

Test Fairness in Traditional and Dynamic Assessment

Afsaneh Baharloo
Shiraz University, Shiraz, Iran

Abstract—This paper aims to unfold the conceptualizations of test fairness from two different perspectives: traditional and dynamic assessment. First, it goes over a variety of definitions presented for fairness. The paper then discusses three views regarding the relationship between test fairness and test validity in order to get better insights into the nature of the intended concept. It further investigates Kunnan's test fairness framework (TFF) as one of the most comprehensive models presented for test fairness. It tries to review and criticize this model. It is worth noting that the three views, discussed in this paper, represent fairness from a traditional perspective. Furthermore, the study elaborates on dynamic assessment and its main tenets since it intends to compare the conceptualizations of fairness within traditional and dynamic assessment. In fact, fairness is viewed from a completely different perspective in dynamic assessment in which instruction and assessment are integrated and dialectically related to form an approach which prioritizes development over measurement.

Index Terms—traditional assessment, dynamic assessment, test fairness, Kunnan's test fairness framework, language development

I. INTRODUCTION

Testing is a multi-faceted and intricate field in which right decision-making is very complicated. In order for any evaluation to be reliable, a number of considerations should be taken into account. In fact, evaluation usually leads into making decisions about individuals and situations; therefore, several consequences will follow as a result of the decisions. Some of these consequences are social or psychological, affecting individuals' motivation, goal, and even social status. As Bachman (1990) states, "since testing takes place in an educational or social context, we must also consider the educational and social consequences of the uses we make of tests" (p. 237). Thus, one can easily notice that testing involves many intricacies because it eventually requires raters to judge test takers on their potentials and causes certain changes in their life path. Thus, thorough attention should be paid to consider as many relevant aspects as possible in order to make fair judgments.

Fair judgment requires measurement professionals to be aware of the concept of test fairness and its characteristics as well as other pertinent testing concepts so that they know how to observe this feature in different testing contexts as much as possible. Though many test developers and raters think that they know what 'fairness' is, they simply consider it as a test quality which pertains to a test itself and guarantees its content validity. However, one should notice that "test fairness" is a multi-faceted issue which is not confined to the content of a test and covers other aspects of testing as well. In fact, most test developers and raters attend to superficial levels and certain dimensions of test fairness which are easier to reach and do not make any attempt to achieve fairness in its full sense and at a higher level.

Fairness should not be restricted to either test development or administration. According to Willingham and Cole (1997), fairness should apply to all assessment stages including the conceptualization underlying assessment as well as the way the assessment results are used. One should not simply view the concept of test fairness as being confined to the test itself. As a matter of fact, having so simplistic a point of view about such an important and complex issue results in unfair testing contexts, violation of test takers' rights and finally lack of sufficient research in this area.

The concept of test fairness is so complicated and controversial that yet no agreed-upon definition is provided. In addition, some fairness models have been proposed but none lends itself easily to practical investigation of fairness. Furthermore, to the best of the researcher's knowledge, no previous study has investigated the trend that fairness follows to see how its concept differs in traditional and dynamic assessment. Therefore, the current study is an endeavor to provide a comprehensive portrait of 'test fairness' and discusses Kunnan's (2004) framework as the most comprehensive model available for test fairness. It also presents different views about fairness and elaborates on each view. Finally, the study compares fairness in traditional and dynamic assessments and hence fills the aforementioned gap.

II. TEST FAIRNESS

Test fairness has not been paid due attention for a long time. People believed that differences among groups reflected reality, and fairness was not really a concern for them. Gradually, measurement professionals began to study group differences in terms of score and fair testing. They "began to pay increasing attention" to test and item fairness almost at

the beginning of 1970s (Cole and Zieky, 2001, p. 370). Kunnan (2010) thinks that test fairness as one of the most fundamental concepts in evaluation entered the forefront of investigations and discussions in the field of language assessment in 1990s.

In fact, fairness is such a complicated concept that a variety of definitions has been proposed to clarify its broad and controversial nature. According to Webster's Ninth New Collegiate Dictionary (1988), 'fairness' means 'being free from having favor toward either or any side'. Xi (2010) believes that such a definition indicates that fair testing mainly focuses on comparing testing practices and test results across different groups. Therefore, test fairness mainly arises from the way group differences are perceived and treated. Similarly, Xi (2010) defines fairness "as comparable validity for all the identifiable and relevant groups across all stages of assessment, from assessment conceptualization to the use of assessment results" (p. 154).

