EFL Teachers' Knowledge, Beliefs, and Practices Regarding Fairness and Justice in Technology-Enhanced Classroom Assessment: A Duoethnography

Teymour Rahmati [a,*], Musa Nushi [b]

[a] *Guilan University of Medical Sciences, Iran*

[b] *Shahid Beheshti University, Iran*

A B S T R A C T

Drawing on duoethnography, the teacher researchers in the present study interacted with the relevant literature, engaged in dialogs, and shared artifacts to examine their knowledge, beliefs, and practices regarding fairness and justice considerations in technology-enhanced language classroom assessment. Under the domain of knowledge, they conceptualized fairness and justice and identified their components. Within beliefs, the difference between high-stakes and low-stakes assessments, the significance of students' perceptions, and the role of computer literacy in relation to fairness and justice in technology-enhanced classroom assessment were debated. To operationalize their knowledge and beliefs, the researchers inspected their assessment practices during and following COVID-19. They agreed that fairness was distinct from justice in that the former pertained to test internal characteristics and its administration procedures while the latter referred to test external consequences at a broader social level. They believed that fairness and justice were equally important in high-stakes and low-stakes assessments, and students' perceptions were valuable sources of feedback regarding fair and just classroom assessments. Moreover, the teachers argued that computer literacy cannot yet be considered an aspect of language ability. Finally, it was revealed that although their practice regarding fairness and justice was affected by the pandemic, they learned valuable lessons (e.g., combining online and paper assessment modalities and giving oral exams) in this respect for the future. The findings imply that language teachers should theoretically adopt a clear conception of fairness and justice while being practically prepared for future developments (e.g., technological advances) and unexpected circumstances (e.g., a pandemic).

Keywords: classroom assessment; duoethnography; EFL teachers' beliefs; EFL teachers' knowledge; fairness; social justice; technology-enhanced language assessment

© Urmia University Press

* Corresponding author: English Department, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran
Email address: trahmati@gums.ac.ir

Introduction

Although closely associated with educational practices, assessment is often imbued with underlying social, political, and economic implications since interpretations of assessment outcomes inform decisions as to, for instance, who should enter a program or hold a position from among a large number of potentially qualified applicants. Brunfaut (2023) illustrates this by briefly referring to the case of an airline that administered a test in only one of the eleven official languages of a country "effectively rendering it a second language reading and writing proficiency test for the majority of the country's nationals" (p. 21). Thus, language assessment experts have been invariably concerned with fairness and social justice considerations of assessment insofar as the preoccupation with accountability led to the dawn of psychometric era or Assessment of Learning (AoL) in the second half of the twentieth century. This era was characterized by rigorous scoring methods and decontextualized measurement of language knowledge through tests which were more psychometrics-driven than construct-directed. Criticisms leveled against AoL gathered momentum when the need was identified for ongoing and contextualized assessment procedures such as Assessment for Learning (AfL) and Assessment as Learning (AaL) (Webb & Ifenthaler, 2018) as well as learning-oriented language assessment (Gebril, 2021). Although these developments were effective in assessing language in the context of use, they brought the issues of fairness and social justice back to the spotlight. Simultaneous with these sometimes pendular shifts from accountability at one end to assessment depth and real-world use at the other, technological innovations were increasingly introduced to language assessment often as a panacea for assessment predicaments. In their introduction to an AaL, AfL, and AoL integrated model, Sadeghi and Rahmati (2017, p. 51) anticipated that the field of language assessment was about "to witness a fundamental turn" rooted in advances in computer technology and the ensuing developments.

The move toward that fundamental turn was expedited by the outbreak of COVID-19 at about the end of 2019. Heavy reliance on technology to administer tests during the stressful conditions of a health crisis foregrounded terms such as *technology-assisted* (Cohen et al., 2023), *technology mediated* (Sadeghi & Douglas, 2023), and *technology-enhanced* (Alavi et al., 2022) in language assessment circles. Prior to the pandemic, technology was utilized in language assessment as an option to capitalize on its efficiency and to streamline the testing process (Chapelle & Voss, 2017). Major large-scale exams such as the International English Language Testing System (IELTS), the Internet-Based Test of English as a Foreign Language (TOEFL-iBT), and Pearson Test of English began to be delivered online in 2000, 2005, and 2008, respectively. The transition from paper-based delivery to online mode, however, was always made prudently by maintaining both modalities available and upgraded. Yet, when the pandemic unfolded in 2020, migration to online assessment became an obligation rather than an option. The urgency to adopt online assessment provoked by the pandemic brought about such radical changes that assessment providers did not have the opportunity to stop and ponder on fairness and social justice considerations. Following an interim suspension of assessment practices as an immediate reaction to the pandemic (Butler-Henderson & Crawford, 2020), technology was finally resorted to in order to meet assessment needs although the conditions were "anything but normal" (Ockey et al., 2021, p. 26). Transition from paper-based tests to technology-delivered versions (e.g., Purpura et al., 2021) was often made with intuitive considerations of fairness, justice, reliability, and validity.

Recently, however, attempts have been made to look back at language assessment practices during the pandemic. This retrospective view is highly valuable as there are lessons to be learned for the future mainly due to at least two reasons. First, with advances in Natural Language Processing (NLP) and the use of Artificial Intelligence (AI) in automated item generation (Attali et al., 2022; Hommel et al., 2022; von Davier, 2020) and automated scoring of performance (Huawei & Aryadoust, 2023), educational assessment in general and language assessment in particular are

awaiting another fundamental turn known as AI-enhanced (Huggins-Manley et al., 2022) or Machine Learning (ML)-enhanced (Tay et al., 2022) assessment. Next, accelerated by an interest in retaining some of the assessment developments which have resulted from a painful struggle with the pandemic (Copeland, 2021; St-Onge et al., 2022), the historical and cautious transition to online modality has resumed following the abatement of the pandemic.

The present study focuses on English as a Foreign Language (EFL) teachers' knowledge, beliefs, and practices regarding fairness and justice as two under-researched areas in technology-enhanced language assessment (TELA). Drawing on Malone's (2013) description of language assessment knowledge as "language teachers' familiarity with testing definitions" (p. 329), knowledge of fairness and justice is operationalized here as EFL teachers' conceptualizations of these concepts and their components, particularly in light of relevant abstractions in the literature. Beliefs are a set of propositions anchored in knowledge and refined by experience which serve as a frame of reference for making decisions (Larenas et al., 2015). For the purposes of the present study, beliefs are operationally defined as EFL teachers' stance regarding issues (e.g., the role of computer literacy) pertaining to fairness and justice in TELA. Practices are also defined as discernable manifestations of teachers' implicit cognitive processes in real teaching context (Lan & Lam, 2020). Thus, practice refers to concrete measures taken by EFL teachers to materialize their underlying knowledge of and beliefs about abstract concepts in a particular context. More specifically, practices in the current research are EFL teachers' activities to observe fairness and justice in their technology-enhanced language classroom assessments.

As to methodology, a duoethnographic approach has been employed to conduct the present research. As a technique espoused with qualitative research strand, duoethnography is a recent methodological innovation closely associated with meta-research era (Amini Farsani et al., 2021) in applied linguistics research, which deals directly with what we know and do. It provides teachers with the opportunity to engage in research (see Amini Farsani & Babaii, 2019) and examine their knowledge, beliefs, and practices through dialogs. Further details on the objectives of the study and duoethnography are presented under Literature Review and Method sections, respectively.

Conceptual Framework

An attempt to establish the relationship between validity and fairness and to conceptualize fairness and justice in language assessment are by far the most recurrent themes in the relevant literature. Three different approaches to the relationship between fairness and validity are *fairness and validity as independent* though inter-related and overlapping test qualities (Educational Testing Service (ETS), 2022; Kane, 2010; Walters, 2012), *fairness as an overarching quality* encompassing validity among other qualities (Kunnan, 2010), and *validity as a comprehensive test quality* subsuming fairness among other qualities (Xi, 2010). It is within the last approach that McNamara and Ryan (2011) make a distinction between fairness and justice.

