Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
45(1) 2026, pp. 81-108
https://doi.org/10.22099/tesl.2025.52797.3395
Online ISSN: 2981-1546
Research Paper

Test Fairness in Online Assessment: Insights from Iranian EFL Teachers' Perspective

Mohammad Ahmadi Safa [*1] **Fatemeh Ansari** [2]

Abstract

The widespread integration of technology into education has led to significant reliance on online assessment. Despite its prevalence, concerns about the fairness of online assessment persist. While this issue has attracted considerable attention, empirical investigations, particularly those exploring teachers' perceptions of fairness in online assessment, remain scarce. Additionally, existing frameworks for assessing fairness in an online context seem inadequate. To address these concerns, the current study investigated Iranian English as a Foreign Language (EFL) teachers' perceptions of fairness in online assessment. Moreover, the study aimed to develop a valid scale to measure the fairness of assessment practices in the online context. Employing a mixed-methods approach, the study conducted a structured interview with 20 EFL teachers, and the resulting data were transcribed and analyzed qualitatively. Ten major codes or themes derived from the data were access, knowledge, personal factors, clear criteria, item quality, safety, policies, teaching quality, assessment type, and administration. The codes were subsequently classified into three overarching categories: student level, teacher level, and organizational level. Then, based on the interview findings and the existing literature, an online assessment fairness questionnaire was developed and administered to a sample of 100 EFL teachers, whose results verified five core factors of online assessment fairness. The study's findings offer a novel understanding of teachers' perspectives on the fairness of online assessment and the factors that contribute to its fairness or unfairness.

Keywords: Online assessment, Fairness, EFL teachers, Iran

Technological advancements have brought significant changes to education, including the emergence of online teaching. In this context, assessment is considered a crucial element that must encompass various aspects to ensure academic integrity, including the principle of fairness.

While according to the Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999), fairness is defined as absence of bias, equitable treatment of all test takers in

[*] Review History:
Received: 23/03/2025 Revised: 09/07/2025 Accepted: 30/08/2025

1. Professor, Department of English, Humanities faculty, Bu-Ali Sina University, Hamedan, Iran (Corresponding Author)
m.ahmadisafa@basu.ac.ir
2. Ph.D. Candidate, Department of English, Humanities faculty, Bu-Ali Sina University, Hamedan, Iran,
f.ansari.1364@gmail.com

How to cite this article:
Ahmadi Safa, M. and Ansari, F. (2026). Test Fairness in Online Assessment: Insights from Iranian EFL Teachers Perspective. *Teaching English as a Second Language Quarterly*, *45*(1), 81-108.
https://doi.org/10.22099/tesl.2025.52797.3395

the testing process, and equity in opportunity to learn the material in an achievement test., others have defined it as an overarching framework encompassing validity (e.g., Kunnan, 2000, 2004), or have taken it as a part of validity framework (e.g., Xi, 2010) or even for some authors (e.g., Davies, 2010) attempts for fairness is futile. Such opposing positions reflect the lack of a universally accepted definition of the concept, given that, as Hamp Lyons (2001) maintains, language testers have no authority to decide what is fair and what is not for individuals. Nevertheless, they are required to apply all available means to make tests as fair as possible.

Fairness in online assessment is becoming a vital factor, given the vast impacts of assessment beyond its immediate contexts, which also influences people's social lives. In this regard, Kunnan (2008) proposes, in his Test Context Framework (TCF), that to thoroughly investigate tests and their consequences, it is essential to examine test contexts. Despite the importance of this issue, there has been a dearth of research on fairness in online assessment, particularly in Iran. Moreover, most of the frameworks proposed for this purpose have been deemed inadequate (Davies, 2010; Xi, 2010). On this basis, the present study aimed to develop a valid scale to measure the fairness of assessment practices in the online context and, accordingly, to explore teachers' perceptions of fairness in online assessment, adopting a mixed-methods design.

Literature Review

Depending on the adopted theoretical framework, test fairness has been defined differently. A group of researchers has interpreted fairness as a broad, all-inclusive construct encompassing validity, the absence of bias, access, administration, and social consequences (Kunnan, 2000, 2004). Others have subsumed fairness under validity (Xi, 2010), while still others have argued that fairness is an independent facet of test quality that does not have a consistent relationship with validity (Educational Testing Service, 2002).

Kunnan (2000) proposed a framework for fairness that comprised validity, access, and justice. Later, Kunnan (2004) revised his framework by introducing two major principles: justice and beneficence. Justice was restricted to construct irrelevant factors that influence test administration and interpretation. By beneficence, he stated that a test must have positive effects on society. Overall, for Kunnan, fairness includes validity, absence of bias, access, administration, and social outcomes, each with its own components. Moreover, according to Kunnan (2008), a framework for fairness should be presented within a wider context encompassing" political and economic; educational, social, and cultural; technological and infrastructure; and the legal and ethical contexts" (p.240).

An alternative conceptualization of fairness is presented by Xi (2010), who describes fairness as a requirement for validity that should be comparable across all stages of assessment. This implies equal testing methods, procedures, content, administration, scoring, and interpretation. However, requiring learners with diverse characteristics to meet the same testing requirements may not be fair. In line with this criticism, Camilli (2013) proposed individual

fairness in assessment, which suggests that students should be treated comparably, but the term 'comparable' does not necessarily mean 'equal'. For instance, students with dyslexia may experience difficulty taking tests, making it unfair to compare their scores with those of normal students. In such cases, test givers should accommodate test scores to increase the fairness of their assessments. This is consistent with the Kantian second principle (Rawls, 2001), which holds that, in cases of inequality, the least well-endowed citizens should be offered the greatest benefits.

On the other hand, the complexities involved in conceptualizing test fairness have even persuaded some experts to abandon the process altogether. Davies (2010), for instance, believes the concept is too complex to define. Hamp-Lyons (2001) also argued that devising a comprehensive model of test fairness seems unlikely. In a less categorical stance, Cole and Zieky (2001) maintained that fairness is not a dichotomous concept but rather a matter of degree. McNamara and Roever (2006) also asserted that while fairness guidelines cannot fully satisfy the needs of all stakeholders, test makers still use them to raise the quality of their assessments.

As Kunnan (2010) argued, any fairness framework must take into account the broader contexts in which it is used. On this basis, there is a concrete call to redefine test fairness in the context of online assessment. Despite its importance, there is a noticeable paucity of research investigating fairness in online assessment contexts, which might be partially due to problems inherent to online assessment contexts, including technological challenges, infrastructural barriers, difficulties in scoring open-ended questions, and security concerns (García-Peñalvo et al., 2021).

García-Peñalvo et al. (2021) studied the challenges of shifting to online assessment encountered by higher education during the COVID-19 pandemic. Their research involved surveys and interviews with students, faculty, and administrators, complemented by a systematic review of the existing literature and case studies of online assessment implemented across various universities. The findings revealed both obstacles and opportunities in online assessment, underscoring the need for flexibility, support, and continuous improvement in these practices. They highlighted the critical need to equip both students and teachers with the essential skills and support to use assessment tools. A paramount concern identified was maintaining academic integrity in online assessment, and to address this, the authors proposed a diverse assessment approach combining open-book examinations, project-based assessments, and oral examinations.

Reflecting similar challenges, studies within the educational context in Iran have also echoed these concerns. For instance, Ghanbari and Nowroozi (2022) investigated how a group of 20 Iranian TEFL teachers at Persian Gulf University experienced the challenges and adaptations of transitioning to online assessment during the COVID-19 pandemic. The researchers conducted semi-structured retrospective interviews with teachers at various points

during their online course. In conclusion, they called for providing technological, pedagogical, and administrative support to optimize post-pandemic online assessment.

In another study, Momeni (2022) explored Iranian EFL teachers' perceptions of online assessment during the COVID-19 lockdown. Using a mixed-methods approach, the research identified key challenges teachers faced during the shift to online assessment. A primary concern was the high risk of academic dishonesty, including cheating and plagiarism, in the online environment. This challenge was compounded by insufficient technological infrastructure and unreliable internet connectivity. To enhance assessment fairness, the author emphasized the need for robust anti-cheating measures, including proctoring tools or well-designed open-book assessments. The findings also recommended flexible assessment formats, transparent grading, timely feedback, clear rubrics, and extending deadlines.

Among the few studies that have recently addressed online assessment fairness, Azizi (2022) explored Iranian university English teachers' perceptions of fairness in online assessment. The findings underscored three main types of justice that must be guaranteed in online assessment contexts. First, distributive justice involves considering equality and equity, as well as matching assessment practices with students' needs. Second, procedural justice, which entails recognizing students' voices, ensuring both consistency and flexibility, and maintaining transparency in assessment procedures; and last, interactional justice, which encompasses interpersonal justice and informational justice. She concluded that, as in traditional classroom settings, ensuring fairness in online assessment must be a priority.

