeping the details of the students' performance assessment rubric confidential	120	8.3	29.2	10.0	26.7	25.8	3.32
13. Clarify policy, procedures, and decisions	120	--	2.5	11.7	56.7	29.2	4.12
14. Illuminating the policy of the class attendance	120	.8	1.7	11.7	62.5	23.3	4.05

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

95

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

As Table 11 displays, for the no harm principle and constructive classroom environment component, respecting students' privacy (item 21) had the highest mean of 4.32 and was rated fair by 50 percent of the teachers and completely fair by 43.3 percent of them. The fairness and complete fairness of slowing down the speed of instruction to adjust to students' understanding (item 16) and announcing the assessment time in advance (item 23) were recognized as fair by most of the teachers (86.7 % & 89.2%, respectively). Additionally, they deemed spending time conferencing with students to explain their performance's strengths and weaknesses (item 15) and showing concern for students' learning (item 20) fair and completely fair assessment practices (81.6 % & 83.4 %, respectively). On the other hand, only 51.2 percent of teachers acknowledged the fairness and complete fairness of employing peer assessment and it had the lowest mean of 3.15 (item 19). Item 18 which called for requiring students to rate others' assignments was admitted fair and completely fair by 65 percent of teachers.

Table 11
Descriptive Statistics for No Harm Principle and Constructive Classroom Environment Component

Item	N	Completely unfair	Unfair	Undecided	Fair	Completely fair	Mean
				Likert Scale			
15. Conferencing with students to explain their performance's strengths and weaknesses	120	--	4.2	14.2	58.3	23.3	4.00
16. Slowing down the speed of instruction to adjust to learners' understanding	120	--	1.7	11.7	51.7	35.0	4.20
17. Sharing with students the rubrics for each task	120	--	2.5	15.0	68.3	14.2	3.94
18. Requiring students to rate each other's assignments	120	2.5	12.5	20.0	50.0	15.0	3.62
19. Employing peer assessment	120	9.2	20.8	20.8	44.2	5.0	3.15
20. Showing concern for students' learning	120	--	2.5	14.2	64.2	19.2	4.00
21. Respecting the privacy of students	120	.8	2.5	3.3	50.0	43.3	4.32
22. Scoring performance based on other modules or giving an alternate version of the same test to students who cheat	120	2.5	15.0	29.2	44.2	9.2	3.42
23. Announcing the test/ assessment time in advance	120	1.7	1.7	7.5	62.5	26.7	4.10

Table 12 indicates that for the avoiding score pollution component, item 31, addressing being immune to parental pressure to alter standards, had the highest mean of 4.06 and was indicated fair and entirely fair by 82.5 percent of the teachers. The results showed that only a few teachers acknowledged the unfairness and complete unfairness of lowering learners' scores for late submission of assignments (26.7 %), showing disruptive behavior (35.9 %), displaying naughty behavior (31.8 %), lack of regular class attendance

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

96

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

(55.9 %), choosing the wrong answer instead of leaving the answer blank (20 %) (Items 24, 26, 27, 29, & 34 respectively). Additionally, only 20 percent of teachers found giving additional credit opportunities to all learners, excluding the more knowledgeable ones (item 35) unfair and completely unfair assessment practice. On the other hand, item 30, referring to dedicating extra credit for volunteering in classroom activities, had the lowest mean of 2.30 and was admitted unfair and completely unfair by 69.2 percent of the teachers.

Table 12
Descriptive Statistics for Avoid Score Pollution Component

Item	N	Completely unfair	Unfair	Undecided	Fair	Completely fair	Mean
				Likert Scale			
24. Lowering learners' scores for belated submission of assignment	120	4.2	22.5	16.7	42.5	14.2	3.40
25. Modifying learners' scores based on other teachers' perceptions of them	120	.8	15.8	15.0	44.2	24.2	3.75
26. Changing students' scores who show disruptive behavior	120	6.7	29.2	17.5	34.2	12.5	3.16
27. Changing students' scores who show naughty behavior	120	3.3	27.5	18.3	34.2	16.7	3.33
28. Do not take into consideration the degree of neatness	120	2.5	30.0	15.8	41.7	10.0	3.26
29. Allocating a part of the score to learners' class attendance	120	9.2	46.7	19.2	22.5	2.5	2.62
30. Dedicating extra credit for volunteering in classroom activities	120	16.7	52.5	17.5	10.0	3.3	2.30
31. Not being influenced by parental pressure to alter standards or bend the rules	120	--	5.8	11.7	52.5	30.0	4.06
32. Basing each pupil's score on the group's outcome	120	.8	21.7	26.7	25.8	25.0	3.52
33. Counting class attendance as a part of the final score	120	9.2	54.2	20.8	15.0	.8	2.44
34. Deducting more points for an incorrect response	120	2.5	17.5	30.0	43.3	6.7	3.34
35. Providing additional credit opportunities to all learners, excluding the more knowledgeable ones	120	.8	19.2	15.0	36.7	28.3	3.72

Finally, a brief return to the descriptive statistic results for all five dimensions of CAFS indicates that the mean scores ranged from 3.24 to 3.92 (Table 13). Accordingly,

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**97** **Somaye Tofighi**

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

the mean scores were 3.92 for the opportunity to learn component, 3.84 for the opportunity to demonstrate learning, 3.88 for transparency, 3.86 for the no harm principle and constructive classroom environment, and for avoiding score pollution component it was 3.24. The total mean score of the teachers' attitude to CAFS was found to be 3.75.