While earlier researches (e.g., Hamp-Lyons, 1997) often used the terms fairness and justice interchangeably or subsumed justice as a component of fairness (e.g., Kunnan, 2000), McNamara and Ryan (2011) have defined fairness as an internal quality of the test which is concerned with equal treatment of all test-takers and tries to be free from psychometric bias. Justice, on the other hand, questions the social uses of a test in the first place and pertains to its consequences and implicit values. They illustrated this distinction by reviewing the Australian Citizenship Test, which was criticized from many grounds, including language assessment experts who believed that it was unfair as the level was far above 'basic' specified by the legislation. Other organizations perceived the test as unjust since it was discriminatory and violated human rights. The revised version of the test specified A1 to A2 English proficiency of Common European Framework of

Reference for Languages (CEFR) as the required English proficiency, but the implicit political motivations remained unchanged. Thus, the use of the test was disputed due to its social consequences. Following McNamara and Ryan (2011), Karami (2013), Deygers (2017), and Hamid et al., (2019) have also adopted a test-internal property versus test-external policies and values view of fairness and justice, respectively. Wallace (2018) and Wallace and Qin (2021) have identified the social entity that administers an assessment (i.e., the language program) as another component of social justice in addition to its social use and consequences. They, however, connected this to fairness and stated that if test-takers perceived a test as fair, then, the language program administering the test was just.

Given the significance attached to the relationship between validity and fairness in assessment literature and the distinction between fairness and justice as opposed to the interchangeability of the two concepts, the discussion above serves as the theoretical framework for the present study. In examining knowledge, the study investigates how EFL teachers define fairness and justice. Under the domain of beliefs, EFL teachers' stance on issues such as students' perceptions of fair and just assessment and how computer literacy affects fairness and justice in TELA are investigated. While inspecting practice, the study attempts to explore how EFL teachers' theoretical knowledge of and personal beliefs about fairness and justice are manifested in their technology-enhanced classroom assessment practices.

Literature Review

Most research studies on fairness and justice draw on the conceptual framework presented above. The reference to psychometric bias in McNamara and Ryan's (2011) conceptualization of fairness implies that it can be measured objectively through, for example, Rasch measurement conducted by computer software such as FACETS. Nevertheless, justice issues cannot be subjected to objective measurement techniques (McNamara et al., 2019). Wallace (2018) rightly observed that fairness could be measured both objectively through statistical procedures and subjectively by examining stakeholders', including students', perceptions of fairness in a given assessment event, but social justice could only be appraised subjectively. This observation confirmed the distinction between fairness and justice as test internal and test external qualities, respectively, while identifying a new source (e.g., students' perceptions) of feedback on assessment fairness. The way EFL teachers' view this observation greatly affects their knowledge, beliefs, and practices with respect to classroom assessment.

Accordingly, a line of research in both high-stakes exams and classroom assessment is surveying students' perceptions of assessment fairness and justice. Test-takers' perspectives of assessment events have a significant bearing on their learning and behavior (Rasooli et al., 2019). Regarding international exams, Puspawati (2014) reported that TOEFL-iBT candidates had negative feelings about the test due to its environment, procedures, and equipment. Moreover, the participants expressed that the test was unfair and unjust mainly because of the subject-specific and culturally loaded topics used in the test as well as the computer skills and speediness required to complete its tasks. For the TOEFL test to become fair, the participants suggested that the test candidates should be given a number of topics to choose from and enough time to show their true language ability. For a similar non-TELA study on the IELTS test, interested readers may see Hamid et al. (2019).

Moving away from high-stakes to classroom assessment, Copeland (2021) explored students' perceptions of fairness in grading video presentations in online language assessment in Korea. Around 90% of the students believed that grading of their video presentations by their teachers was fair since clear and understandable rubrics were designed and communicated to the students,

hard work led to a high score, the conditions were the same for all students, they got the opportunity to submit their performance, and had enough preparation time. Nevertheless, some students believed that the assessment practice was not fair since video presentations were affected by technological factors like editing skills and network problems, and some of their peers cheated by hiding the script somewhere around. Evidently, the findings emphasized prior specification and release of assessment criteria while identifying the threat to academic integrity and the negative impact of computer literacy on learners' performance as the drawbacks of TELA. The implications for EFL teachers' knowledge, beliefs, and practices were the need to adjust their conceptualizations of the relevant concepts and decide how computer literacy had to be treated as well as how test security could be enhanced in online assessment of achievement.

Obviously, teachers' perceptions of fair and just assessments are significant in administering tests which are later perceived as fair and just by their students. According to Azizi (2022), Iranian university English teachers believed that fairness in online assessment encompassed distributive justice (e.g., match between assessment practices and students' needs), procedural justice (the importance of students' voice, consistency, flexibility, and transparency), and interactional justice (respectful and caring teacher-student relationship and observing informational justice). As evident, informational justice was subsumed under interactional justice in this study and the words fairness and justice were used interchangeably. A non-TELA study by Beheshti and Ahmadi Safa (2023) also outlined the role of variable contexts in forming the constituents of fairness. These findings pinpoint the components of assessment fairness which, in turn, signify the need for relevant adjustments in EFL teachers' knowledge, beliefs, and practices.

As assessment providers, Neiriz et al. (2023) described the modifications they made to Iowa State University English Placement Test during COVID-19 and the resultant fairness issues. The modification of the writing test caused a number of challenges, including selection of a time zone which was not too early or too late for test-takers from different geographic regions, monitoring for cheating, computer system requirements, and stable internet connections. A number of fairness issues were also raised, especially in terms of construct representation (completing a timed writing task under the surveillance of a webcam and variable environmental factors such as lighting and noise which were kept uniform in previous administrations of the test), test security (although the test was proctored using technology, it was time-consuming and almost impossible to review the recordings for cheating), and test conditions (e.g., [lack of] access and familiarity with the required technology and taking the test while maintaining a certain physical posture). As to speaking, the test was first administered outdoors and then using live video conferencing through WebEx. Regarding construct representation, outdoor administration of the spoken test raised fairness consideration due to distracting noise. Live conferencing also entailed interaction with an examiner instead of a peer for some part of the test, which caused extra anxiety. With respect to security, the outdoor administration meant that the prompt could possibly be heard by those examinees who were standing nearby while this was controlled in live conferencing. The study revealed that the outbreak of the pandemic further confounded issues such as computer literacy, technological requirements, and academic integrity in TELA. This, in turn, entailed that EFL teachers should constantly engage in refining their knowledge, beliefs, and practices regarding the use of technology in classroom assessment.

The studies reviewed above may conjure up the thought that TELA inherently brings about negative effects in terms of fairness and social consequences. The final two studies from the context of China reported here, however, provide a contradictory image. In the first study, Jin and Yan (2017) found that students who took the online College English Test (CET) obtained higher scores on the writing section, which was attributed to the elimination of the negative impact of bad handwriting on raters' scoring. In the second study, Jin (2022) showed that the online administration of CET benefited the Chinese educational system since it proved to be a rich source of extrinsic motivation for teachers and learners, assisted alignment among involving

organizations, offered criterion and norm referenced information for decision-makers as well as information regarding variable performance by students from economically different regions, and served symbolic values in the form of the need to get a CET certificate for graduation. Given that CET was computerized in the mid-2000s, the results of these two studies imply that the cultivation of positive aspects of TELA and smooth transition to online modality require careful planning and consideration of myriad factors as well as coordination among involving organizations.