Davies (2010) states that "the language testing literature has tended to treat fairness under the heading of bias" (p. 174). According to Elder (1997), test bias studies "are directed to identifying and where possible reducing the effect of any confounding variables on test scores, by making changes to the test" (p. 261). McNamara and Roever (2006) think that the term "bias" in assessment research conveys an unfair tendency for one group or population which results in the detriment of another. Therefore, the notion of bias is highly related to fairness in all stages of assessment: "A biased judgment unduly takes into account factors other than those that should be informing it" (p. 82). Angoff (1993) believes that in traditional terms, bias can be considered as construct-irrelevant variance which threatens the test outcomes and hence reduces the validity of the conclusions made based on scores. Particularly, one can consider a test as biased if test takers who have equal ability but come from different groups score differently on the items of the test due to their group membership.

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 the testing process, and equity in opportunity to learn the material in an achievement test" (Xi, 2010, p. 147). In other words, fairness requires equitable treatment of all test takers in the testing process. However, test fairness is so broad an area that many measurement professionals consider it to encompass management of quality in test design, administering and scoring, appropriate coverage of relevant material, sufficient work to ensure construct validation, equal opportunities for learning and access to testing, and items which measure only the ability under investigation without being affected by factors related to test-takers' background (McNamara and Roever, 2006; Kunnan, 2000; Saville, 2003, 2005; Shohamy, 2000).

Davies (2010) believes that the concept of fairness has been deeply studied by John Rawls (2001) who equates fairness with justice. Rawls proposes two principles underlying his argument. "The first is that everyone has the same claim to the basic liberties. The second is that where there are inequalities they must satisfy two conditions, that offices and positions must be open to everyone on the basis of equality of opportunity, and that the least-advantaged members of society should benefit most from these inequalities" (Davies, 2010, p. 174). Having a similar idea, Velasquez *et al.* (2008) link fairness to justice and define justice as providing everyone with what they deserve, or, in other words, offering each individual their due.

As it was mentioned, measurement professionals have defined test fairness in different ways. Such definitions may not clarify the concept of test fairness and its relevant aspects to the extent that practitioners can observe fairness in actual testing contexts. They may need a concrete model through which fairness can really be observed and applied in testing situations. Among all the available models proposed for test fairness, Kunnan's (2004) framework can be considered as the most comprehensive fairness model. In what follows this model is explained and criticized.

III. THE TEST FAIRNESS FRAMEWORK

A number of approaches have been proposed to investigate fairness. However, the test fairness framework, proposed by Kunnan (2004), is the main concern of this study since this model has been at the forefront of attention regarding test fairness for several years. Kunnan (2010) puts forward an ethics-inspired rationale for his framework and claims that this model considers the whole testing system, not only the test itself, hence it seems to be more comprehensive than the other existing models. Kunnan's (2004) framework was the first attempt made to "propose an overarching framework for fairness research" in language testing (Xi, 2010, p. 150). In his previous work on test fairness (Kunnan, 2000), he considers fairness as a concept encompassing three facets to deal with validity, justice and access. Kunnan agrees with Jensen (1980) who thinks that "the concepts of fairness, social justice, and equal protection of the laws are moral, legal, and philosophical ideas and therefore must be evaluated in these terms" (Jensen, p. 376). Xi (2010) also thinks that such an idea mainly arises from social justice theories. However, Kunnan tried to expand his ideas and develop a more comprehensive model that was later proposed in 2004. In this framework, he views fairness as an overarching concept which includes five testing qualities: Validity, absence of test bias, equal access for learning and testing, test administration, and social consequences. According to this approach, validity of a test score interpretation, which is considered as part of the test fairness framework, can be supported through four types of evidences: *Content representativeness or coverage evidence* which refers to the adequacy with which test items represent the test domain, *construct or theory-based validity evidence* which refers to the extent to which test items represent the construct or the underlying trait being measured in a test, *criterion-related validity evidence* which refers to whether the test scores meet some criterion variables, and *reliability evidence* which refers to the consistency of test scores. The second quality