Table 13
Descriptive Statistics of Teachers' Score on Five Components of CAFS

	N	Minimum	Maximum	Mean	Std. Deviation
Opportunity to learn	120	2.20	5.00	3.92	.51
Opportunity to demonstrate learning	120	2.25	5.00	3.84	.67
Transparency	120	2.40	5.00	3.88	.44
No harm principle and constructive classroom environment	120	2.33	4.67	3.86	.45
Avoid score pollution	120	2.33	4.67	3.24	.50
Fairness	120	2.81	4.58	3.75	.36

The second research question addressed the possibility of any statistically significant differences among Iranian EFL teachers' attitudes to the fairness of classroom assessment practices in terms of their gender, educational degree, teaching experience, and teaching context. First, to address the differences between male and female EFL teachers' attitudes to the fairness of classroom assessment practices, Independent samples *t*-Test was performed. Preliminary checks guaranteed that no assumptions were violated as Levene's test showed that there was no significant difference between the variances of the two groups of male and female teachers on CAFS (p=.07, p> .05). In addition, the results of the Kolmogorov-Smirnov test displayed that male and female EFL teachers' scores on CAFS were normally distributed (p=. 20,. 20, p>. 05). The results of *t*-test revealed that there was a statistically significant difference in scores for male (M= 3.55, SD= .29) and female (3.74, SD= .37; t(118)= -2.93, p= .00) teachers on CAFS (Table 14).

Table 14
Independent Samples Test Comparing Male and Female Teachers' Attitudes to CAFS

	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	3.21	.075	-2.93	118	.00	-.18	.06	-.31	-.06
Equal variances not assumed			-3.04	117.57	.00	-.18	.06	-.30	-.06

In addition, a one-way ANOVA was applied to investigate the possibility of any significant difference among Iranian EFL teachers' attitudes towards fairness in classroom assessment practices with respect to their educational degree. Concerning the data distribution normality and the equality of variances assumption of ANOVA analysis, Kolmogorov-Smirnov test verified that B.A, M.A., and Ph.D. holder teachers' scores on CAFS enjoyed normal distribution (p= .20, .15, p> .05). Furthermore, as evident in Table 15, there was no significant difference between the variances of the three groups of teachers on CAFS (p= .058, p> .05).

Table 15
Test of Homogeneity of Variances

	Statistic	df1	df2	Sig.
Degree	2.91	2	117	.058
Years of experience	2.46	2	117	.08
Brown-Forsythe	2.924	2	28.956	.07

Moreover, as is evident in Table 15, the one-way ANOVA results verified that B.A, M.A., and Ph.D. holding teachers' scores on CAFS were significantly different, $F(2, 117)$= 10.53, p= .00, p< .05. Further, in order to investigate the role of teaching experience in Iranian EFL teachers' attitudes towards fairness in classroom assessment, another one-way ANOVA was performed. Levene statistic (Table 15) showed that there was no significant difference among the variances of the three groups of less experienced, semi-experienced, and experienced EFL teachers on CAFS (p= .08, p> .05). In addition, KMS results indicated that less experienced, semi-experienced, and experienced EFL teachers' scores on CAFS were normally distributed (p= .20, .18, .20, p> .05). The ANOVA analysis results (Table 16) verified that there was a statistically significant difference in attitudes towards fairness in classroom assessment techniques among teachers with different years of experience ($F(2, 117)$= 4.72, p=.01, p< .05).

Finally, another one-way ANOVA was run to determine whether there was a statistically significant difference among Iranian EFL teachers' attitudes towards fairness in classroom assessment practices with respect to their teaching context. Brown-Forsythe test results (Table 15) revealed no significant difference between the variances of the three groups of EFL public schools, private language institutes, and university teachers (p=.07, p> .05).

KMS test results also showed that EFL public school, private language institute, and university teachers' scores on CAFS were all normally distributed (p= .20, .25, .21, p> .05). The ANOVA results confirmed that there were statistically significant differences among EFL public school, private language institute, and university teachers regarding their attitudes towards fairness in classroom assessment practices ($F(2, 117)$= 3.59, p= .03, p< .05).