The review of the literature presented here indicates that researching language teachers' knowledge, beliefs, and practices with respect to fair and just assessment warrants due attention. In classroom assessment, which affects almost all students, teachers are responsible for providing their learners with fair assessment and alleviating the negative consequences of their tests. Hence, the present study used a duoethnographic approach to examine EFL teachers' knowledge, beliefs, and practices regarding fairness and justice in classroom assessment with a particular focus on TELA during and following the COVID-19 pandemic. One of the main objectives of the study was to delve into teachers' knowledge and inspect how they conceptualized fairness and justice and the relevant components of these concepts. Another aim was to elicit teachers' beliefs about the significance of fairness and justice in classroom assessment, the value of students' perceptions of fair and just assessment, and the role of computer literacy in relation to fairness and justice in TELA. Regarding practices, the study attempted to investigate teachers' day to day measures to run fair and just online assessments during the pandemic and to document the lessons they learned for the future. Accordingly, the present research was guided by the following questions:

1) How do EFL teachers conceptualize fairness and justice in TELA in relation to classroom assessment?

2) What are EFL teachers' beliefs about fairness and justice in classroom assessment, students' perceptions of fair and just assessment, and the role of computer literacy in TELA and in relation to classroom assessment?

3) How are EFL teachers' knowledge and beliefs about fairness and justice in TELA manifested in their classroom assessment practices?

Method

The study can be framed within assessment literacy as it examines knowledge of fair assessment (Xu & Brown, 2016), language teacher cognition research since it deals with what teachers know, believe, and do (Borg, 2003), constructivist paradigm because meaning is co-constructed here through dialogs (Lowe & Kiczkowiak, 2016), and reflective practice due to teachers' deliberations of their artifacts and classroom practices (Banegas & del Pozo Beamud, 2022).

Duoethnography

The present study deploys duoethnography to attain its objectives. As an emerging and promising qualitative research method, duoethnography provides teachers with the opportunity to inspect their own beliefs and practices. Duoethnography is a combination of two autoethnographies (Banegas & del Pozo Beamud, 2022) in which two teachers form the site of their research (Karas & Uchihara, 2021) and "utilize dialogue to juxtapose their individual life histories in order to come to new understandings of the world" (Lawrence & Lowe, 2020, p. 9). Written and/or face-to-face dialogs are the primary sources of data in duoethnography although teachers' artifacts (Huang & Karas, 2020) and literature can also be drawn upon. Of course, duoethnographic dialogs, unlike everyday dialogs, are re-examined and refined to convey meaning accurately. In

language education, duoethnography has been used to investigate a range of topics, including native-speakerism (Lowe & Kiczkowiak, 2016), research-informed innovations (Rose & Montakantiwong, 2018), co-authoring research papers (Yeo & Lewis, 2019), critical language teacher education (Banegas & Gerlach, 2021), language teachers' perspectives of silence (Karas & Uchihara, 2021), postmemory and multilingual identities in English language teaching (Ahmed & Morgan, 2021), and teacher encounters with learner spelling errors (Jing & Reynolds, 2022). The main rationale behind the selection of duoethnography for the purposes of the present study was the assumption that it could accurately present insiders' meaning on sensitive issues such as knowledge and beliefs. Moreover, duoethnography requires deep interaction with literature and peers or mentors, which leads to professional learning as a result of doing research. The key to success in duoethnographic research is that the two teachers should be critical friends (Lawrence & Lowe, 2020) who trust each other and are willing to share ideas. In light of this, we present short biographies of ourselves, below, descriptive enough for the purposes of the current research.

Participants

Teymour is an assistant professor of applied linguistics at Guilan University of Medical Sciences, Rasht, Iran. He teaches general English and medical English to students of medical sciences and has the experience of teaching courses such as *Essay Writing*, *Language Teaching Methodology*, *Research Methodology*, and *Language Testing* to English major students. His main research interests are language assessment, language teacher motivation, and professional development, and he has published papers in these areas. Teymour is an advocate of technology integration into language teaching and assessment; he delivered almost all of his classroom assessments through technology during the COVID-19 pandemic.

Musa is an associate professor of applied linguistics at Shahid Beheshti University, Tehran, Iran. He specializes in teaching various subjects related to language and linguistics at both the undergraduate and graduate levels. At the undergraduate level, Musa teaches courses such as *Academic Writing, Linguistics,* and *Phonetics and Phonology.* At the graduate level, Musa offers courses in *Psycholinguistics, SLA Theories, and Technology in Second Language Teaching and Learning.* Musa's research interests revolve around the intersection of instruction and L2 learning, with a particular focus on the roles of technology and personality traits.

Teymour and Musa were classmates and close friends during their Bachelor of Arts studies. Following graduation, they were admitted to different universities to pursue their Master of Arts and Doctor of Philosophy degrees. In 2018, they reunited at a conference on language teaching and literature at a university in northern Iran. During the stringent conditions of health crisis, they were each engrossed in making COVID-19 modifications in their practices and keeping up with their teaching responsibilities. Since 2022, Teymour and Musa have been involved in writing joint research papers. They have a respectful relationship and have always been willing to learn from each other.

Data Collection and Analysis

For the purposes of the present study, Teymour and Musa interacted with the relevant literature to identify the key concepts investigated by previous research. As a result of this stage, a number of themes emerged at each domain. Under the domain of knowledge, conceptualization of fairness and justice as well as their components in TELA and in relation to classroom assessment were found to be significant. These were, then, formulated into the following key questions:

- *How do you conceptualize fairness in TELA in relation to classroom assessment?*

- *What are the components of fairness in TELA and its application in classroom assessment?*

- *How do you conceptualize social justice in TELA in relation to classroom assessment?*

- *What are the components of social justice in TELA and its application in classroom assessment?*

- Similarly, significant themes under the domain of beliefs were used to pose the following questions:

- *Do you believe that fairness and justice considerations are more relevant to high-stakes exams than classroom assessment?*

- *Do you believe that students' perceptions are reliable sources of information regarding how fair and just a classroom assessment event is?*

- *Do you believe that computer literacy should be regarded as part of language ability?*

- As to practice, the following questions were derived from our personal experiences of administering TELA during the pandemic:

- *Did you experience any fairness and justice issues in your technology-enhanced classroom assessment during COVID-19?*

- *Did you receive any feedback regarding your classroom assessment practices during COVID-19?*

- *What measures did you take to administer fair and just classroom assessment during COVID-19?*

- *What modifications have you made in your technology-enhanced classroom assessment practices following the COVID-19 pandemic?*

The above questions were developed form the sources reviewed above in the Literature Review section. For validation purposes, the questions were reviewed by two expert colleagues, whose comments resulted in further refinement of the questions. In our original questions, for instance, "TELA" was missing as we had taken it for granted, but our colleagues insisted on its inclusion to reiterate the focus of the study.

Subsequently, we engaged in eleven conversations over a course of two months to answer the above questions. Some conversations, particularly the ones on beliefs and practices, lasted around two hours. Additionally, we shared Word documents and exchanged artifacts during the course of this study. Regarding data analysis, it should be noted that our dialogs were theme-based by nature as they focused on the specific questions posed in the domains of knowledge, beliefs, and practices.

Findings

In this section, we present the dialogs we engaged in to answer our specific questions under the domains of knowledge, beliefs, and practices. It should be noted that delineating the boundary between knowledge and beliefs was not easy to establish. For the purposes of the current study, we decided to classify questions pertaining to conceptualizations and components of fairness and justice under the domain of knowledge while the views held toward assessment issues rooted in EFL teachers' personal experiences were grouped under beliefs.

Knowledge

Due to the significance attached to the conceptualization of fairness and justice in the relevant literature, the first question we answered was: *How do you conceptualize fairness in TELA in relation to classroom assessment?*

> **Teymour:** *To conceptualize fairness, I regard it as an independent test quality distinct from validity since subsuming fairness under validity renders it secondary to validity while fairness issues require due attention in language assessment. Hence, I broadly define fair language assessment as an assessment practice which is free from all forms of bias. Expanding this conceptualization to TELA, I would like to add computer literacy as a form of bias. As to classroom assessment, a test should focus on the objectives of a given course. Therefore, a fair technology-enhanced language classroom assessment is one which is construct-representative and free from any form of bias (differential item functioning), including gender, ethnicity, culture, socioeconomic status, and computer literacy.*

> **Musa:** *Fairness is an attribute of a language assessment event at the level of design, administration, and interpretation of test results, which ensures all students have equal opportunities to demonstrate their knowledge and skills. Fairness encompasses more than just the (content) validity of the test. When it comes to TELA, fairness means that the technological tools and platforms that are incorporated into language assessment do not introduce discrimination or disadvantage for test-takers based on their personal characteristics such as race, gender, socioeconomic status, or cultural background. For instance, lack of access to certain technology can present students from low-income families with a challenge or disadvantage. A fair technology-enhanced language classroom assessment is, therefore, one in which technological aids are employed to conduct an unbiased and equitable evaluation of learners' language proficiency.*

The second question we focused on was: *What are the components of fairness in TELA and its application in classroom assessment?* In addition to engaging in dialogs, we interacted with the literature to answer this question. As shown by the above literature review, the four components of classroom assessment were distributive, interactional, informational, and procedural fairness. Of course, most previous studies (e.g., Wallace, 2018) subsumed informational fairness under interactional fairness.