Drawing upon the reviewed literature- and particularly in light of scarce empirical inquiry into fairness in online assessment- the current study sought to explore the complexities of fairness of online assessment through foregrounding the perspectives of Iranian EFL teachers. While previous studies in online assessment have primarily focused on challenges such as technological constraints and academic dishonesty, they have often treated fairness as a secondary issue at most. In addition, although scholars have discussed fairness from varied perspectives, there remains a dearth of a validated instrument developed to measure fairness in online assessment practices. Addressing these gaps, the present study aimed to explore how Iranian EFL teachers conceptualize fairness and unfairness in online assessment and to develop and validate a scale to evaluate fairness in the online context. For these stated purposes, the following research questions were posed:

1. What are Iranian EFL teachers' perceptions of un/fairness in online assessment?
2. What are the factors of a valid measure of online assessment fairness?

Methodology

Participants

In the qualitative phase, the project used a convenient sample of 20 EFL school, institute, and university teachers holding BA, MA, and PhD degrees in TEFL, English literature, and translation, with experience in online teaching and assessment. 15 (75 percent) of the respondents were males, and 5 (25 percent) were females. They were aged between 25 and 37. In the quantitative phase, the pilot sample consisted of 100 conveniently available online English teachers teaching in schools, institutes, and universities who held BA, MA, and PhD degrees in TEFL, English literature, and translation. 27% were males and 73% were females. All participants were aged 22-40 years. Announcements for both phases of the study were issued in the Iranian EFL WhatsApp and Telegram groups. Upon receiving the agreement, the qualitative phase items and the questionnaire were sent to the volunteers. Demographic information for participants in both phases is presented in Table 1.

Table 1

Demographic information of the participants

		Gender		Years of teaching experience			Educational Major			Teaching context			
		Male	Female	1-5	6-10	More than 10	TEFL	English Literature	English translation	Non-English	School	Language Institute	University
Qualitative Phase	N	5	15	6	9	5	13	3	3	0	3	13	4
	%												
Quantitative Phase	N	27	73	27	28	45	59	19	8	14	6	81	13

Instruments

Interview. A structured interview consisting of 21 questions was designed based on the existing literature to explore participants' perspectives on fairness in online assessment (see Appendix A). Dornyei (2007) asserts that interviewing is an acceptable method of data collection as most individuals feel comfortable participating in interviews. It can also be employed as a valuable tool for deeply exploring a wide range of topics across diverse contexts. The questions were reviewed by three TEFL experts and revised based on their comments. The revised version was administered to participants via social media platforms such as WhatsApp and Telegram. To increase the validity of responses and participants' comfort and willingness to participate, the interview questions were designed in Farsi, regardless of participants' English proficiency levels. Furthermore, the participants were required to voice-record their answers and comments to each question item whenever they felt comfortable, for a limited period.

Online Assessment Fairness Scale. Based on the related literature and interview findings, a 55-item questionnaire was developed to investigate the topic further quantitatively. Dornyei (2007) proposes that questionnaires offer a structured and systematic way of data collection applicable across a variety of contexts and topics. To address the need for item validity, the

questionnaire was subjected to the expert judgement of three TEFL university professors with expertise in assessment and evaluation. Following the revision of the scale based on the experts' comments, the questionnaire was pilot-tested with 100 participants, and the data were subjected to exploratory and confirmatory factor analysis. The resulting indices finally confirmed the scale's acceptable validity.

Data Collection

A mixed-methods research approach was employed, comprising both qualitative and quantitative phases. In the qualitative phase, the interview questions were developed after conducting a thorough review of the literature. To ensure content validity and item clarity, the items were reviewed by three TEFL university professors with expertise in assessment and subsequently revised based on their comments. Prior to commencing the study, participation requests were sent to potential interview candidates via WhatsApp and Telegram. They were briefed on the study's purpose and assured that their participation was voluntary. Upon receiving their consent, the questions were sent to them via WhatsApp and Telegram. To safeguard the ethicality of the process, the participants were assured of data anonymity, results confidentiality, and the exclusive use of the data for research purposes only. Participants were instructed to respond in Farsi via voice messages. They were requested to elaborate on their understanding of fairness in online assessment. Any unclear responses were followed up with requests for clarification. The responses were first transcribed, and then, following Grounded Theory (GT) rubrics, were analyzed and codified into open, axial, and selective coding. Dornyei (2007) proposes that grounded theory provides a robust tool for a thorough analysis of a phenomenon and generates theoretical insights, particularly in areas where existing knowledge is scarce.

Following a thorough examination of the transcripts, recurring patterns and themes were identified and labeled as initial codes. As the analysis progressed and new insights emerged, the initial codes were revised and refined to ensure alignment with the underlying concepts. These codes were analyzed to determine their thematic significance and relevance, as well as their underlying themes. They were examined and assigned to their appropriate category based on their content and relationship to other codes. Throughout the entire process, reflective memos were also written and integrated into the analysis to foster data understanding and develop themes by linking codes to broader categories. This process of iterative data analysis, aimed at achieving data saturation, resulted in the identification of ten sub-codes and three overarching codes. Two coders reviewed the codes and sub-codes, and their inter-coder agreement was 83.33 percent, which was deemed satisfactory according to the recommended threshold of 80 percent agreement (Miles & Huberman, 1994). Following the qualitative data analysis and drawing on its findings, a questionnaire was developed to delve deeper into participants' understanding of fairness in online assessment. To ensure its content validity, the questionnaire was reviewed by three TEFL university professors with expertise in assessment

and subsequently revised based on their feedback. Next, to evaluate the modified questionnaire's construct validity and internal consistency, a pilot administration was conducted with 100 participants. To safeguard ethical standards, the same ethical procedure implemented in the qualitative phase was also maintained in this phase. Additionally, prior to answering the questionnaire items, the participants provided their informed consent by selecting the relevant agreement option. The collected data were analyzed using exploratory and confirmatory factor analysis to assess the questionnaire's validity, identify its underlying constructs, and examine their relationships.

Results

Qualitative Phase

The examination of the transcribed interview data from 20 participants identified three overarching levels of themes: the student level, the teacher level, and the organizational level.

The Grounded Theory approach was applied to examine the data and extract categories at every level, resulting in the identification of ten sub-codes: access, knowledge, personal factors, criteria, item quality, safety, policies, teaching quality, assessment type, and administration.

Given the findings of memo writing, initial data reading, reflections, and emerging concepts, the subcodes are examined at two levels: the reasons cited by the participants for the perceived unfairness in online assessment, and the potential measures that need to be taken to improve fairness in online assessment. In addition, referring to the literature which discussed fairness mainly from the views of the main stakeholders- i.e., students, teachers, and policy makers- and also patterns arising from the data, the findings are discussed at student, teacher, and organizational levels.

Concerning the first research question, the findings were examined through two analytical dimensions: the reasons participants mentioned for unfair online assessment and the measures they cited to increase fairness. The reasons are illustrated at three levels: student, teacher, and organization. Each level is composed of several reasons, elaborated in the following tables.

Table 2 displays the open and axial codes related to the student-level reasons.

Table 2

Reasons for unfairness at the student level

Student Level			
Knowledge	**Access**	**Personal Factors**	**Administration**
No technological and administrative knowledge	No access to tools Poor quality of tools and connection	Techno- phobia Financial, Cultural, Physical, Emotional, and Psychological conditions Different ability levels and types	Distracters in physical environment e.g., uncomfortable temperature, noise, lighting

As can be seen from the table above, several open codes were used to create the four axial codes of "Knowledge, Access, Personal factors, Administration", concerned with student level. For instance, concerning "Administration", participant 3 said that "sometimes the existence of unavoidable distractors in students' places influences their performance to the extent that it makes it unfair to compare their performance with those who take exams in places void of distractors".

Table 3 presents the reasons participants mentioned for the teacher level.