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

99

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Table 16
Cumulative display of the ANOVA results concerning educational degree, experience, and teaching context

		Sum of Squares	Df	Mean Square	F	Sig.
	Between Groups	2.25	2	1.12	10.53	.00
Educational degree	Within Groups	12.51	117	.10		
	Total	14.77	119			
	Between Groups	1.10	2	.55	4.72	.01
Teaching experience	Within Groups	13.66	117	.11		
	Total	14.77	119			
	Between Groups	.85	2	.42	3.59	.030
Teaching context	Within Groups	13.91	117	.11		
	Total	14.77	119			

Discussion

This study focused on Iranian EFL teachers' attitudes toward fairness in classroom assessment practices. It uncovered that teachers were reasonably conscious of fairness issues in classroom assessment. They were mindful of the opportunity to learn, the opportunity to demonstrate learning, the no-harm principle, and the transparency components. However, they were not deservingly aware of the score pollution component. Regarding the opportunity to learn component, most teachers acknowledged the fairness of aligning activities included in a test with activities presented in class, postponing the final exam till the complete coverage of the material, and administering parallel forms of the test. Generally, EFL teachers admitted the significance of assessing learners on the material they had mastered. The congruence between assessments and syllabi is also suggested as a fair classroom assessment practice in studies like Rasooli et al. (2019), Torkey and Sayed Haider (2017), McMillan (2011), Russell and Airasian (2012), and Tierney (2014). However, some of the teachers deemed adding a few surprise items fair, which is consistent with the findings of Fan et al. (2020), Fan et al. (2017), and Liu et al. (2016), that reported that college students, professors, and pre-service teachers did not show agreement with experts and considered the inclusion of few surprise items in assessment ethical.

Regarding the opportunity to demonstrate the learning dimension, most teachers admitted the fairness of drawing on many assessment forms and including a variety of activities through which they could demonstrate their learning. Prior research has also verified that when learners are given multiple assessment opportunities, they feel treated fairly (Alm & Colnerud, 2015; Mauldin, 2009; Scott et al., 2014). To achieve fairness in assessment, the use of multiple assessments is strongly advised (Camilli, 2006), and has been welcomed by pre-service, in-service, and university instructors (Fan et al., 2019; Fan et al., 2020; Green et al., 2007; Liu et al. 2016). Buzzelli and Johnston (2002) contended that the fairness of the test cannot be morally acknowledged if there are insufficient opportunities for learning to be demonstrated.

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

100

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Descriptive statistic results for the transparency component disclosed that most of the teachers viewed illuminating the policy, procedures, and decisions, as well as why they are essential at the start of the course as fair practices. Their view is in line with that of Tierney (2014), who showed that teachers unanimously advocated communicating the learning requirements and assessment criteria to students beforehand. Nevertheless, although teachers reckoned specifying how a task will be graded as fair, they did not have the same attitude towards sharing rubrics with learners. The findings revealed that nearly half of the teachers were of the opinion that the details of the students' performance assessment rubric must be kept confidential and the rubrics should not be shared with learners, which contrasted with Fan et al. (2020) which reported that college students admitted that engaging students in the development of an assessment rubric was an ethical practice. The findings in this regard contradict assessment professionals' recommendations for clearly communicating the rubrics with students (DeLuca et al., 2016a; Kippers et al., 2018; McMillan, 2011). The results highlight the need for clear communication between teachers and learners to ensure that everyone is aware of what is happening in the classroom, what is expected from them, and how they will be assessed.

For the no-harm principle and constructive classroom environment component, respecting students' privacy was viewed by most of the teachers as quite fair. The significance of respect and trust in the educational process has already been emphasized (Cowie, 2005; Shepard, 2006), as research indicates that supportive teacher-student relationships are significantly related to positive student academic achievement (Kaufman, & Killen, 2022). "Do No Harm" is underlined when assessing students (Taylor & Nolen, 2005, p. 7) since poor assessment has an adverse effect on students. Studies on students' perceptions of fairness have also indicated that respect and trust in assessment exchanges are highly appreciated (Gordon & Fay, 2010; Wendorf & Alexander, 2005). Furthermore, most teachers agreed that slowing down the teaching speed to accommodate students' knowledge of the content and announcing the assessment time beforehand was fair. Furthermore, they considered spending time conferencing with students to discuss the strengths and weaknesses of their performance, as well as demonstrating concern for students' learning by carefully examining their complaints to be fair assessment techniques. Empirical research has shown that taking students' voices into account increases their opportunity to participate in the assessment process actively (Flores et al., 2015; Murillo & Hidalgo, 2017), and lack of voice in assessment processes has a detrimental influence (Murillo & Hidalgo, 2017).

For avoiding score pollution, teachers appeared to have a lot of misconceptions concerning the fairness of score assignments. This piece of finding is consistent with those of Torkey and Sayed Haider (2017), Green et al. (2007), and Pope et al. (2009), reporting that score pollution made up the majority of ethical dilemmas teachers experienced. The results revealed that Iranian EFL teachers were inclined to change learners' scores based on non-academic performances, such as their late delivery of assignments, disruptive behavior, and lack of regular class attendance. These findings are in line with those of Alm and Colnerud (2015), Duncan and Noonan (2007); Liu et al.