> **Teymour:** *Fairness in classroom assessment begins with test design. Care should be exercised in the design of a classroom test so that it revolves around the skills practiced in the classroom and the content focused on during instruction. Additionally, test items should not advantage or disadvantage any individual because of construct-irrelevant factors such as cultural background and the speed with which someone can work with computers. Test design features are directly related to distributive fairness defined as both teachers' and students' observations that assessment outcomes reflect test-takers' true ability. Following test design, test administration is a significant issue to consider. Sharing information regarding test time and date, item types, the weight attributed to each section as well as notification of scoring rubrics, performance standards, and technological requirements prior to test administration form the essential components of informational fairness while caring for and respectful treating of every test-taker prior, during, and after test administration comprise the interactional fairness of a classroom assessment. Furthermore, employing unanimous scoring strategies, timely distribution of the results, and adopting uniform test score interpretation procedures form the procedural fairness of classroom assessment. In TELA, extra attention should be paid to sharing information on technological requirements and taking care of test security.*

> **Musa:** *Fairness is a characteristic of a language assessment event at the level of design, administration, and interpretation of test results. In the design phase of TELA, fairness involves selecting or creating technology-aided test materials and tasks that are aligned with the learning objectives of the course. This entails meticulous review of the content of the test materials and tasks selected or created to ensure that they are not tainted by construct-irrelevant factors such as race, gender, linguistic or cultural background. During the administration phase, fairness requires following standardized procedures and protocols to guarantee consistent test conditions for all test-takers. Furthermore, fairness should be observed in the interpretation of test outcomes, and this entails employing proper statistical techniques and equating procedures to ensure that scores accurately demonstrate test-takers' knowledge and skills.*

Following on, we tried to answer the question of: *How do you conceptualize social justice in TELA in relation to classroom assessment?*

> **Teymour:** *First of all, I want to reiterate the distinction between fairness as a test internal characteristic and its administration specification, on the one hand, and social justice as the consequences of a test for individual test-takers and the society as a whole, on the other. Social justice in TELA and its application in classroom assessment can be defined as no harm caused to individual test-takers and their families by an assessment event. An important consequence of TELA in classroom assessment is the increased anxiety experienced by students and their families regarding smooth running of an online assessment event.*

> **Musa:** *Social justice in TELA includes but goes beyond fairness and relates to the realization that language assessments may have significant implications for test-takers' current or future educational and professional opportunities. The relation between social justice in TELA and classroom assessment lies in the alignment of principles and goals of different stakeholders, including teachers, researchers, policymakers, parents, representatives from marginalized communities, and test-takers themselves, in the design, administration, and interpretation of language assessments.*

We, then, addressed the question of: *What are the components of social justice in TELA and its application in classroom assessment?*

> **Teymour:** *The answer to this question is highly context specific. For instance, in the context of Iran and at the level of tertiary education, some regions did not have a reliable power supply while other regions did not have a stable internet connection during the pandemic. Anxiety and cost were the important consequences of these challenges as students and their parents had to provide both Wi-Fi and mobile data connections to use the most effective one in a given circumstance. At an individual level, access to the required equipment was another significant social justice consideration, which brought about even further anxiety. Another noticeable consideration in TELA is academic dishonesty. In conditions where test security measures are compromised, some students may have access to test items prior to test administration. A serious consequence of this phenomenon is score pollution and the ensuing implications regarding having access to opportunities for higher education studies and positions. Thus, anxiety, cost, and academic dishonesty are the most notable examples of social consequences of TELA at classroom level in the context of Iran.*

> **Musa:** *I regard equitable, reliable access to assessment technologies as the most crucial aspect of a socially just TELA. The issue of 'digital divide' is in fact a major problem in contexts such as Iran where equitable access to devices, secure/stable internet connection, and necessary software are not guaranteed for many, especially for those who may face financial barriers. A related problem is that many websites are not accessible without a VPN and this in itself creates negative emotions in students. Digital literacy is another important component, so teachers or administrators should provide training and resources to ensure students have the necessary skills to navigate and use assessment technologies effectively. Third is transparency in terms of assessment expectations, procedures, and grading criteria. Students should have a clear understanding of how their assessments will be conducted and evaluated using technology.*

The dialogs above show that the teachers make a distinction between fairness and justice in that the former is a test quality rooted in design, administration, and interpretation procedures whereas the latter pertains to the consequences of an assessment at social level. As to the components of fairness, the teachers highlight the importance of distributive, interactional, informational, and procedural fairness. Regarding justice, the major components are computer literacy, availability of required technology and infrastructures (e.g., a stable internet connectivity), cost, and anxiety.

Beliefs

Language teachers' beliefs form an important part of their cognition (Borg, 2003). Mimirinis (2019) argues that teachers' decisions about utilizing technology are informed by their beliefs regarding the benefits of technology use, its facilitative role in assessment practices, and its effects on learners' engagement with assessment and feedback. Below, we present our beliefs about

fairness and justice in TELA with a particular focus on classroom assessment. The first relevant question we contemplated on was: *Do you believe that fairness and justice considerations are more relevant to high-stakes exams than classroom assessment?*

> **Teymour:** *Although unfairness and social consequences cause wide range effects through high-stakes assessments and the political, hegemonic, and economic implications are more conspicuous in high-stakes tests than classroom assessment practices, I believe that fairness and justice issues in so-called low-stakes classroom tests deserve equal attention. This is a critical measure to observe individuals' rights and reduce the negative academic and behavioral effects of unfair and unjust assessment on our learners. In utilizing TELA for classroom purposes, fairness and justice considerations gain even more significance, particularly due to issues of access to technology and computer literacy.*

> **Musa:** *No, I think fairness and social justice considerations are important in both high-stakes exams and classroom assessments. While the impact of assessment outcomes can have significant ramifications for individuals' educational or professional future (for instance for those planning to migrate to another country), fairness and social justice are highly relevant in classroom assessments, even though the impact may be more localized. For the latter, one may think of a learner whose motivation and emotions are negatively impacted by inappropriate classroom assessment procedures such as vague scoring or lack of feedback. Such a bad taste may linger in the learner's memory and negatively affect his or her subsequent language learning experiences.*

Regarding the significant role of students' perceptions, the next question we investigated was: *Do you believe that students' perceptions are reliable sources of information regarding how fair and just a classroom assessment event is?*

> **Teymour:** *My answer to this question is absolutely positive. I do believe that students can provide teachers with valuable feedback on assessment practices. This is particularly so with respect to informational, interactional, and procedural fairness. In my opinion, the only issue which needs careful examination is students' perceptions of distributive fairness or how assessment outcomes represent students' true ability. I have examples of students in my mind with a high score on a test though with the perception that the score was below their proficiency level and vice versa. Broadly speaking, students often believe that assessments lack distributive fairness unless sufficient and accessible feedback is offered as to their performance on the test.*

> *Students' perceptions of the consequences of classroom assessment practices also help teachers in alleviating the problems with social justice considerations of their tests. This is of paramount importance in terms of briefing teachers on the range of students' access to computers, stable internet connections, and computer literacy.*