Table 3

Reasons for unfairness at the teacher level

Teacher Level									
Access	**Safety**	**Assessment Type**	**Teaching Quality**	**Criteria**	**Knowledge**	**Personal Factors**	**Items Quality**	**Policies**	**Administration**
Teachers' access to poor tools and connection No access to webcam No access to teachers during assessment	Questions with same answers No time limit Same forms of questions	No continuous assessment Multiple choice assessment	Poor quality Teaching to the test Poor relationship between teaching and assessment	Involving personal feeling in assessment No criteria for planning, administering, scoring, reporting, using, needed remedies Not considering class activities, only final score	No knowledge about students' personalities, families, cultural, financial status	Neglecting students' physical, psychological status No rapport with students, so more stress for students .Not considering individual differences	Asking superficial questions No oral questions Ambiguous questions Poor reliability and validity	Not involving students in assessment Strict limit One shot end term exams Not sharing the criteria with students	No uniformity in administration e.g., time, test rubrics

As shown in Table 3, the open codes were summarized into 9 axial codes: "Access, Safety of Assessment, Type of Assessment, Teaching Quality, Criteria, Knowledge, Personal Factors, Items Quality, and Policies". For example, regarding "Criteria," participant 6 maintained that "for some teachers, online assessment does not run using clear criteria, neither in design nor in administration." In terms of the axial code of "Quality," participant 14 said: "Assessment items that are ambiguous and poorly designed decline the fairness of assessment".

Table 4 illustrates the reasons mentioned concerning the organizational level.

Table 4

Reasons for un/fairness at the organizational level

Organizational Level							
Safety	**Personal Factors**	**Items Quality**	**Criteria**	**Teaching Quality**	**Assessment Type**	**Policies**	**Administration**
Questions with the	Not considering	No oral questions	No criteria for planning,	Poor relationship	No continuous assessment	Not involving teachers and	No uniformity in administrative

same answers No time limit Same forms of questions for all	individual differences	Ambiguous questions Poor reliability and validity	administering, scoring, reporting, using, and needed remedies Not considering class activities, only final score	between teaching, course content and assessment	Multiple choice assessment	students in assessment Aim for results and money rather than learning Setting strict time limit One shot end term exams Not sharing the criteria with students	procedure e.g. time, test rubrics

As shown in Table 4, the reasons related to "organizational level" are classified into 7 axial codes: "Safety, Personal Factors, Items Quality, Criteria, Teaching Quality, Assessment Type and Policies". Concerning "Policies", participant 10 mentioned that "no involvement of teachers and students in the process of online assessment threatens its fairness". With regards to "Assessment Type", participant 6 stated that "focusing solely on final assessment and disregarding continuous assessment lowers the fairness of the evaluating individuals' performance".

In terms of the measures proposed by the respondents to enhance the fairness of online evaluation, the data are presented at three levels: Student Level, Teacher Level, and Organizational Level. The measures of every level will be illustrated in the following tables.

Table 5 indicates the measures the students need to take to increase assessment fairness.

Table 5

Measures to increase fairness at the student level

Student Level
Access
Providing students with the tools and requirements by their families
Providing a webcam and a microphone
Checking the internet connection before the assessment

The measures raised by the participants at this level were related to "Access". For instance, participant 5 held that "families should provide the necessary tools to take an online assessment for students". Table 6 presents the measures teachers can take to improve assessment fairness.

Table 6

Measures to increase fairness at the teacher level

Teacher Level Measures									
Personal factors	**Criteria**	**Knowledge**	**Access**	**Administration**	**Assessment Type**	**Teaching Quality**	**Items Quality**	**Safety**	**Policies**
Encouraging students Considering students' physical, psychological conditions Considering students' differences	Scores for class participation No involvement of personal feeling Clear criteria	Getting information about learners' family, financial and cultural status Tool use knowledge	Teachers' access to right tools and connection Students' access to teachers during assessment	Taking remedial measures in the case of distracters Uniformity in time, tools, administrative instruction	Using different modes Continuous assessment Open-ended assessment	Relationship between teaching and assessment	Clarifying ambiguous items Assuring item validity	Asking for students' webcams Making a rapport with students Asking on the spot Explaining the harms of cheating Being in touch with students' parents Shuffling answer choices Comparing answers Devising oral exams Devising different forms of questions Setting a reasonable time limit Asking questions with different answers Open-ended questions	Involving students in assessment Offering backup support Pre-sharing the criteria Aiming for learning in assessment

As shown in Table 6, the open codes were categorized into 10 axial codes of "Personal Factors, Criteria, Knowledge, Access, Administration, Assessment Type, Teaching Quality, Items Quality, Safety, and Policies". In the case of Personal Factors, for instance, participant 7 mentioned that "teachers should take into account the differences existing in individuals about their needs and wants which influence their performance". Regarding "Safety", interviewee 12 said that "to ensure assessment safety, students' ability needs to be assessed through different forms of assessment, both orally and written, and both multiple choice and open-ended". Participant 9 also commented: "Having a friendly relationship and building trust among students and teachers can help reduce cheating in online assessment". Table 7 provides the measures to increase fairness at the organizational level.

Table 7

Measures to increase fairness at the *organizational level*

Organizational Level									
Safety	**Access**	**Assessment Type**	**Administration**	**Teaching Quality**	**Items Quality**	**Personal Factors**	**Policies**	**Criteria**	**Knowledge**
Shuffling answer choices Comparing answers Devising oral exams Devising different forms of multiple choice questions Setting reasonable time limit Asking questions with different answers Asking open-ended questions in written assessment	Providing access to right tools and connection	Using different modes Continuous assessment Open-ended assessment	Uniformity in time, required devices, type of administrative instruction	Relationship between teaching, course content and assessment	Valid items Clarifying ambiguous items Unbiased items	Considering learners' individual differences in their abilities, preferences and needs	Involving students and teachers in assessment Offering back up support e.g., remedial measures Pre-sharing criteria Aiming for learning	Setting criteria Clear criteria for planning, administering, scoring, reporting, using, needed remedies	Raising students' knowledge about assessment tools

As Table 7 shows, 10 axial codes were identified as "Safety of Assessment, Access, Type of Assessment, Administration, Items Quality, Personal Factors, Policies, Criteria, and Knowledge". With respect to "Teaching Quality", participant 10 said that "whatever is taught must be assessed". Additionally, regarding Policies, interviewee 13 mentioned that "students must be informed about the assessment criteria before holding it".

Quantitative Phase Results

To answer research question two, the Online Assessment Fairness Scale (OAFS) was developed based on thematic insights from 20 interviews. Analyzed using grounded theory, interview data yielded 10 axial codes, which informed the construction of 55 items for the initial scale. The items aimed to reflect qualitative themes, ensuring theoretical and content validity. The scale was piloted with 100 EFL teachers, and an Exploratory Factor Analysis (EFA) was conducted on the collected data. The initial results of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) indicated a value of 0.49, which fell below the acceptable threshold of 0.60. As a result, 30 of the items with low factor loadings- especially those lacking

theoretical distinctiveness- were subsequently eliminated and upon re-evaluation using KMO Measure of Sampling Adequacy and Bartlett's Test of Sphericity, the results showed improvement (see Table 8)

Table 8

KMO and Bartlett's test for the OAFS

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.622
Bartlett's Test of Sphericity	Approx. Chi-Square	1055.513
	df	465
	Sig.	.000

As is evident in Table 8 above, the KMO Measure of Sampling Adequacy was 0.62, and Bartlett's Test of Sphericity was statistically significant (X^2 (465) = 1055.51, $P < 0.05$). Using Kaiser's criterion (eigenvalues > 1), a total of 11 factors were initially extracted, accounting for 70.16% of the total variance, as shown in Table 9.