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

101

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

(2016), Murillo and Hidalgo (2017), Resh (2009), Scott et al. (2014), Tierney (2015), and Torkey and Sayed Haider (2017) on teachers' orientation to consider non-mastery factors in grading. Most assessment professionals believe that scores should solely represent learners' mastery of instructional goals (Brookhart, 2004; Smith et al., 2001); however, Iranian EFL teachers presumed considering students' non-academic performances, such as belated submission of assignments and disruptive behavior to be fair practices. The tendency of Iranian EFL teachers to take non-mastery factors into account while grading may be connected to their educational background. For example, it is common in Iranian education for instructors to include non-mastery factors such as class participation as part of students' final grades.

The findings also indicated that teachers welcomed providing extra credit opportunities to all learners, excluding the more knowledgeable ones, which is consistent with Tierney (2014) who reported that teachers considered increasing grades for at-risk students as a fair practice. Such persisted confusion among the teachers indicates that teachers lacked a clear mind about the score pollution aspect of assessment fairness, a point which reflects Maclellan (2004) where he maintains that in spite of the significance of grades and the intricacies of the grading procedure, it is an issue that has been overlooked in teacher education, leaving newly certified instructors typically ill-equipped for their jobs.

The findings of the study concerning the second research question verified that gender, educational degree, years of teaching experience, and teaching context made statistically significant differences among Iranian EFL teachers' attitudes toward fairness in classroom assessment practices. Accordingly, female teachers were more conscious about classroom assessment fairness than male teachers. This finding corroborates Fan et al. (2020), who found that female professors obtained a slightly higher score on ethical dilemmas in assessment.

The results also showed that higher education levels might be associated with more awareness of classroom assessment fairness. This finding is partially consistent with those of Hamzelou et al. (2022), Soodmand Afshar and Ranjbar (2021), Soodmand Afshar et al. (2018), and DeLuca et al. (2013), which highlighted the significant role of educational degree in teachers' assessment literacy. Furthermore, years of teaching experience contributed to teachers' awareness of fairness in classroom assessment, and experienced teachers were better at discerning fair classroom assessment practices. This finding is also consistent with those of DeLuca et al. (2016), Homayounzadeh and Razmjoo (2021), and Zolfaghari and Ashraf (2015), who found that more experienced teachers reported higher skill levels in assessment practices than less experienced teachers. However, Soodmand Afshar et al. (2018) reported that teachers' years of teaching experience had no bearing on how assessment literate they were. Lastly, the results showed statistically significant differences among public schools, private language institutes, and university teachers regarding their attitudes toward fairness in classroom assessment practices. This piece finding highlights the impact of context on assessment practices assumptions (Brown et al., 2011; James & Pedder, 2006; Rubie-Davies et al., 2012; Vandeyar & Killen, 2007).

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**102** Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Conclusion and Implications

The study's findings indicated that Iranian EFL teachers were rather concerned with fairness in classroom assessment practices. They appeared to be conscious of the opportunity to learn, the opportunity to demonstrate learning, the no-harm principle, and the constructive learning environment, as well as the transparency components. However, they seemed oblivious to the avoiding score pollution component. Such findings contribute to the existing literature and further document EFL teachers' attitudes toward fairness in classroom assessment. As for the theoretical implications of the study findings, an improved understanding of what fair assessment is helps theoreticians in defining the basics of fair assessment practice and demarcating fairness from unfairness in theoretical terms. On the other hand, as it is the instructors who are the ones to establish the assessment culture in the classroom, their attitudes toward fair and equitable assessments exert influence on the psychosocial and learning outcomes of learners (Elwood & Murphy, 2015), from a pedagogical perspective, the findings underscore the necessity of the teachers prompt attention to the factors that may pollute their classroom assessment practices.

Furthermore, from a teacher education perspective, the findings underscore the necessity and significance of professional development programs for in-service teachers in which the theoretical principles and practical aspects of classroom assessment fairness are put into focus.

In addition, as teacher education programs inadequately prepare teachers for well-grounded classroom assessment (McGee & Colby, 2014), and Iranian teachers receive no specific instruction about fair assessment during preservice training (Rasooli et al., 2022), PD programs, as well as further studies on innovative continuing professional development program for boosting language teachers' knowledge of fair assessment, are encouraged. Indeed, adding the discussion about what constitutes a fair evaluation and its implications in teachers' initial and ongoing training could help them become more conscious of their assessment choices and work toward a fairer assessment.

Finally, as for the limitations, this study was limited to Iranian EFL teachers, excluding other subject matter teachers, test takers, and even their parents' perspectives and attitudes towards classroom assessment fairness. On this basis, the researchers are suggested to explore the attitudes of various stakeholders including the teacher trainers, educationalists, educational policy makers, the learners, and the learners' parents as well.

References

Aitken, N. (2012). Student voice in fair assessment practice. In C. F. Webber & J. Lupart (Eds.), *Leading student assessment,* (pp. 175-200). Springer.

Alm, F. & Colnerud, G. (2015). Teachers' experiences of unfair grading. *Educational Assessment, 20*(2), 132–150. http://dx.doi.org/10.1080/10627197.2015.1028620.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**103** Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Association, American Psychological Association, & National Council on Measurement in Education.