> **Musa:** *Yes, students' first-hand experiences can offer unique perspectives on various aspects of the assessment process. Their feedback can shed light on potential biases, inconsistencies, or areas where improvements can be made to promote fairness and justice. Nevertheless, it is important to recognize that students' viewpoints alone may not be the sole determinant of the fairness and justice of an assessment event. A range of factors, including established assessment principles, validity evidence, and the perspectives of other stakeholders, such as teachers and assessment experts, need to be considered to ensure a thorough assessment of fairness and justice.*

The issue of computer literacy, which was referred to above and singled out in the literature, was what we attended to next: *Do you believe that computer literacy should be regarded as part of language ability?*

> **Teymour:** *The answer to this question is highly context specific and has a significant bearing on a teacher's practices regarding their use of TELA for classroom assessment. I believe that the ability to use computers is a continuum with computer familiarity at one end, levels of computer intimacy spanning in the mid-point, and computer literacy at the other end. By computer familiarity, I mean the basic knowledge of computers and how they are used for simple tasks like typing. Computer familiarity comes from having computers around somebody though not necessarily using them routinely. Computer intimacy is a state in which a person is quite skillful in using a particular computer such as a personal laptop. This level of skill may not be carried over onto other computers. Computer literacy, on the other hand, is the ability to demonstrate knowledge of computer use across tasks and computers. In the context of Iran and in my own personal experience, most students can be classified as being computer-familiar or having intimate knowledge of a personal computer use. Based on my personal experience during*

the pandemic, I believe that computer literacy cannot form part of the construct of language ability, at least for the time being. Of course, the landscape will hopefully change in the future.

Musa: *In today's increasingly digital world, much communication happens through various digital platforms and channels. Such developments require new forms of literacy and computer literacy should be seen an essential component of ability to communicate in L2. Of course, I realize computer literacy and language ability are distinct skills, yet they are becoming increasingly interconnected in today's technology-dominated global communication. Within our field, there are proposals such as 'Technological Communicative Competence' which indicate that the ability to use computers and related technology efficiently is being regarded as a complementary skill to language ability. I hasten to add that I prefer 'digital literacy' to 'computer literacy' as the former is a more inclusive term covering not only the ability to utilize various digital devices, software applications, and online platforms efficiently but also the skills, knowledge, and attitudes required "to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital technologies for employment, decent jobs, and entrepreneurship" as stipulated by UNESCO (2018).*

The dialogs above reflect how personal experience and contextual factors bring about modifications in teachers' knowledge to form their beliefs. This process of modification does not stop here, however. In fact, beliefs mediate between knowledge and practice and are further modified in light of the realities of the context of practice, which we turn to next.

Practices

To operationalize our conceptualizations of fairness and justice in technology-enhanced language classroom assessment, we inspected our practices. In addition to dialogs, our artifacts served as rich sources of data in this part of the study. The first question we answered here was: *Did you experience any fairness and justice issues in your technology-enhanced classroom assessment during COVID-19?*

Teymour: *Actually, yes. It was a really difficult situation which entailed radical changes in our assessment practices. We had to design tests based on the requirements of the test delivery system run by the university IT center. At the end of the semester, we had to design tests which could be completed in 60 minutes because there were other tests in line scheduled for the day. This meant a lot of pressure on language teachers who had to make modifications in their assessment tasks, particularly writing tasks. For general courses such as English, students took tests at home without being proctored. There was no way whatsoever to ensure academic integrity.*

Musa: *Yes, I did. The first major challenge was how to prevent cheating. To minimize chances of cheating, I had to administer most of my classroom exams orally with the students' webcams turned on, which potentially impacted students' emotional well-being and performance during assessments. Aside from threat to validity and fairness, oral assessment also took an unconscionable amount of time from both teachers and students. Another major issue was the unequal access to resources among students from disadvantaged backgrounds. Moreover, not all students, and this sometimes included me, were equally prepared or possessed the necessary digital literacy skills, which impacted effective administration and engagement in technology-enhanced assessments. Lastly, the absence of visual cues during assessments posed an additional obstacle. Due to slow internet speed, we often had to turn off our device camera. The limitations made it difficult to interact effectively during assessment as everything was based on verbal cues. They also made the assessment process tedious and boring for students.*

The next question we inspected was: *Did you receive any feedback regarding your classroom assessment practices during COVID-19?*

Teymour: *Immediately after a test, some students used to call me and complain about the limited time available to complete the tasks. Most students said that they had limited computer skills and were very slow in typing. There were others who said that they could not take the test on time due to a problem with their internet connection. Some students also expressed that they were not accustomed to brainstorming and recollecting ideas at a computer, and they had to spend extra time on writing their answers on a piece of paper and then keying their answers in. This strategy often did not leave them enough time to submit their answers to writing tasks. On the other hand, there were some students who performed at a level well above their classroom performance. The IT center staff used to shuffle items and options, and they sometimes did so for English tests. This meant that sometimes the items on a reading*

comprehension test appeared before the relevant text, and students had to spend extra effort to relocate reading texts and their relevant items in order to be able to answer them correctly.

Musa: *Teaching in Iran presented unique challenges due to the country's inadequate technological infrastructure, particularly the slow and restricted internet access. The primary concern expressed by students was their anxiety over potential disruptions in internet connectivity, which could hinder their ability to demonstrate their true ability. This issue was particularly problematic when evaluating their speaking skills. Second, was their persistent complaints about the nature of tasks assigned and the insufficient time allocated for completion of the tasks, especially for those with limited computer skills. Additionally, students sometimes expressed their concerns about scoring specifications and how the technical glitches affected the way their performance would be assessed.*

Given the chaotic situation described above, we then present the question of: *What measures did you take to administer fair and just classroom assessment during COVID-19?*

Teymour: *We knew that we had to be flexible and realistic rather than idealistic. Health was the first priority with managing anxiety during a highly stressful condition following next. In my own personal practice, the first question I asked myself was: Should I give quizzes and mid-term tests? The answer was: No. The next strategy, I adopted was to reduce the weight of final exams and place greater emphasis on online class participation. I allowed those students who had not been able to key in their answers to writing tasks to send me images of their hand-written answers through social networking sites. Sometimes, I allowed those who had not been able to take a test because of issues with their internet connection to sit for a test on the same course at a different department. The main aim was to keep assessment practices alive.*

Musa: *Prior to the assessment, I took proactive steps to support students. I provided them with sample materials and tasks that closely aligned with the content and format of the assessment. Also, I actively encouraged them to share any apprehensions they had about the assessment and tried to address their worries as much as I could. During the written exams, I adopted a flexible approach to accommodate students' unique circumstances. For instance, I did not limit the platforms through which they could submit their answer sheets; I accepted responses through various channels, including popular social media platforms like WhatsApp or Telegram. Recognizing the possibility of unexpected technical difficulties or interruptions during online exams, I allowed students a second chance if they experienced connectivity issues. In addition, since online assessments required a transition from traditional paper-and-pencil exams to digital formats, I gave students slightly more time to finish the tasks.*

Finally, we aimed at answering the important question of: *What modifications have you made in your technology-enhanced classroom assessment practices following the COVID-19 pandemic?*

Teymour: *I have made significant modifications in my practice, and I hope I can manage to recall them here. Now that the pandemic has subsided, I am trying to improve the quality of my assessment practices by focusing on all test qualities. I have also tried to consult professional textbooks and journal papers on TELA to update myself in this regard. As to my post-pandemic assessment practices, I notify students of test time and date and other relevant information prior to the test as a step toward informational fairness. I design test tasks and items within the topics which are immediately relevant to my students' major and experience. Figure 1 shows part of a reading comprehension text as part of a test designed for General English for the Students of Medicine course. This text was used to control differential item functioning rooted in students' personal interests.*

Like all individuals, medical students also need to do regular exercise. Medical students often spend long hours sitting in lecture halls or studying at desks. This ***sedentary*** lifestyle can lead to health problems such as obesity, high blood pressure, and diabetes. However, regular exercise can help medical students to maintain a healthy weight, reduce stress levels, and improve their overall physical and mental health. Exercise can also increase energy levels and improve cognitive function, which can be beneficial during exams and clinical rotations. Therefore, it is important for medical students to make time for regular exercise, even if it means sacrificing some study time.