Table 9

Total Variance Explained by Extracted Factors of the OAFS

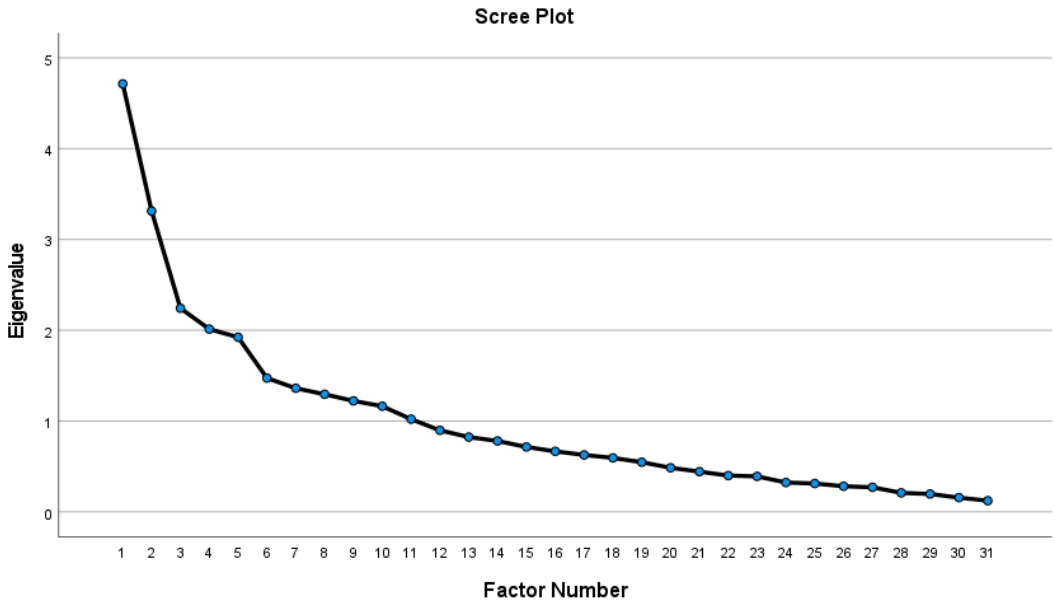
Total Variance Explained						
	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.714	15.208	15.208	2.833	9.137	9.137
2	3.314	10.689	25.897	2.524	8.143	17.281
3	2.243	7.235	33.132	2.284	7.369	24.649
4	2.014	6.495	39.628	1.835	5.919	30.568
5	1.924	6.206	45.833	1.796	5.795	36.363
6	1.474	4.755	50.588			
7	1.362	4.393	54.981			
8	1.295	4.179	59.160			
9	1.223	3.944	63.105			
10	1.164	3.755	66.860			
11	1.021	3.295	70.155			

Extraction Method: Principal Axis Factoring

However, an inspection of the screeplot (Figure 1) revealed a clear break after the fifth component.

Figure 1

Scree Plot



Also, a parallel analysis was conducted to determine the number of factors to retain for the OAFS. Comparing the actual eigenvalues to those derived from randomly generated eigenvalues at the 95th percentile indicated that only the eigenvalues of the first five factors exceeded the criterion value from parallel analysis, supporting the retention of five factors. (Table 10).

Table 10

Comparison of Actual Eigenvalues from EFA, and Criterion Values from Parallel Analysis

Component number	Actual eigenvalue	Criterion value from parallel analysis	Decision
1	4.714	1.958302	Accept
2	3.314	1.828117	Accept
3	2.243	1.723099	Accept
4	2.014	1.644127	Accept
5	1.924	1.554686	Accept
6	1.474	1.482716	Reject
7	1.362	1.416175	Reject
8	1.295	1.352853	Reject
9	1.223	1.291736	Reject
10	1.164	1.235872	Reject
11	1.021	1.181243	Reject

The results of the parallel analysis, along with the total variance explained and the screeplot, support retaining only the first five factors.

In Table 9, the analyses yielded five factors, explaining a total of 45.83 percent of the variance in the entire set of variables. Following rotation, the five retained factors accounted

for 36.36% of the total variance, with individual contributions of 9.14%, 8.14%, 7.37%, 5.92%, and 5.80%, respectively (see Table 9). The factor loadings of the revised scale are presented in Table 11 below.

Table 11

Factor loadings for the rotated factors of the OAFS

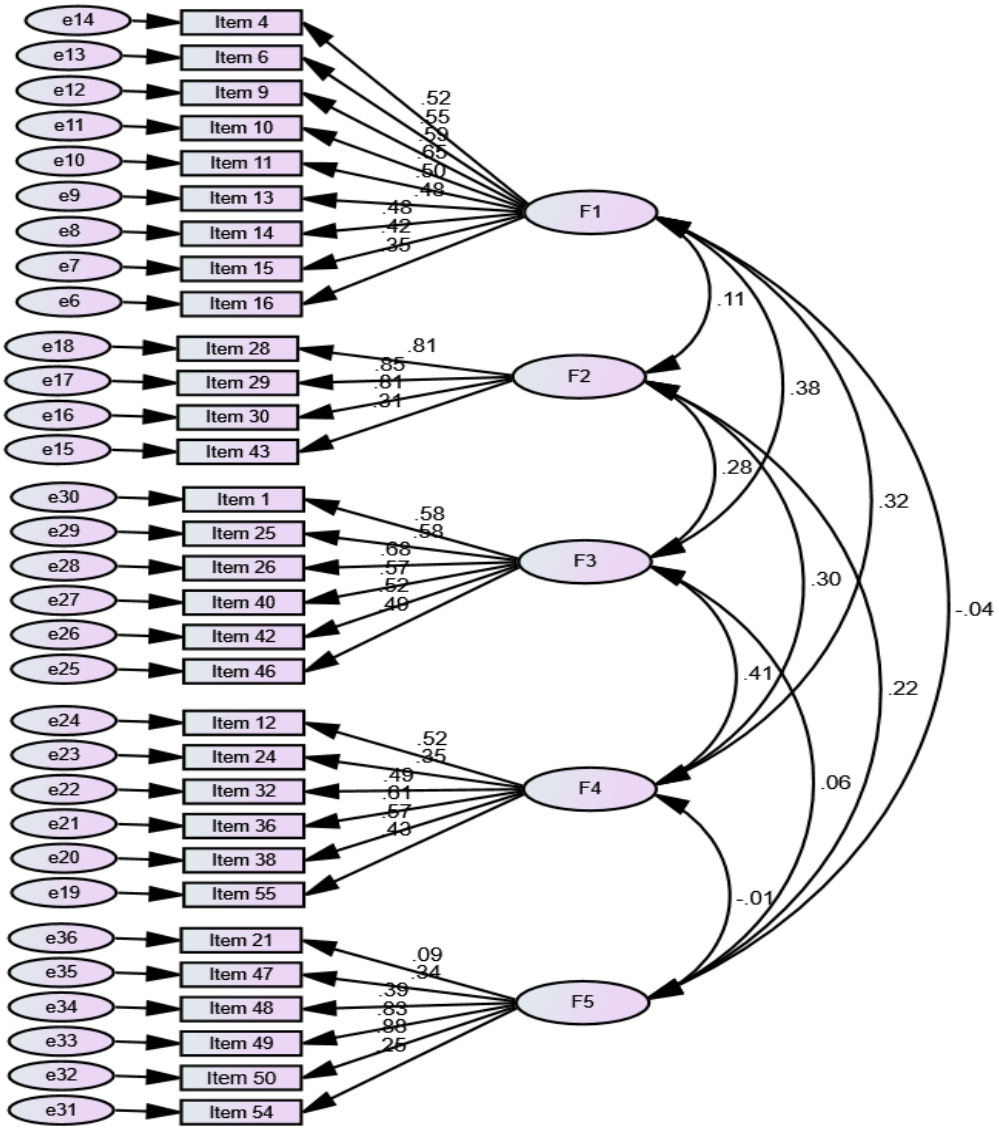
Item	Factor loading				
	1	2	3	4	5
Item 1			0.523		
Item 4	0.509				
Item 6	0.612				
Item 9	0.567				
Item 10	0.702				
Item 11	0.621				
Item 12				0.59	
Item 13	0.592				
Item 14	0.482				
Item 15	0.563				
Item 16	0.413				
Item 21					0.442
Item 24				0.572	
Item 25			0.74		
Item 26			0.681		
Item 28		0.849			
Item 29		0.808			
Item 30		0.813			
Item 32				0.569	
Item 36				0.634	
Item 38				0.61	
Item 40			0.534		
Item 42			0.621		
Item 43		0.409			
Item 46			0.566		
Item 47					0.507
Item 48					0.562
Item 49					0.725
Item 50					0.784
Item 54					0.477
Item 55				0.54	
Eigenvalues	4.714	3.314	2.243	2.014	1.924
% of variance	15.208	10.689	7.235	6.495	6.206

As indicated in Table 11, the factors explained 15.20, 10.68, and 7.23, 6, 49, 6, 20 percent of the total variance, respectively.

To confirm the construct validity of the Online Assessment Fairness Scale (OAFS), a CFA was conducted using the Amos 24 statistical package**.** The results are presented in Figure 2 below.

Figure 2

The CFA of the OAFS (Before Modification)



Several fit indices were examined to evaluate the model fit (Table 12).

Table 12

Goodness of fit indices for the OAFS (before modification)

	X^2/df	GFI	IFI	TLI	CFI	RMSEA
Acceptable fit	<3	>.90	>.90	>.90	>.90	<.08
Model	1.433	.746	.760	.722	.747	.066

As shown in Table 12, most fit indices (i.e., GFI, IFI, TLI, and CFI) did not fall within acceptable fit thresholds. Therefore, the model underwent modification. To modify the model, low-factor-loading items were removed, and residual covariances were considered. On this basis, 30 items were made redundant, and the resulting scale was rechecked. However, their exclusion was not based solely on statistical misfit; each redundant item was also examined against the themes and data in the qualitative phase. These items were excluded only if their underlying concepts were represented elsewhere within the scale.