Azizi, Z. (2022). Fairness in assessment practices in online education: Iranian University English teachers' perceptions. *Language Testing in Asia, 12*(14), 1-17. https://doi.org/10.1186/s40468-022-00164-7

Baniasadi, A., Salehi, K., Khodaie, E., Bagheri Noaparast, Kh. & Balal Izanloo, B. (2023). Fairness in classroom assessment: A systematic review. *Asia-Pacific Edu Res, 32*(1), 91–109. https://doi.org/10.1007/s40299-021-00636-z

Berg, B. L. (2001). *Qualitative research methods for the social sciences*. Allyn and Bacon.

Berti, C., Molinari, L., & Speltini, G. (2010). Classroom justice and psychological engagement: Students' and teachers' representations. *Social Psychology of Education, 13*(4), 541–556. http://dx.doi.org/10.1007/s11218-010-9128-9

Brookhart, S. (2004). *Grading*. Pearson.

Brown, G. T. L., Hui, S. K., Flora, W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50*(5), 307–320. http://dx.doi.org/10.1016/j.ijer.2011.10.003

Buzzelli, C. A. & Johnston, B. (2002). *The moral dimensions of teaching: Language, power, and culture in classroom interaction*. Routledge Falmer.

Camilli, G. (2006). Test fairness. In R. Brennan (Ed.). *Educational measurement* (pp. 221-256) (4th Ed.). American Council on Education and Praeger.

Catell, R. B. (1966). The scree test for number of factors. *Multivariate Behavioral Research, 1*(2), 245-276. http://dx.doi.org/10.1207/s15327906mbr0102_10

Colnerud, G. (1997). Ethical conflicts in teaching. *Teaching and Teacher Education, 13*(6), 627–635. https://doi.org/10.1016/S0742-051X(97)80005-4

Colnerud, G. (2015). Moral stress in teaching practice. *Teachers and Teaching: Theory and Practice, 21*(3), 346–360. https://doi.org/10.1080/13540602.2014.953820

Cowie, B. (2005). Student commentary on classroom assessment in science: A sociocultural interpretation. *International Journal of Science Education, 27*(2), 199-214. https://doi.org/10.1080/0950069042000276721

Cowie, B. (2015). Equity, ethics and engagement: Principles for quality formative assessment in primary science classrooms. In C. Milne, K. Tobin, & D. DeGennaro (Eds.). *Sociocultural studies and implications for science education* (pp. 117–133). Springer.

Chory-Assad, R. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly, 50*(1), 58-77. http://dx.doi.org/10.1080/01463370209385646

Chory, R. M. (2007). Enhancing student perceptions of fairness: The relationship between instructor credibility and classroom justice. *Communication Education, 56*(1), 89-105. http://dx.doi.org/10.1080/03634520600994300

Chory-Assad, R., & Paulsel, M. (2004). Antisocial classroom communication: Instructor influence and interactional justice as predictors of student aggression. *Communication Quarterly, 52*, 98–114. https://doi.org/10.1080/01463 37040 9370184

Darabi Bazvand, A. & Rasooli, A. (2022). Students' experiences of fairness in summative assessment: A study in a higher education context. *Studies in Educational Evaluation, 72*, 101118. https://doi.org/10.1016/j.stueduc.2021.101118

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

104

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Teacher Education Yearbook XXI: A Special Issue of Action in Teacher Education, 34*(5-6), 576–591. https://doi.org/10.1080/01626620.2012.730347

DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for pre-service assessment Education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator, 48*(2), 128-142. https://doi.org/10.1080/08878730.2012.760024

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

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251–272. https://link.springer.com/article/10.1007/s11092-015-9233-6

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D. & Luhanga, U. (2016). Teachers' approaches to classroom assessment: a large-scale survey. *Assessment in Education: Principles, Policy & Practice*, *25*(4), 355-375. https://doi.org/10.1080/0969594X.2016.1244514

Dorans, N. J., & Cook, L. L. (2016). *Fairness in educational assessment and measurement*. Routledge.

Dornyei, Z. (2003). *Questionnaires in second language research: Construction, administration and processing*. Lawrence Erlbaum Associates.

Downing, S. M. & Haladyna, T. M. (2004). Validity threats: Overcoming interference with proposed interpretations of assessment data. *Medical Education, 38*(3), 327-333. DOI: 10.1046/j.1365-2923.2004.01777.x

Duncan, C. R., & Noonan. B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research, 53*(1), 1–21. https://doi.org/10.11575/ajer.v53i1.55195

Fan, X., Johnson, R. & Liu, X. (2017). Chinese university professors' perceptions about ethical issues in classroom assessment practices. *New Waves Educational Research & Development, 20*(2), 1–19. https://files.eric.ed.gov/fulltext/EJ1211389.pdf

Fan, X., Johnson, R., Liu, X., 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), 309–32. https://link.springer.com/article/10.1007/s11516-019-0015-7