Figure 1. Part of a Reading Text Immediately Relevant to Medical Students

To observe procedural fairness, I have tried to make scoring rubrics for written tests as objective as possible. For instance, a common item type in medical education context is paraphrasing, in which students have to rewrite a text using their own words. Lately, I have marked parts of the texts which students have to replace with equivalents from their knowledge of English words and phrases. Figure 2 is an artifact that illustrates how this has been done. I have also tried to treat my students respectfully during the whole assessment process to observe interactional fairness as well as design tests rooted in classroom practice to help them feel that the test outcome is an image of their true ability (distributive fairness). Honestly, however, I have not yet systematically surveyed their perceptions of my assessment practices although I have plans to do so in the future.

c) Paraphrase the following text. The parts which have been **_marked_** have to be replaced with your own words (3 points).

Although smallpox **_has ceased to kill_**, it remains a **_potential threat_** to humanity. The **_accidental release_** of material traditionally **_set aside_** for variolation in **_remote communities_** is still possible. However, the most **_pressing_** fear is that **_stocks_** of the variola virus **_stored_** for research purposes could someday be used as a biological **_warfare agent_**. The **_proposal_**, in 2003, to **_inoculate_** health care staff in some countries shows that the potential of smallpox to kill remains **_as strong as ever_**.

Figure 2. A Paraphrasing Task with Marked Parts to Make Its Scoring Process Objective

As to delivery mode, I provide my students with a choice regarding their preferences for a particular delivery mode. Now, my students can decide to take the whole test online or using the traditional paper-and-pencil modality. They are free to take some sections (e.g., grammar and vocabulary) of a test online and some other sections (e.g., writing) of the same test in paper-and-pencil format. Those who choose to take the test online, attend the university IT center and take it while being proctored. I arrange with the IT center to administer tests when there is sufficient interval between them so that I can give my students enough time to complete the tasks. These measures are taken to accommodate individual students' level of computer literacy, reduce anxiety, and enhance academic integrity. Hopefully, these modifications can alleviate the social consequences of my assessment practices.

Musa: *I gained valuable insights. Firstly, I became more acutely aware of the importance of fair assessment and that I should provide students with clear instructions as to the content, type (format) of assessment tasks, and scoring criteria. Secondly, I strive to design innovative and interactive assessment methods (e.g., interactive quizzes) to enhance students' engagement and learning. Thirdly, I have increased the variety of assessment tasks and formats to accommodate students' individual needs and learning styles. Fourthly, I try to employ more authentic assessment tasks to not only be more stimulating for learners but also gain a better understanding of students' knowledge and skills. Finally, I realized the value of immediate feedback, so now I try to use timely feedback practices such as peer and self-assessment so that students identify gaps in their knowledge and take the initiative to address those gaps.*

As evident from the above dialogs, teachers' practices have been extensively modified following the chaotic situations during the pandemic. In some cases, the circumstances resulted in heightened awareness of the significance of administering fair and just assessment while in some other cases teachers had to add new features to their practical arsenal. These new features have seemingly helped them navigate present issues and highlighted the need to be watchful for similar issues in the future.

Discussion and Conclusion

Using a duoethnographic approach, the two EFL teacher researchers in the present study examined their knowledge, beliefs, and practices regarding fairness and justice in TELA. Specific questions informed by the literature were proposed under each domain and the teachers answered those questions through engaging in dialogs and sharing artifacts. With respect to what teachers know about fairness and justice, it was revealed that they treated the two concepts as distinct assessment qualities each made up of their own particular components. As to beliefs, the teachers

emphasized the significance of fairness and justice considerations in classroom assessment and the value of students' feedback on how fair and just assessment practices are. Moreover, the teachers believed that computer literacy cannot be yet considered part of language ability although calls (e.g., Jin & Yan, 2017) have already been made to incorporate it into models of communicative competence. Regarding practice, lessons were learned from the chaotic conditions of the pandemic, including providing students with the opportunity to choose their preferred assessment delivery mode and to run oral exams to observe academic integrity. Furthermore, delivery mode fairness was identified as a new aspect of fairness in TELA in addition to distributive, informational, interactional, and procedural fairness.

On the knowledge domain, the two researchers aimed at conceptualizing fairness and justice as well as identifying the components of each concept with a focus on TELA and its use in classroom assessment. In line with Kane (2010), Walters (2012), and ETS (2022), one of the researchers viewed fairness as an independent test quality distinct from validity and reliability while the other researcher, following Kunnan (2010), regarded fairness as an umbrella quality encompassing validity. Another point of contention between the two researchers was the distinction between fairness and justice; while one of the authors agreed with the differentiation made between the two concepts by McNamara and Ryan (2011), the other researcher used fairness and justice interchangeably. To reach a consensus on these issues, the two researchers engaged in further dialogs and interacted with the relevant literature. Finally, they both agreed on the independent nature of fairness and the distinction between fairness and justice in that fairness was a test-internal quality whereas justice pertained to the consequences of assessment use and its benefits for the society. The phenomenon of "changes in thinking as a result of dialoguing with a co-researcher" has been documented by Yeo and Lewis (2019, p. 110). As to the definition of fairness, the two researchers already agreed that it referred to the quality of a test which required it to be free from gender, ethnic, cultural, and socioeconomic status bias. Moreover, the researchers were of the same opinion as to the components of fairness (e.g., distributive, informational, interactional, and procedural fairness) and social justice (e.g., access to computer and the availability of a stable internet connection). Similar findings have been reported by Wallace and Qin (2021) and Azizi (2022). Of course, the present researchers regarded informational fairness as an independent component of classroom assessment fairness rather than subsuming it under interactional fairness mainly to raise awareness of its importance and the special attention it needs to receive.

As to the domain of beliefs, the two teacher researchers observed that considerations of fairness and social justice warranted equal attention in both high-stakes and classroom assessments. This piece of finding confirmed Wallace's (2018) observation that most assessment practices occur in the classroom and, hence, fairness and justice are as important for classroom assessment as high-stakes exams. With respect to students' perspectives, the present study, confirming Wallace and Qin (2021), emphasized the value of students' perceptions in administering fair and just assessments and modifying relevant practices. Nevertheless, it highlighted the point that students' perceptions should be drawn upon as only one among an array of sources of feedback on assessment fairness and justice. Regarding the effect of computer literacy in technology-enhanced language classroom assessment, the researchers agreed that computer literacy was an integral part of modern communication, but it was too early to consider it an aspect of the construct of language ability, at least in the context of Iran. Puspawati (2014) and Chung (2017) have also proposed that computer literacy affects assessment fairness due to the limited access to computers and relevant technologies in some contexts. Nevertheless, counter arguments have been forwarded by Douglas (2013), who has called for a reconceptualization of language ability with a focus on technology integration, and Jin and Yan (2017), who argue that, due to its widespread use, computer literacy is now part of language ability.

In terms of practice, the researchers recounted that during the stringent conditions of the pandemic they resorted to TELA in their classroom assessment while they took fairness and social justice issues for granted. They used online oral exams to guarantee academic integrity and allowed their students to retake a test or submit their responses through social networking sites to reduce their anxiety. These measures were taken to eliminate social consequences of assessments during a health crisis. Following the pandemic, the two researchers tried to pay closer attention to fairness and social considerations of their assessment practices by taking into consideration the feedback they received from their students and the lessons they learned during COVID-19. Accordingly, they now provide their students with a choice to take a whole test or part of a test through the delivery mode of their own preference. This practice suggests that delivery mode fairness can be identified as a new component of classroom assessment fairness or as another important aspect of procedural fairness. They also try to be creative in designing assessment tasks and incorporate more real-life tasks in their assessment practice. The study identifies the need to adjust practice based on the lessons learned, contextual realities, and feedback provided by members of the community, including students and colleagues.