The modified model is presented in Figure 3, and the goodness-of-fit indices for the CFA results of the modified scale are presented in Table 13 below.

Figure 3

The CFA of the OAFS (After Modification)

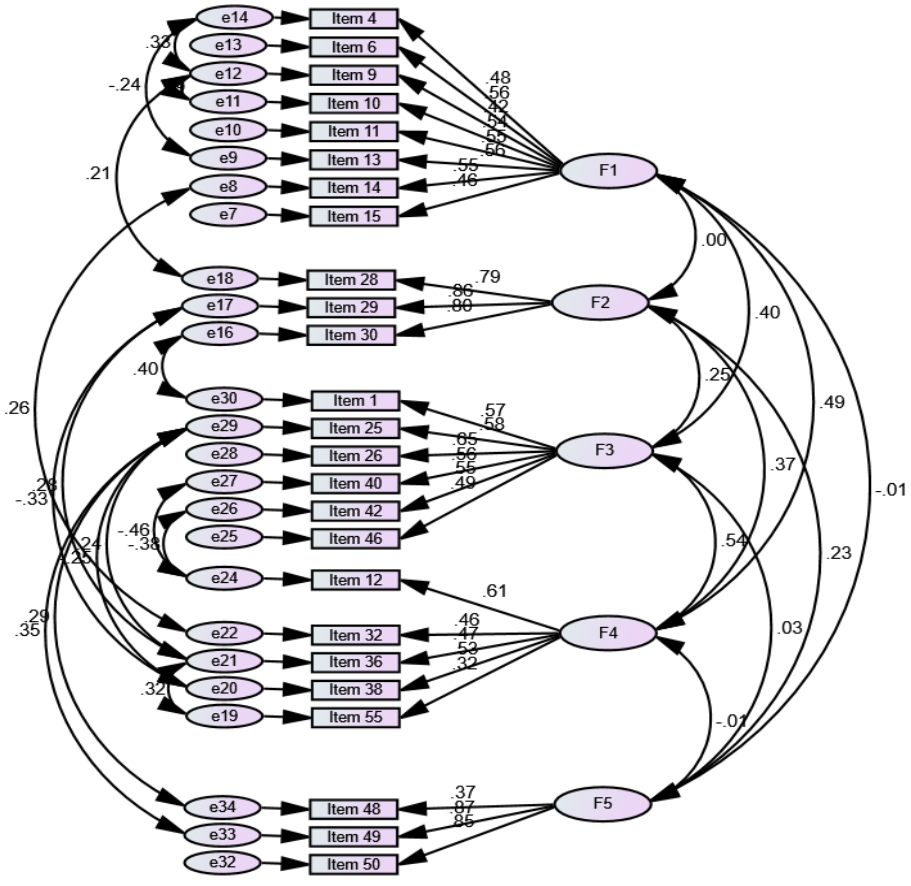


Table 13

Goodness of fit indices for the OAFS (after modification)

	X^2/df	GFI	IFI	TLI	CFI	RMSEA
Acceptable fit	<3	>.90	>.90	>.90	>.90	<.08
Model	.989	.852	1.004	1.005	1.000	.000

As shown in Table 13, almost all fit indices (i.e., chi-square/df ratio (.989), RMSEA (.000), IFI (1.004), TLI (1.005), and CFI (1.000)) lay within the acceptable fit thresholds. Furthermore, as evident in GFI=.852, it did not fall within the acceptable fit thresholds. However, it has been verified that GFI<.90 in small samples is acceptable (Mulaik et al., 1989). Consequently, it can be concluded that the model had a perfect fit with the empirical data after modification.

The analysis yielded 25 items, categorized into five distinct factors, while retaining all major domain content. In addition, the concepts raised in the quantitative phase were shown to align with those in the qualitative part (Table 14). See Appendix B for the remaining items after factor analysis.

Table 14

Mapping of Qualitative Axial Codes to Final CFA Factors.

Final CFA Factors	Representative Axial Codes from Qualitative Phase
Uniformity and Consistency in Administration, Access, Knowledge/ Quality of the Items	Administration, Access, Knowledge. Item Quality
Administrative Safety and Monitoring	Safety
Criteria and Policies	Policy, Criteria, Teaching Quality
Personal Requirements and Feedback	Personal factors, Assessment Type
Item Design Safeguards	Safety

Discussion

This study aimed to assess Iranian EFL teachers' perception of fairness in online assessment. The findings demonstrated that EFL teachers were generally cognizant of the significance of fairness in the online assessment context. The findings further underscored that the fairness of online assessment needs to be considered at the student, teacher, and organizational levels, and factors including Access, Knowledge, Personal Factors, Clear Criteria, Items Quality, Safety, Policies, Teaching Quality, Assessment Type, and Administration were found as the componential aspects of online assessment fairness.

With respect to "Administration", the participants pointed out several physical setting distractors, such as uncomfortable temperature, noise, and lighting, which may prevent learners from demonstrating their full potential. Heissel et al. (2021) reported that environmental distractors can influence test takers' psychological state and, consequently, their test performance. To mitigate such negative effects, several participants suggested that test developers and administrators anticipate these intervening factors and be prepared to implement remedial measures. This aligns with Bachman and Palmer (2010), who proposed that, prior to administering an assessment, test developers should consider the type of assistance they expect proctors to provide throughout the test administration. Additionally, to ensure fair assessment, the participants emphasized the importance of uniformity in time, devices, and the type of administrative instruction, as well as clear instructions prior to assessment, which are underscored in the literature (see Bchman & Palmer, 2010; Kunnan, 2000).

In relation to "Safety", which pertains to assessment security, CRADLE (Centre for Research in Assessment and Digital Learning) states that it aims at detecting, reducing, and eliminating cheating in assessment (Dawson et al., 2020). According to Kunnan's (2004) test fairness framework (TFF), appropriate test security is a crucial aspect of the administration module. It is also highlighted as a major challenge in online platforms (García-Peñalvo et al., 2021). One key aspect of assessment security is preventing cheating, which undermines the fairness of the assessment. However, Murdock et al. (2007) and Newton and Essex (2023) argued that cheating may increase when students perceive assessment practices as unfair. As a result, there seems to be a reciprocal relationship between the fairness and security of assessments.

Consistent with Bretag et al. (2018), participants in this study proposed various measures in this regard, including maintaining high-quality courses, teaching, and assessment. In addition, the participants suggested fostering good relationships with learners as highlighted by Allen and Kizilcec (2023), who noted that the absence of mutual trust and relationships between teachers and students can lead to increased dishonesty among students. Further, the findings depicted that raising students' awareness of the harms of cheating can contribute to its reduction. The elimination of unethical practices such as cheating and fraudulent test-taking behavior needs to be a top educational goal. According to Denisova-Schmidt (2017), students might carry over their unethical educational practices to their future professional and personal lives.

Similar to what Perkins et al. (2020) called for, the findings also confirmed that the teachers need to be trained and instructed on how to use the internet and detect various forms of online cheating. Nevertheless, awareness and detection of cheating do not ensure that educational organizations and authorities will take action against such misconduct, as their motives might be driven by financial or political interests (Allen & Kizilcec, 2023). In line with this, the third principle of the International Language Testing Association (ILTA) Codes of Ethics (2018) implies that the human rights of test takers should be prioritized over the interests of society.

The participants also emphasized the importance of equal access to tools, internet connections, teachers, and prior instruction on tool use, which were collectively identified as "Access". In this respect, Kunnan (2004) states that a fair test should have educational, financial, and equipment access. Also, in line with the ETS Standards for Fairness and Quality (ETS, 2002), individuals need impartial and unbiased access to products and services. Furthermore, McNamara and Roever (2006) and Bchman and Palmer (2010) underscore the significance of fair and equitable access for all test takers to resources, equipment, costs, and assessment procedures.

As noted by the participants and emphasized in earlier studies, to ensure fair assessment, it is necessary to evaluate learners' performance using comparable testing methods, procedures, content, administration, scoring, and interpretation (Bchman & Palmer, 2010; Kunnan, 2008, 2018). However, requiring all learners, with inherent distinctive characteristics, to meet the same testing requirements may not necessarily represent fairness. In this regard, the participants

discussed a range of "Personal Factors" that affect students' performance, including physical, psychological, and cognitive ones. Partially consistent with this finding, Lam (1995) underscored the importance of equity in assessment, advocating that each student be assessed as an individual. This includes taking into account each student's instructional context and personal history. Additionally, Camilli (2013) proposes the concept of individual fairness in assessment, which involves treating students comparably rather than equally. As mentioned in Tierney (2016), democratic values now advocate for educational systems that are inclusive and supportive of student diversity. To address these individual differences, Rawls (2001) suggests that when inequalities exist, the least advantaged individuals should benefit most from positive discrimination.