Fan, X., Liu, X., & Johnson, R. (2020). A mixed method study of ethical issues in classroom assessment in Chinese higher education. *Asia Pacific Education Review, 21*, 183-195. https://doi.org/10.1007/s12564-019-09623-y

Fan, X., Johnson, R., Liu, X., & Gao, R. (2020). College students' views of ethical issues in classroom assessment in Chinese higher education. *Studies in Higher Education*, *47*(1), 131-145. https://doi.org/10.1080/03075079.2020.1732908

Flores, M. A., Veiga Simão, A. M., Barros, A., & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: A study in higher education. *Studies in Higher Education, 40*(9), 1523–1534. http://dx.doi.org/10.1080/03075079.2014.881348

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

105

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching, 58*(3), 93–98. https://www.jstor.org/stable/41305091

Green, S. K. & Johnson, R. L. (2010). *Assessment is essential*. McGraw-Hill Higher Education.

Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: issues and attitudes. *Teaching and Teacher Education, 23*(7), 999-1011. https://doi.org/10.1016/j.tate.2006.04.042

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Prentice hall.

Hair, J. F. J., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publication.

Hamzelou, Z., Alavi, S. M., & Nemati, M. (2022). Integration of learning-oriented assessment into writing instruction: A case of IELTS writing preparation class. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *41*(3), 57-87. doi: 10.22099/tesl.2022.35693.2761

Herman, J., & Cook, L. (2019). Fairness in classroom assessment. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 243-264). Routledge.

Hidri, S. (2016). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics, 1*(1), 19–43.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *SEM, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Homayounzadeh, Z., & Razmjoo, S. A. (2021). Examining 'assessment literacy in practice' in an Iranian context: Does it differ for instructors and learners? *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *40*(2), 1-45. doi: 10.22099/jtls.2021.40269.2978

Ishak, Z., & Fin, L. (2013). Truants' and teachers' behaviors in the classroom. *Procedia-Social and Behavioral Sciences, 103*, 1228–1237. https://doi.org/10.1016/j.sbspro.2013.10.451

James, M. & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal, 17*(2), 109–138. https://doi.org/10.1080/09585170600792712

Kaufman, E., & Killen, M. (2022). Children's perspectives on fairness and inclusivity in the classroom. *The Spanish Journal of Psychology, 25*, 28. 1-11. DOI:10.1017/SJP.2022.24

Kippers, W. B., Wolterinck, Ch. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education, 75*, 199-213. https://doi.org/10.1016/j.tate.2018.06.015

Kunnan, A. J. (2018). *Evaluating language assessments*. Routledge.

Lantolf, J., & Poehner, M. (2013). The unfairness of equal treatment: Objectivity in L2 testing and dynamic assessment. *Educational Research and Evaluation, 19*, 141–157. http://dx.doi.org/10.1080/13803611.2013.76761

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills)	**106**
	42(2), Spring 2023, pp. 81-110	Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Liu, J., Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers' perceptions about ethical issues in classroom assessment. *Studies in Educational Evaluation, 48*, 57–66. http://dx.doi.org/10.1016/j.stueduc.2016.01.002

Lupart, J., & Webber, C. F. (1996). Schools in transition: Issues and prospects. In J. Lupart, A. Mckeough, & C. Yewchuk (Eds.), *Schools in transition* (pp. 3–39). Nelson Canada.

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education, 20*(5), 523-535. https://doi.org/10.1016/j.tate.2004.04.008

Mauldin, R. K. (2009). Gendered perceptions of learning and fairness when choice between exam types is offered. *Active Learning in Higher Education, 10*(3), 253-264. http://dx.doi.org/10.1177/1469787409343191

McGee, J. & Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education, 36*(5-6), 522-532. https://doi.org/10.1080/01626620.2014.977753

McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed.). Pearson Education Inc.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge.

Molinari, L., Speltini, G., & Passini, S. (2013). Do perceptions of being treated fairly increase students' outcomes? Teacher–student interactions and classroom justice in Italian adolescents. *Educational Research and Evaluation, 19*(1), 58–76. http://dx.doi.org/10.1080/13803611.2012.748254

Murdock, T., Miller, A., & Goetzinger, A. (2007). Effects of classroom context on university students' judgments about cheating: Mediating and moderating processes. *Social Psychology of Education, 10*, 141–169. https://doi.org/10.1007/s11218-007-9015-1

Murillo, F. J., & Hidalgo, N. (2017). Students' conceptions about a fair assessment of their learning. *Studies in Educational Evaluation, 53*, 10-16. http://dx.doi.org/101016/j.stueduc.2017.01.001

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

Nisbet, I. (2017). Fairness takes center stage. *Assessment in Education: Principles, Policy & Practice*, *26*(1), 1–7. https://doi.org/10.1080/0969594X.2017.1358151

Nisbet, I. & Shaw, S. D. (2019). Fair assessment viewed through the lenses of measurement theory. *Assessment in Education: Principles, Policy & Practice*, *26*(5), 612-629. https://doi.org/10.1080/0969594X.2019.1586643

Pallant, J. (2013). *A step by step guide to data analysis using IBM SPSS*. Mc Graw Hill.