Our experience of duoethnography illustrates that it is a highly effective method in raising teachers' awareness of (gaps in) their knowledge, beliefs, and practices. Moreover, unlike Banegas and del Pozo Beamud (2022) who think of duoethnography as a combination of two autoethnographies, we believe, given the changes in our knowledge and practices, that the sum of duoethnography is greater than the combination of two autoethnographies. Duoethnography has also the potential to serve as an efficient method of peer learning, professional development, and even teacher training, through which teachers can improve their efficacy and instructional practices. As a research method, duoethnography is a highly generative approach since the interaction with the relevant literature generates ideas for research while exposing personal dialogs to criticism from experts and colleagues stirs further dialogs. Last, but not least, we found duoethnography greatly useful in systematically documenting our knowledge, beliefs, and practices, which we held intuitively and took for granted prior to engaging in this study.

The findings of the present study imply that we should be on full alert for the factors that may affect our classroom assessments, particularly when we use technology. In terms of fairness, for instance, we should inform our students of the technological requirements of our assessments and eliminate the effects of computer literacy and (lack of) access to technology on the distributive fairness of our classroom assessment events. Similarly, we should be cautious about the consequences of technology-enhanced classroom assessment such as increased anxiety and cost. The study also implies that respectful interaction prior to, during, and following an assessment procedure plays a key role in our students' perceptions of fair and just classroom assessment. As to teacher learning, the results indicate that engaging in dialogs with our colleagues and peers about our assessment beliefs and practices can be a reliable source of continuing professional development. Methodologically, the study implies that duoethnography has the potential to enrich applied linguistics research as the fruit of conducting research in terms of teacher learning is immediate.

The limitations of the present study are the inherent limitations of duoethnographic research. First, dialogs may be excessively reconstructed since the researchers involved in duoethnography are well aware that their dialogs may be criticized by scholars and colleagues. We navigated this limitation by the comfort that knowledge, beliefs, and practices are subject to constant change. Another limitation of duoethnography, persistent in most qualitative research methods, is that it generates overwhelming bulk of data which are too huge to present through a journal article in a coherent manner. We resolved the issue by agreeing on a number of key questions specifically designed for the purposes of the present study. Hence, we had dialogs on other relevant questions which were raised during the course of the study, but we did not include them in the final report. Examples of these questions are the ones we focused on to resolve our disagreement on our

approach to the relationship between fairness and validity and the distinction between fairness and justice. Furthermore, the questions addressed in a duoethnography may be biased by researchers' preferences. Here, we allowed the questions to emerge from the literature based on the most recurrent themes attended to by other researchers. Finally, the value of a duoethnographic study lies in further dialogs that it generates rather than the generalizability power of its findings.

Future researchers are encouraged to investigate language teachers' classroom assessment practices since this is an overlooked area of research. The findings of such research can provide us with valuable insights into different aspects of language classroom assessment, including the use of technology. Furthermore, studies which have specifically focused on justice and social consequences of language assessment are rare (e.g., Shohamy, 2022), and the field is awaiting models of examining social consequences of language assessment and empirical studies based on such models, be it TELA or non-TELA and high-stakes or classroom assessments. Language teachers are also encouraged to exploit the potentials of duoethnography for the purposes of their own professional development in areas of their own interests. The findings of the current study and future similar studies provide us with knowledge and skills to observe our students' rights by administering classroom assessments that are fair and just and serve the benefits of our societies as a whole. Hopefully, the findings can raise EFL teachers' awareness of the need to engage in constant reconsiderations and modifications of their knowledge, beliefs, and practices in the use of technology for the purposes of language classroom assessment.

References

Ahmed, A., & Morgan, B. (2021). Postmemory and multilingual identities in English language teaching: A duoethnography. *The Language Learning Journal, 49*(4), 483-498. https://doi.org/10.1080/09571736.2021.1906301

Alavi, S. M., Dashtestani, R., & Mellati, M. (2022). Crisis and changes in learning behaviours: Technology-enhanced assessment in language learning contexts. *Journal of Further and Higher Education, 46*(4), 461-474. https://doi.org/10.1080/0309877X.2021.1985977

Amini Farsani, M., Jamali, H. R., Beikmohammadi, M., Daneshvar Ghorbani, B., & Soleimani, L. (2021). Methodological orientations, academic citations, and scientific collaboration in applied linguistics: What do research synthesis and bibliometrics indicate? *System, 100*, 1-17. https://doi.org/10.1016/j.system.2021.102547

Amini Farsani, M., & Babaii, E. (2019). EFL faculty members' research engagement: Voices from Iran. *Research in Post-Compulsory Education, 24*(1), 37-59. https://doi.org/10.1080/13596748.2019.1584437

Attali, Y., Runge, A., LaFlair, G. T., Yancey, K., Goodwin, S., Park, Y., & von Davier, A. A. (2022). The interactive reading task: Transformer-based automatic item generation. *Frontiers in Artificial Intelligence, 5,* 1-13. https://doi.org/10.3389/frai.2022.903077

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

Beheshti, S., & 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

Banegas, D. L., & del Pozo Beamud, M. (2022). Content and language integrated learning: A duoethnographic study about CLIL pre-service teacher education in Argentina and Spain. *RELC Journal, 53*(1), 151-164. https://doi.org/10.1177/0033688220930442

Banegas, D. L., & Gerlach, D. (2021). Critical language teacher education: A duoethnography of teacher educators' identities and agency. *System, 98*, 1-12. https://doi.org/10.1016/j.system.2021.102474

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*(2), 81-109. https://doi.org/10.1017/S0261444803001903

Brunfaut, T. (2023). Future challenges and opportunities in language testing and assessment: Basic questions and principles at the forefront. *Language Testing, 40*(1), 15-23. https://doi.org/10.1177/02655322221127896

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education, 159*, 104024. https://doi.org/10.1016/j.compedu.2020.104024

Chapelle, C. A., & Voss, E. (2017). Utilizing technology in language assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 149-161). Springer. https://doi.org/10.1007/978-3-319-02261-1_10

Chung, Y-R. (2017). Validation of technology-assisted language tests. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 332-347). Wiley Blackwell.

Cohen, A. D., Rahmati, T., & Sadeghi, K. (2023). Test-taking strategies in technology-assisted language assessment. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated language assessment* (pp. 235-254). Routledge. https://doi.org/10.4324/9781003292395-19

Copeland, C. (2021). Student perceptions of the fairness of video presentation grading in the online EFL classroom. *Journal of English Teaching through Movies and Media, 22*(4), 27-38. https://doi.org/10.16875/stem.2021.22.4.27

Deygers, B. (2017). Just testing: Applying theories of justice to high-stakes language tests. *ITL-International Journal of Applied Linguistics, 168*(2), 143-163. https://doi.org/10.1075/itl.00001.dey

Douglas, D. (2013). Technology and language testing. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-7). Blackwell. https://doi.org/10.1002/9781405198431.wbeal1182

Educational Testing Service. (2022). *ETS guidelines for developing fair tests and communications.* https://www.ets.org/content/dam/ets-org/pdfs/about/fair-tests-and-communications.pdf

Gebril, A. (Ed.). (2021). *Learning-oriented language assessment: Putting theory into practice.* Routledge.

Hamid, M. O., Hardy, I., & Reyes, V. (2019). Test-takers' perspectives on a global test of English: Questions of fairness, justice and validity. *Language Testing in Asia, 9,* 1-20. https://doi.org/10.1186/s40468-019-0092-9

Hamp-Lyons, L. (1997). Ethics in language testing. In C. M. Clapham & D. Corson (Eds.), *Language testing and assessment: (Encyclopaedia of language and education* Vol. 7, pp. 323-333). Kluwer Academic.