Further, participants noted that teachers' lack of interaction with students during assessment could heighten students' stress levels. Assessment must be at the service of learning, and to this end, there must be a trustworthy, respectful, and engaging learning context in which learners feel psychologically secure to make mistakes (Tierney, 2014; 2016).

With respect to the "Knowledge" factor, the participants highlighted deficiencies in both technological and administrative knowledge of online assessment among learners, as well as a lack of prior instruction on how to use the tools. In line with this, earlier studies in their definitions of assessment fairness underscore the significance of prerequisite knowledge, skills and abilities (Bachman & Palmer, 2010; Kunnan, 2018; McMillan, 2011, Tierney, 2014). Additionally, participants noted that teachers often lack adequate knowledge of their students' emotional well-being, family backgrounds, cultural identities, and financial status.

In the case of "Item Quality", the participants expressed concerns about the ambiguity of the items. Volante (2006) emphasizes that to ensure a fair assessment, it is necessary to ensure the quality of items through careful item construction, the application of multiple test formats, and reviews prior to assessment. In addition, for a test to be fair and just, both the test as a whole and its items must be statistically robust and free of bias (Tierney, 2016; Xi, 2010).

In terms of "Criteria", some participants noted that the absence of specific, predetermined criteria for assessment might lead test givers to base their assessments on intuition. According to Elder et al. (2005), intuition is generally not recommended because it reduces reliability. Earlier studies also highlight the need to clearly and reasonably specify criteria to ensure fair classroom assessment (Camili, 2006; Russell & Airasian, 2012; Tierney, 2014, 2016; Volante, 2006). Also, as Tierney (2016) proposes, students' engagement enhances when they are aware of assessment criteria. Therefore, the principle of transparency raises fairness in assessment by minimizing the influence of irrelevant factors. In line with this, Tofighi and Ahmadi Safa (2023) highlighted the significance of transparency in their study on fairness in classroom assessment. They emphasized the importance of clarifying policies, procedures, and decisions, and of stating their significance at the beginning of the course as requirements for fair practices. Moreover Beheshti and Ahmadi Safa (2023), in their investigation of the empirical validity of Kunnan's (2008) Test Fairness Framework stressed the importance of explicit standards. They

demonstrated that accuracy and standard-setting must be achieved not only during the test development process and selection decisions, but also throughout all stages of testing, including test development, administration, scoring, reporting, and decision-making.

Furthermore, the respondents reported that teachers' feelings and biases must not interfere with any part of the assessment process, as emphasized by Kunnan (2004) and Russell and Airasian (2012). This finding implies that assessors need training in setting criteria and maintaining consistency in scoring which as Eckes (2005) maintains is one of the basic assumptions of fair testing.

Regarding "Policy", several participants commented that students are not usually consulted and shared the type of assessment. In line with this, Suskie (2002) recommends the inclusion of the test takers in defining and determining aspects of the assessment process. Additionally, the findings showed that to enhance assessment fairness, students should be adequately informed about expectations, learning targets, and course requirements prior to instruction and assessment. Similar to this piece of finding, McMillan (2011) and Russell and Airasian (2012) also suggested such test takers' participation in the assessment process.

Another important finding indicated that teachers should be involved in the processes of test design, administration, and scoring. However, it should be noted that the kinds of actions teachers can take regarding assessment are constrained by the perceptions of society and those in power (Fulcher, 2009). Consistent with this, Hamp-Lyons (2007) contends that the status and recognition of teachers by societies influence the success of teacher-based and classroom-based assessment. Furthermore, a number of participants noted that some testing authorities and organizations tend to prioritize financial gains over genuine learning. In these cases, tests or assessments reflect the interests of those in power and are used to exercise control over test takers. Therefore, it is unreasonable to expect a test or assessment driven by unfair motives to promote fairness.

With respect to "Teaching Quality", the participants referred to inadequate teaching quality and the lack of alignment between teaching, the content of course books, and assessment as sources of assessment unfairness. This finding is consistent with Tierney (2014, 2016), who argues that ensuring fairness in classroom assessment is an ongoing process that must run through the cycles of teaching, learning, and assessing. Additionally, Camilli (2006) and McMillan (2011) assert that, for fair classroom assessment, the assessment process should provide learning opportunities for learners.

Regarding "Assessment Type", the participants referred to the use of different modes, continuous assessment, and open-ended assessment as factors behind test unfairness. Such statements confirm the findings of Suskie (2002), Tierney (2014, 2016), and Volante (2006) who emphasized the use of a range of summative and formative assessment strategies. Additionally, Tofighi and Ahmadi Safa (2023) highlighted the need to provide learners with diverse assessment forms and activities to demonstrate their learning. Likewise, Azizi (2022) suggests that by incorporating multiple online assessments using a variety of instruments,

individual differences among students can be considered. On this basis, having a range of assessment approaches in online assessment seems to provide more opportunities to meet students' needs, styles, and wants.

Concerning the second research question i.e. "What are the factors of a valid measure of online assessment fairness?", the piloting phase resulted in 25 items of OAFS that were subsequently categorized into 5 factors of (1) Uniformity and Consistency in Administration, Access, Knowledge and Quality of the Items, (2) Administrative Safety and Monitoring, (3) Criteria, and Policies, (4) Personal Requirements and Feedback, (5) Item Design Safeguards. These factors were also confirmed by the insights participants shared during the interview phase. The outcome of this phase of the study aligned with the findings from the qualitative phase, reflecting the thematic patterns derived from the qualitative interviews. The five factors identified correspond to the ten axial codes extracted from the interview. For instance, factor 1 represents concerns about Administration, Access, and Knowledge and Item Quality, while factor 2 mirrors Safety protocols and tools during test administration. Likewise, factor 3 mirrors Policy, Criteria, and Teaching Quality, emphasizing points such as transparency, the involvement of teachers and students in assessment, and the alignment between teaching and assessment. Factor 4 reflects the concepts of Personal factors and Assessment Type, such as offering accommodations when required, providing feedback, caring about test-takers' physical and psychological well-being during the assessment, building rapport, and using formative assessment. Similarly, factor 5 focuses on the structural design of item formats, such as time restrictions and shuffling responses and questions to safeguard the item format. This consistency between the findings of both phases reinforces the theoretical validity of the scale and confirms that the extracted model is rooted in participants' lived experiences and perceptions.

With respect to the first factor, Kunnan (2004), in his TFF, highlights the importance of uniformity and consistency in administration across test sites. However, it is crucial to note that uniformity should go beyond test administration and reliability and extend to all phases of design and decision-making about assessment. Moreover, ensuring uniform and consistent access for all test takers to assessments and the required knowledge is necessary. In line with this, Camilli (2006) operationalizes fairness as the equity in both assessment and instruction. Furthermore, Aryadoust et al. (2011) assert that in a fair test, all test takers should be provided with equal opportunity to demonstrate their knowledge, abilities, and skills.

Regarding the factors of "Administrative Safety and Monitoring" and "Item Design Safeguards", in the quantitative phase, the safety of the testing procedure and the safety of items were identified as distinct factors. However, in the qualitative phase, both of these aspects were discussed under the umbrella term Safety. As previously stated, test security or safety has been considered as an aspect of the administration facet in Kunnan's TFF (2004), encompassing issues such as fraud, misrepresentation, cheating, and plagiarism. Regarding online assessment, cheating appears to pose a greater threat to security than other concerns (García-Peñalvo et al.,

2021). As discussed in the qualitative phase of the current study, a large number of measures need to be taken in this respect, ranging from raising awareness among all stakeholders involved about the harms of cheating and the ways to detect and confront it to establishing negative views in society towards it.

Concerning "Criteria and Policies", the comparison of the findings with those of other studies highlighted the necessity of transparent criteria for fair assessment (see Beheshti & Ahmadi Safa, 2023; Camilli, 2006; Suskie, 2002; Tierney, 2014; Tofighi & Ahmadi Safa, 2023). Emphasizing criteria, Camilli (2006) operationalizes fairness as the establishment of clear, sensible assessment criteria. Furthermore, Suskie (2002) emphasizes, in the definition of fairness, the importance of clearly expressing learning outcomes. Moreover, Volante (2006) underscores students' involvement in the development of assessment criteria as an essential aspect of ensuring fairness.