Paulsel, M., Chory-Assad, R., & Dunleavy, K. N. (2005). The relationship between student perceptions of instructor power and classroom justice. *Communication Research Reports, 22*(3), 207–215. https://doi.org/10.1080/00036810500207030

Pepper, M. B., & Pathak, S. (2008). Classroom contribution: What do students perceive as fair assessment? *Journal of Education for Business, 83*(6), 360–368. http://dx.doi.org/10.3200/JOEB.83.6.360-368

Peters, R., Kruse, J., Buckmiller, T., & Townsley, M. (2017). It's just not fair! making sense of secondary students' resistance to a standards-based grading. *American Secondary Education, 45*(3), 9–27.

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 42(2), Spring 2023, pp. 81-110	**107** **Somaye Tofighi**

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Pettifor, J. L. & Saklofske, D. H. (2012). Fair and ethical student assessment practices. In Webber, C. F., Lupart, J. (Eds.), *Leading student assessment* (pp. 87-106). Springer.

Pope, N., Green, S., Johnson, R., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education, 25*(5), 778-782.

Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation, 56*, 164-181. https://doi.org/10.1016/j.stueduc.2017.12.008

Rasooli, A., DeLuca, Ch., Resegh, A., & Fathi, S. (2019). Students' critical incidents of fairness in classroom assessment: an empirical study. *Social Psychology of Education, 22*, 701-722. https://doi.org/10.1007/s11218-019-09491-9

Rasooli, A., Rasegh, A., Zandi, H. & Firoozi, T. (2022). Teachers' conceptions of fairness in classroom assessment: An empirical study. *Journal of Teacher Education, 74*(3), 260-273. https://doi.org/10.1177/00224871221130742

Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education, 12*(3), 315-25. https://link.springer.com/article/10.1007%2Fs11218-008-9073-z

Rezai, A., Alibakhshi, G., Farokhipour, S., & Miri, M. (2021). A phenomenographic study on language assessment literacy: Hearing from Iranian university teachers. *Language Testing in Asia, 11*, 1–18. 10.1186/s40468-021-00142-5

Rubie-Davies, C. M., Flint, A. Y., & McDonald, L. G. (2012). Teacher beliefs, teacher characteristics, and school contextual factors: What are the relationships? *British Journal of Educational Psychology, 82*(2), 270–288. https://doi.org/10.1111/j.2044-8279.2011.02025.x

Russell, M. K. & Airasian, P. W. (2012). *CA: Concepts and applications* (7th ed.). McGraw-Hill Higher Education.

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

Shah Ahmadi, M., & Ketabi, S. (2019). Features of language assessment literacy in Iranian English language teachers' perceptions and practices. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *38*(1), 191-223. doi: 10.22099/jtls.2020.34843.2739

Shepard, L. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (pp. 623–646). National Council on Measurement in Education, American Council on Education/Praeger Publishers.

Smith, J. K., Smith, L. F., & DeLisi, R. (2001). *Natural classroom assessment: Designing seamless instruction and assessment*. Corwin Press.

Soodmand Afshar, H. & Ranjbar, N. (2021). EAP teachers' assessment literacy: From theory to practice. *Studies in Educational Evaluation, 70*(4), 1-14 https://doi.org/10.1016/j.stueduc.2021.101042

Soodmand Afshar, H., Tofighi, S., Assodeh, M. & Ranjbar, N. (2018). The impact of alternative assessment knowledge, teaching experience, gender, and academic degree on EAP teachers' assessment literacy. *Iranian Journal of English for Academic Purposes, 7*(2), 1-15. 20.1001.1.24763187.2018.7.2.1.6
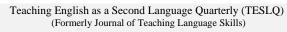
Taylor, K. & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Pearson Education.

Tierney, R. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.). *Sage handbook of research on classroom assessment* (pp. 125-144). Thousand Oaks, CA: SAGE Publications.

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

Tierney, R. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters, 8*, 5–30. http://dx.doi.org/10.18296/am.0002

Tierney, R. (2016). Fairness in educational assessment. In M. A. Peters (Ed.). *Encyclopedia of educational philosophy and theory* (pp. 1-6). Springer Singapore.

Torkey, S. A. & Sayed Haider, N. A. (2017). Ethicality in EFL classroom assessment: Bridging the gap between theory and practice. In S. Hidri & C. Coombe (Eds.), *Evaluation in foreign language education in the Middle East and North Africa* (pp. 59-85). Springer.

Vandeyar, S. Y. & Killen, R. (2007). Educators' conceptions and practice of classroom assessment in post-apartheid South Africa. *South African Journal of Education, 27*(1), 101–115. https://www.ajol.info/index.php/saje/article/view/25101

Warnick, B. R. & Silverman, S. K. (2011). A framework for professional ethics courses in teacher education. *Journal of Teacher Education, 62*(3), 273-285. https://doi.org/10.1177/0022487110398002

Weber, R. P. (1990). *Basic content analysis*. Sage Publications.