Hommel, B. E., Wollang, F-J. M., Kotova, V., Zacher, H., & Schmukle, S. C. (2022). Transformer-based deep neural language modeling for construct-specific automatic item generation. *Psychometrika, 87*(2), 749-772. https://doi.org/10.1007/s11336-021-09823-9

Huang, P., & Karas, M. (2020). Artefacts as 'co-participants' in duoethnography. *TESL Canada Journal, 37*(3), 64-74. https://doi.org/10.18806/tesl.v37i3.1344

Huawei, S., & Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies, 28*, 771-795. https://doi.org/10.1007/s10639-022-11200-7

Huggins-Manley, A. C., Booth, B. M., & D'Mello, S. K. (2022). Toward argument-based fairness with an application to AI-enhanced educational assessments. *Journal of Educational Measurement, 59*(3), 362-388. https://doi.org/10.1111/jedm.12334

Jin, Y. (2022). Consequential research of accountability testing: The case of the CET. *Language Testing in Asia, 12,* 1-19. https://doi.org/10.1186/s40468-022-00165-6

Jin, Y., & Yan, M. (2017). Computer literacy and the construct validity of a high-stakes computer-based writing assessment. *Language Assessment Quarterly, 14,* 1-19. https://doi.org/10.1080/15434303.2016.1261293

Jing, Y., & Reynolds, B. L. (2022). A duoethnography of a Chinese pre-service teacher's encounters with young learner spelling errors in the English classroom. *Languages, 7*(4), 1-19. https://doi.org/10.3390/languages7040298

Karami, H. (2013). The quest for fairness in language testing. *Educational Research and Evaluation: An International Journal on Theory and Practice, 19*(2-3), 158-169. https://doi.org/10.1080/13803611.2013.767618

Karas, M., & Uchihara, T. (2021). Silence: A duoethnography. *Journal of Silence Studies in Education, 1*(1), 64-75. https://doi.org/10.31763/jsse.v1i1.5

Kane, M. (2010). Validity and fairness. *Language Testing, 27*(2), 177-182. https://doi.org/10.1177/0265532209349467

Kunnan, A. J. (2000). Fairness and justice for all. In A. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). 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

Lan, W., & Lam, R. (2020). Exploring an EFL teacher's beliefs and practices in teaching topical debates in mainland China. *Iranian Journal of Language Teaching Research, 8*(1), 25-44. doi: 10.30466/ijltr.2020.120806

Larenas, C. D., Hernandez, P. A., & Navarrete, M. O. (2015). A case study on EFL teachers' beliefs about the teaching and learning of English in public education. *Porta Linguarum, 23*, 171-186. https://doi.org/10.30827/Digibug.53765

Lawrence, L., & Lowe, R. J. (2020). *Duoethnography in English language teaching research, reflection and classroom application.* Multilingual Matters.

Lowe, R. J., & Kiczkowiak, M. (2016). Native-speakerism and the complexity of personal experience: A duoethnographic study. *Cogent Education, 3*(1), 1-16. https://doi.org/10.1080/2331186X.2016.1264171

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testin, 30*(3), 329-344. https://doi.org/10.1177/0265532213480129

McNamara, T., Knoch, U., & Fan, J. (2019). *Fairness, justice, and language assessment: The role of measurement.* Oxford University Press.

McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: The place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly, 8*(2), 161–178. https://doi.org/10.1080/15434303.2011.565438

Mimirinis, M. (2019). Qualitative differences in academics' conceptions of e-assessment. *Assessment & Evaluation in Higher Education, 44*(2), 233-248. https://doi.org/10.1080/02602938.2018.1493087

Neiriz, R., Baghestani, S., Muhammad, A. A., & Ranalli, J. (2023). Fairness in remote English placement testing at Iowa State University during the covid-19 pandemic. In K. Sadeghi, (Ed.), *Technology-assisted language assessment in diverse contexts: Lessons from the transition to online testing during COVID-19* (pp. 148-162). Routledge.

Ockey, G. J., Muhammad, A. A., Prasetyo, A. H., Elnegahy, S., Kochem, T., Neiriz, R., Kim, H., & Beck, J. (2021). Iowa State University's English placement test of oral communication in times of COVID-19. *Language Assessment Quarterly, 18*(1), 26-35. https://doi.org/10.1080/15434303.2020.1862122

Purpura, J. E., Davoodifard, M., & Voss, E. (2021). Conversion to remote proctoring of the Community English Language Program Online Placement Exam at Teachers College, Columbia University. *Language Assessment Quarterly, 18*(1), 42-50. https://doi.org/10.1080/15434303.2020.1867145

Puspawati, I. (2014). Fairness issues in a standardized English test for nonnative speakers of English. *TESOL Journal, 5*(3), 555-572. https://doi.org/10.1002/tesj.157

Rasooli, A., DeLuca, C., Rasegh, 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

Rose, H., & Montakantiwong, A. (2018). A tale of two teachers: A duoethnography of the realistic and idealistic successes and failures of teaching English as an international language. *RELC Journal, 49*(1), 88-101. https://doi.org/10.1177/0033688217746206

Sadeghi, K., & Douglas, D. (Eds). (2023). *Fundamental considerations in technology mediated language assessment.* Routledge. https://doi.org/10.4324/9781003292395

Sadeghi, K., & Rahmati, T. (2017). Integrating assessment *as*, *for*, and *of* learning in a large-scale exam preparation course. *Assessing Writing, 34*, 50-61. https://doi.org/10.1016/j.asw.2017.09.003

Shohamy, E. (2022). Critical language testing, multilingualism and Social Justice. *TESOL Quarterly, 56*(4), 1445-1457. https://doi.org/10.1002/tesq.3185

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2022). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology, 53*(2), 349-366. https://doi.org/10.1111/bjet.13169

Tay, L., Woo, S. E., Hickman, L., Booth, B. M., & D'Mello, S. K. (2022). A conceptual framework for investigating and mitigating machine-learning measurement bias (MLMB) in psychological assessment. *Advances in Methods and Practices in Psychological Science, 5*(1), 1-30. https://doi.org/10.1177/25152459211061337

UNESCO Institute for Statistics. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2.* https://unesdoc.unesco.org/ark:/48223/pf0000265403.locale=en

von Davier, A. A. (2020). Sphinx: An automated generation system for English reading Comprehension assessment. *International Conference on Learning Analytics and Knowledge.* https://www.semanticscholar.org/author/A.-V.-Davier/3437408?q=Sphinx&sort=influence

Wallace, M. P. (2018). Fairness and justice in L2 classroom assessment: Perceptions from test-takers. *Journal of Asia TEFL, 15*(4), 1051-1064. http://www.doi.org/10.18823/asiatefl.2018.15.4.11.1051

Wallace, M. P., & Qin, C. Y. (2021). Language classroom assessment fairness: Perceptions from students. *LEARN Journal: Language Education and Acquisition Research Network, 14*(1), 492-521. https://so04.tci-thaijo.org/index.php/LEARN/index

Walters, F. S. (2012). Fairness. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 469–478). Routledge.

Webb, M., & Ifenthaler, D. (2018). Assessment as, for, and of twenty-first century learning using information technology: An overview. In J. Voogt, G. Knezek, R. Christensen, & K-W. Lai (Eds.), *Second handbook of information technology in primary and secondary education* (pp. 581-600). Springer. https://doi.org/10.1007/978-3-319-53803-7_37-1

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

Xu, Y., & Brown, G. T. L. (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

Yeo, M., & Lewis, M. (2019). Co-authoring in action: Practice, problems and possibilities. *Iranian Journal of Language Teaching Research, 7*(3). 109-123. doi: 10.30466/ijltr.2019.120739

Teymour Rahmati is an assistant professor of applied linguistics at Guilan University of Medical Sciences, Rasht, Iran. His research interests include language teacher education, professional development, teacher motivation, and technology-enhanced language assessment. He has published papers in these areas with *Assessing Writing, RELC Journal, and The Qualitative Report,* among others.

Musa Nushi is an associate professor of applied linguistics at Shahid Behehsti University, Tehran, Iran. His research interests lie mainly in the interface of second language acquisition and second language instruction, with particular emphasis on the role of technology and positive psychology in the second language development.