Finally, regarding "Personal Requirements and Feedback", prior investigations have emphasized the need for different samples and also types of measures to ensure fair assessment (Suskie, 2002; Volante, 2006). This issue is referred to as continuous assessment in the current study, as discussed earlier.

Conclusion

The findings of this study revealed that Iranian EFL teachers were conscious of the factors contributing to un/fair online assessment and the necessary measures to address them. They voiced concerns regarding access, knowledge, personal factors, clear criteria, item quality, safety, policies, teaching quality, assessment type, and administration in online assessments. These insights offer valuable implications for a range of stakeholders. Researchers can use the Online Assessment Fairness Scale (OAFS) to explore fairness across online contexts, while teachers and teacher educators can gain insights about enhancing fairness in assessment. Applying the results and the scale, policymakers will be able to craft institutional guidelines to ensure fairness across assessment practices. Further, the results illuminate areas for_test designers and developers to develop clearer and more inclusive assessment formats. Likewise, scorers and evaluators will benefit from the defined constructs that inform more equitable and ethical assessment practices.

Regarding the limitations of the present study, it should be noted that both phases employed rather small sample sizes, which might have led to findings that are not necessarily reflected. Therefore, future investigations with large sample sizes are required to examine all potential factors involved in online assessment unfairness. Collectively, further empirical studies into the perceptions of various stakeholders, using a range of methods, are recommended to shed light on this multifaceted construct.

Acknowledgments

We would like to thank the editorial team of TESL Quarterly for granting us the opportunity to submit and publish the current synthesis. We would also like to express our appreciation to the anonymous reviewers for their careful, detailed reading of our manuscript and their many insightful comments and suggestions. We also acknowledge all the participants who took part in this study.

Declaration of conflicting interests

The authors declare no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for this article's research, authorship, and/or publication.

References

Allen, S., & Kizilcec, R. F. (2023). A systemic model of academic (mis)conduct to curb cheating in higher education. *Higher Education*, 87(5), 1-21. https://doi.org/10.1007/s10734-023-01077-x

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Aryadoust, V., Goh, C. C. M, & Kim, L. O. (2011). An investigation of differential item functioning in the MELAB listening test. *Language Assessment Quarterly*, 8(4), 361–385. https://doi.org/10.1080/15434303.2011.628632

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

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Beheshti, Sh., & Ahmadi Safa, M. (2023). Reconceptualization of test fairness model: A Grounded Theory approach. *Iranian Journal of Language Teaching Research 11*(2),119-146. DOI: 10.30466/ijltr.2023.121333

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Saddiqui, S., Rozenberg, P., & van Haeringen, K. (2018). Contract cheating: A survey of Australian university students. *Studies in Higher Education*, 44 (1), 1-20.DOI:10.1080/03075079.2018.1462788

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Praeger

Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research and Evaluation*, *19*(2-3), 104-120. https://doi.org/10.1080/13803611.2013.767602

Cole, N.S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38(4), 369–382. https://psycnet.apa.org/doi/10.1111/j.1745-3984.2001.tb01132.x

Comrey, A.L., & Lee, H.B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum

Creswell, J.W., & Poth, C.N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE.

Davies, A. (2010). Test fairness: A response. *Language Testing*, *27*(2), 171–176. https://doi.org/10.1177/0265532209349466

Dawson, P., Sutherland-Smith, W., & Dullaghan, K. (2020). *CRADLE suggests… academic integrity, assessment security and digital assessment.* Centre for Research in Assessment and Digital Learning, Deakin University. https://doi.org/10.6084/m9.figshare.12585443

Denisova-Schmidt, E. (2017). *The challenges of academic integrity in higher education: Current trends and prospects.* CIHE Perspectives No. 5. The Boston College Center. International Higher Education.

Deygers, B. (2017). Just testing: Applying theories of justice to high-stakes language tests. *ITL: International Journal of Applied Linguistics*, *168*(2), 143–163.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford University Press.

Eckes, T. (2005) Examining rater effects in Test DaF writing and speaking performance assessments: a many-facet Rasch analysis. *Language Assessment Quarterly 2*(3), 197–221. https://doi.org/10.1207/s15434311laq0203 2

Educational Testing Service. (2002). *ETS standards for quality and fairness*. Author.

Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly 2* (3), 175–196. https://doi.org/10.1207/s15434311laq0203_1

Fulcher, G. (2009). Test use and political philosophy. *Annual Review of Applied Linguistics 29*, 3–20. https://doi.org/10.1017/S0267190509090023

García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2021). Recommendations for mandatory online assessment in higher education during the COVID-19 pandemic. In D. Burgos, A. Tlili, & A. Tabacco (Eds.), *Radical solutions for education in a crisis context* (pp. 85–98). Springer: https://doi.org/10.1007/978-981-15-7869-4_6

Ghanbari, N., & Nowroozi, S. (2022). Iranian EFL teachers' challenges and coping strategies during COVID-19 pandemic: A case study. *The Qualitative Report, 27(*3), 605-625. https://doi.org/10.46743/2160-3715/2022.5066

Hamp-Lyons, L. (2001). Ethics, fairness, and developments in language testing. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honor of Allan Davies* (pp. 222-227). Cambridge: Cambridge University Press.

Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching* (pp. 487–504). Springer.

Heissel, J. A., Adam, E. K., Doleac, J. L., Figlio, D. N., & Meer, J. (2021). Testing, stress, and performance: How students respond physiologically to high-stakes testing. *Education Finance and Policy. 16*(2). 183-208. 10.1162/edfp_a_00306.

International Language Testing Association. (2018). *Code of ethics for ILTA.* http://www.iltaonline.com/code.pdf

Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1-13). Cambridge University Press.

Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context* (pp. 27–48). Cambridge University Press.

Kunnan, A. J. (2008). Towards a model of test evaluation: Using the test fairness and test context frameworks. In L. Taylor & C. Weir (Eds.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity* (pp. 229–251). Cambridge University Press.

Kunnan, A. J. (2010). Test fairness and Toulmin's argument structure. *Language Testing*, *27*(2), 183–189. https://doi.org/10.1177/0265532209349468

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

Lam, T. C. M. (1995). Fairness in performance assessment (ERIC Digest No. ED391982). ERIC. http://www.ericdigests.org/1996–4/fairness.htm

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

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications.

Momeni, A. (2022). Online assessment in times of COVID-19 Lockdown: Iranian EFL teachers' perceptions. *International Journal of Language Testing, 12*(2), 1-24. doi: 10.22034/ijlt.2022.157122

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*(3), 430-445.

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*(2), 141–169. https://doi.org/10.1007/s11218-007-9015-1.

Newton, P. M., & Essex, K. (2023). How common is cheating in online exams and did it increase during the COVID-19 pandemic? A systematic review. *Journal of Academic Ethics, 22*(3), 323-343. https://doi.org/10.1007/s10805-023-09485-5

Perkins, M., Roe, J., Chonua, G. K., & Bhatia, A. (2020). Student perceptions of peer cheating behavior during COVID-19 induced online teaching and assessment. *Higher Education Research & Development, 43*(6), 966-1-15. https://doi.org/10.1080/07294360.2023.2258820

Rawls, J. (2001) *Justice as fairness: A restatement* (2nd ed.). Belknap Press.

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

Sen, A. (2010). *The idea of justice*. Penguin Books.

Stein, Z. (2016). *Social justice and educational measurement: John Rawls, the history of testing and the future of education* (1st ed.). Routledge.

Suskie, L. (2002). Fair assessment practices: Giving students equitable opportunities to demonstrate learning. *Adventures in Assessment 14*, 5-10.

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. D. (2016). Fairness in educational assessment. In M. A. Peters (Ed.), *Encyclopedia of Educational Philosophy and Theory*. Springer. https://doi.org/10.1007/978-981-287-532-7_400

Tofighi, S., & Ahmadi Safa, M. (2023). Fairness in classroom language assessment from EFL teachers' perspective. *Teaching English as a Second Language Quarterly, 42*(2), 81-110. Https://doi.org/10.22099/tesl.2023.46825.3173

Volante, L. (2006). Principles for effective classroom assessment. *Brock Education Journal, 15*(2), 134-147.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147–170. http://doi.org/10.1177/0265532209349465

Appendix A

Interview Items

۱. به نظر شما باید چه ارتباطی بین محتوای تدریس و ارزیابی آنلابن زبان آموزان وجود داشته باشد؟

۲. به نظر شما نمرات زبان آموزان در دو آزمون آنلاین که به لحاظ محتوا معادل یکدیگر هستند اما به لحاظ شکلی متفاوتند چگونه باید باشد؟

۳. به نظر شما نمرات زبان آموزان در دو آزمون آنلاین که در دو موعد زمانی مختلف برگزار می گردند اما از نظر محتوا و شکل آزمون یکسان هستند چگونه باید باشد؟

۴.آیا شما معیار یا مدلی را برای ارزیابی آنلاین زبان آموزان در نظر میگیرید ؟در صورتی که پاسخ شما مثبت است لطفا توضیح بفرمایید.