Wendorf, C. A., & Alexander, S. (2005). The influence of individual- and class-level fairness-related perceptions on student satisfaction. *Contemporary Educational Psychology, 30*(2), 190–206. https://doi.org/10.1016/j.cedpsych.2004.07.003

Xu, Y., & Brown, G. (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

Zhang, Z., & Burry-Stock, J. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323-342. https://doi.org/10.1207/S15324818AME1604_4

Zolfaghari, S., & Ashraf, H. (2015). The relationship between EFL teachers' assessment literacy, their teaching experience, and their age: A case of Iranian EFL teachers. *Theory and Practice in Language Studies, 5*(12), 2550-2556.

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(2), Spring 2023, pp. 81-110

109

Somaye Tofighi

FAIRNESS IN CLASSROOM LANGUAGE ASSESSMENT FROM EFL

Appendix A
Classroom Assessment Fairness Scale (CAFS)

Items	1	2	3	4	5
1. I make sure that the activities included in a test are quite similar to activities presented in class. (Opportunity to Learn +)					
2. I do not administer the final exam until I make sure that I have covered the material comprehensively. (Opportunity to Learn +)					
3. For the final exam, I always use a few surprise items that are not presented in class (Opportunity to Learn -)					
4. To prepare students for an upcoming test, I administer a parallel form of the test. (Opportunity to Learn +).					
5. To assess students' achievement, I use methods that students have regularly encountered in class. (Opportunity to Learn +)					
6. I assess students' learning by drawing on many types of assessment forms such as self and peer assessment, classroom discussion, presentations, doing projects, portfolios (Opportunity to demonstrate learning +)					
7. I use observation as the sole method to assess what students have learned (Opportunity to demonstrate learning -)					
8. I assess oral proficiency through visual or audio recording of oral performances of students engaged in different activities such as role play, interview, discussion, and comparing them with recordings of subsequent performances to document improvements. (Opportunity to demonstrate learning +)					
9. I give students a variety of activities through which they would be able to demonstrate their learning (Opportunity to demonstrate Learning +)					
10. I state how I will grade a task when I assign it. (transparency +)					
11. At the beginning of the semester, I share with students the rubrics for each task to guide their completion of the tasks (transparency +)					
12. I keep the details of the students' performance assessment rubric confidential (transparency -)					
13. I clarify my policy, procedures, and decisions, as well as why they are important, at the start of the course. (transparency +)					
14. I clarify why I am concerned about the class attendance policy and why I believe it is important because attendance is linked to improved learning and better grades. (transparency +)					
15. I spend time conferencing with students to explain their performance's strengths and weaknesses. (No Harm Principle and Constructive Classroom Environment +)					
16. Based on the students' understanding of the material, I would slow down my teaching pace to adapt to students' needs. (No Harm Principle and Constructive Classroom Environment +)					
17. I share with students the rubrics for each task. I take into account students' feedback on the rubrics for each assignment					

Items	1	2	3	4	5
and make adjustments to the rubrics as required. (No Harm Principle and Constructive Classroom Environment +)					
18. I do not grade all assignments. Instead, I have students rate each other's assignments and then share the results in groups. Teamwork, in my view, would aid students' learning. (No Harm Principle and Constructive Classroom Environment +)					
19. I employ peer assessment as a part of the final exam (No Harm Principle and Constructive Classroom Environment +)					
20. I show my concern for students' learning by carefully considering student complaints and taking remedial action when necessary. (No Harm Principle and Constructive Classroom Environment +)					
21. I respect the privacy of my students; I do not require them to reveal highly personal information in a class discussion. (No Harm Principle and Constructive Classroom Environment +)					
22. When facing a student or students cheating on a test, I either score their performance based on other modules of the test or give an alternate version of the same test. (No Harm Principle and Constructive Classroom Environment +)					
23. I announce the test/ assessment time days before I administer it so that the student can have sufficient preparation time (No Harm Principle and Constructive Classroom Environment +)					
24. If someone hands in their assignments late, I will give him/her a lower score. (score pollution -)					
25. I do not modify learners' score based on other teachers' perceptions of them. (score pollution +)					
26. If a learner shows a disruptive behavior, I change his/her score. (score pollution -)					
27. My perception of learners is influenced by their naughty behavior. (score pollution -)					
28. In giving scores, I do not take into consideration the degree of neatness with which learners do their tasks. (score pollution +)					
29. I allocate a small part of the total score to learners' class attendance. (score pollution -)					
30. I dedicate extra credit for volunteering in classroom activities. (score pollution -)					
31. I am not influenced by parent pressure to alter standards or bend the rules. (score pollution +)					
32. For a group project, I base each student's score on the group's product and ignore group members' individual abilities. (score pollution -)					
33. To encourage a dynamic and active atmosphere, I count class participation as some part of the final score. (score pollution -)					
34. To minimize guessing, I deduct more points for a wrong answer than for leaving the answer blank. (score pollution -)					
35. I offer extra credit opportunities to all leaners except the more knowledgeable ones (score pollution -)					