۵. آیا نمره ای که زبان آموزان در ارزیابی آنلاین از شما دریافت میکنند ضرورتا بر اساس معیارهای شماست ؟

۶.نظر شما در مورد تاثیر عوامل غیر مرتبط با موضوع ازمون بر نمرات زبان آموزان در ارزیابی آنلاین چیست؟ (مانند استرس، ابهام در سوالات ،خستگی زبان اموزان اموزان، شرایط فیزیکی محیط از قبیل گرما و سرما و ...)

۷.در صورتیکه زبان آموزی از نظر توانایی ذهنی در یادگیری مشکلی داشته باشد (learning disability) ، شما چگونه این موضوع را در ارزیابی اتلاین خود لحاظ میکنید؟

۸. ایا ارزیابی آنلاین شما از زبان آموزان تاثیری برنحوه تدریس و اموزش شما به زبان آموزان دارد ؟ اگر بله لطفا توضیح بفرمایید.

۹.نظر شما در مورد تاثیرات اجتماعی ارزیابی آنلاین چیست؟(social consequence)

10.نظر شما در مورد ضرورت آشنایی قبلی زبان آموزان با تجهیزات و شرایط آزمون و ارزیابی آنلاین چیست؟برای مثال چطور از موبایل ، تبلت و یا کامپیوتر برای پاسخ دهی استفاده کنند.

11.اگر بعد از ارزیابی آنلاین مشخص شود که در ارزیابی شما مشکلاتی وجود داشته که بر عملکرد زبان آموزان تاثیر گذار بوده ، شما چه اقداماتی انجام میدهید ؟

12.آیا شما به عنوان یک معلم در طراحی سوالات و ارزیابی آنلاین زبان آموزان احساسات شخصی خود را نیز دخیل می‌بینید ؟

۱۳.به نظر شما آیا معلم باید این اختیار را داشته باشد که بر اساس معیارهای خود نوع ارزیابی آنلاین خود را انتخاب کند؟

۱۴ .به نظر شما آیا زبان آموزان باید این اختیار را داشته باشند که بر اساس معیارهای خود نوع ارزیابی آنلاین خود را انتخاب کنند؟

۱۵.آیا سنجش عادلانه در ارزیابیهای آنلاین به معنای ارزیابی برابر و یکسان همه زبان آموزان است؟

۱۶.چه خطراتی سنجش عادلانه را در ارزیابی های آنلاین تحت تاثیر قرار میدهد؟

۱۷. به نظر شما در مبحث عدالت آموزشی چه تفاوتی بین ارزیابی آنلاین و حضوری وجود دارد؟

۱۸.اگر در حین آزمون آنلاین دسترسی زبان آموز به آزمون قطع شود(به دلیل قطع برق و یا وضعیت اینترنت) ، شما چه اقدامی را صورت میدهید؟

۱۹.اگر در حین آزمون آنلاین ابهامی در سوالات برای زبان آموزان به وجود بیاید شما چگونه این ابهامات را برطرف میکنید؟

۲۰.به نظر شما چه اقداماتی را برای **شناسایی** تقلب در آزمون های آنلاین میتوان صورت داد؟

۲۱.به نظر شما چه اقداماتی را برای **کاهش** تقلب در آزمون های آنلاین میتوان صورت داد؟

Appendix B

Online Assessment Fairness Scale (OAFS) in EFL Context

Dear Teachers

The aim of this survey is to investigate your fairness in EFL online assessment. Please, be informed that all the collected data is confidential and will only be used for research purposes. So, we will be grateful if you respond to the following sentences honestly as your answers will help teachers and educators to better understand the value of fairness in online assessment.

Thank you.

Name: …… (**Optional**)
Gender: Male: Female:
Teaching at: university/ school/ institute
Experience in teaching:…years
Degree: BA/ MA/ PhD in TEFL
BA/ MA/ PhD in English literature
BA/ MA/ PhD in English translation
BA/ MA/ PhD in non- English fields
E-mail address: …….. (**Optional**)

- I hereby confirm that I am informed about the voluntary basis of participation in this research project.
  Yes No
- I willing give the right to the researchers to use the information for only research purposes.
  Yes No
- I hereby express my consent to be contacted by the researchers in case needed concerning the information I provide below. Yes No

* Please circle ONE choice which best reflects your actions, preferences or opinion on the given statements.

1. Do your EFL online assessment items and tasks relate to your EFL classroom instruction?				
Very frequently	frequently	occasionally	rarely	never
2. On the basis of your experience, do you think that your EFL online assessments produce similar results when they are administered to the same people in different time and place conditions?				
Very frequently	frequently	occasionally	rarely	never
3. Do the items of your EFL online assessment measure the same subject matter, knowledge or skills domain?				
Very frequently	frequently	occasionally	rarely	never
4. Do you make sure that the scores your learners achieve are related to the subject area, knowledge or skills domain, or the syllabus behind the EFL assessment and not to other irrelevant factors?				
Very frequently	frequently	occasionally	rarely	never
5. Do your EFL learners have equal access to the content and tasks required for EFL online assessment?				
Very frequently	frequently	occasionally	rarely	never
6. Can all your learners afford the EFL online assessment financially?				
Very frequently	frequently	occasionally	rarely	never
7. Does your EFL online assessment offer accommodations to your students who have learning problems?(for examples, the students whose ability is low in learning language skills)				
Very frequently	frequently	occasionally	rarely	never

8. Are your learners equipped with facility, procedures, App literacy and conditions required for the EFL online assessment beforehand? (For example using a computer, mobile phone or tablet)				
Very frequently	frequently	occasionally	rarely	never
9. Do you administer your EFL online assessments in a consistent and similar manner with all of your learners? (For example assessment environment, equipment, supervision...)				
Very frequently	frequently	occasionally	rarely	never
10. Do you usually use similar language (in terms of vocabulary, grammar …) in your oral online assessment of all EFL earners?				
Very frequently	frequently	occasionally	rarely	never
11. Do you share your scoring criteria for the EFL online assessments clearly with your students?				
Very frequently	frequently	occasionally	rarely	never
12.Do you take any remedial measures if your learners lose access to the EFL online assessment during the assessment or tests?(For example, if the electricity cuts out , or the learners lose the internet connection)				
Very frequently	frequently	occasionally	rarely	never
13. Do you take any measures to identify cheating by your learners in the EFL online assessments?				
Very frequently	frequently	occasionally	rarely	never
14. Do you take any measures to reduce cheating by your learners during the EFL online assessments?				
Very frequently	frequently	occasionally	rarely	never
15. Do you make sure about the safety of different tools of your EFL online assessments regarding cheating?				
Very frequently	frequently	occasionally	rarely	never
16. Do you offer feedback to your students on their EFL online performance?				
Very frequently	frequently	occasionally	rarely	never
17. Do you consider your students' physical and psychological conditions such as stress and sicknesses while carrying out the EFL online assessments?				
Very frequently	frequently	occasionally	rarely	never
18. Do you try to establish a good rapport with your learners during your EFL online assessments?				
Very frequently	frequently	occasionally	rarely	never
19. Do you consider any share for activities like class activities, participation, presentations, in the final scores of your EFL online assessments?				
Very frequently	frequently	occasionally	rarely	never
20. Do teachers need to be involved in devising EFL online assessment items?				
Very frequently	frequently	occasionally	rarely	never
21. Do you explain the reasons behind the EFL online assessments to your students to reduce cheating?				
Very frequently	frequently	occasionally	rarely	never
22. Do you shorten the time for students to answer multiple-choice items?				
Very frequently	frequently	occasionally	rarely	never
23. Do you shuffle or randomize the order of responses for multiple - choice items?				
Very frequently	frequently	occasionally	rarely	never
24. Do you shuffle or randomize the order of questions for multiple-choice items?				
Very frequently	frequently	occasionally	rarely	never
25. Do you use online assessment during the course to for example get feedback on your own teaching style, or to identify the poorly learnt material so that you can improve them?				
Very frequently	frequently	occasionally	rarely	never

A: Add any comments concerning the clarity of the items in the questionnaire.

B: Add any comments related to the items in the questionnaire that you think is necessary.