refers to absence of any source of bias such as offensive content or language, unfair penalization based on test takers background, and disparate impact and standard setting. In fact, offensive content can cause bias for test takers from different backgrounds because it may conflict with their beliefs or it may be needlessly controversial (McNamara and Roever, 2006). A test is also biased if it causes unfair penalization due to a test taker's group membership. In addition, disparate impact and standard setting can bring about different performances by test takers from various group memberships. The third quality of the fairness framework refers to test takers' access to the test in terms of educational, financial, geographical, personal, and equipment access. In other words, all the test takers should be provided with opportunities to learn the test content and get familiar with the kind of tasks and cognitive demands required by the test. Furthermore, the test should be affordable for test takers and the site should also be accessible in terms of distance as well as test takers' physical and learning conditions. In addition, test takers should be familiar with the test taking equipments, procedures, and conditions. The fourth feature of test fairness framework is related to administration conditions. This quality refers to the physical conditions of test administration such as optimum light and temperature as well as uniformity and consistency across test sites and in equivalent forms. Finally, social consequences of a test should be studied as contributing to test fairness. These consequences refer either to the effect of a test on instructional practices or the remedies offered to test takers to compensate for the detrimental consequences of a test.

Although Kunnan's test fairness framework considers many relevant factors and thus seems to be a comprehensive model, it has several shortcomings. First, it mainly arises from theories and is not practical enough to provide a principled guideline to ensure all the aspects of test fairness. Having a similar idea, Xi (2010) also criticizes Kunnan's (2004) framework for not providing "practical guidance on how to go about developing the relevant evidence to support fairness" (p. 148). He thinks that although this approach may be helpful in dealing with general aspects of research, it does not provide a practical "mechanism to integrate all aspects of fairness investigations into a fairness argument", nor does it "offer a means to plan fairness research" (p. 148). Xi believes that a framework should identify and prioritize research needs so that one can practically observe and measure the progress made in fairness investigations.

Second, although Kunnan (2004) claims that this fairness framework can apply to the whole system of a testing practice, it does not actually consider all stages regarding assessment. The Code of Fair Testing Practices in Education (1988), modified in 2004, highlights the role of test fairness as a quality that pertains to all assessment stages. Therefore, fairness issues are not confined to the development of appropriate and adequate test items, their administration and scoring but they also pertain to providing test takers with the accurate and sufficient reporting of test results since individuals have rights to receive feedback on their performance so that they get aware of their strength and weakness. However, there is no concern for the latter issues in Kunnan's test fairness framework; thus, it is not comprehensive enough to consider all the aspects and consequences of a testing practice. In addition, this test fairness framework does not specifically define any qualities devoted to the responsibilities of test developers and users regarding the importance of their roles. On the one hand, test developers should provide test users with sufficient information and supporting evidence to help them select appropriate tests. They should also explain the procedures needed for administering and scoring tests appropriately and fairly. On the other hand, test users should inform test takers about their responsibilities and rights, the nature and purpose of the test, the appropriate use of test results, and procedures used for resolving challenges encountered in the evaluation process (McNamara and Roever, 2006).

Finally, Kunnan's test fairness framework mainly focuses on *group differences* and the kind of bias that may stem from test takers' membership in different groups, but it does ignore the important issue of *individual differences*. This lack of attention to individual differences may result in having test items and tasks, which are more suitable and convenient for some individuals with certain learning styles; but are not appropriate for all the members of the same group. For instance, the same test given to two groups of men and women can yield different results that may be attributed to gender differences. However, one should note that there are some intra-group differences regarding the ability being tested, for example: Test taking strategies or learning styles that can bring about different performances. Highlighting the importance of individual differences, Cole and Zieky (2001) state that, "all testing data show far more individual variation of scores within groups than variation between groups. Individual variation, not group variation, is the dominant influence on scores and should therefore be the dominant fairness concern" (p. 11). Therefore, considering the qualities and aspects included in Kunnan's test fairness framework, one eventually finds out that this model does not propose a practical means to investigate test fairness in its full sense.

Xi (2010) believes that establishing a fairness framework that would be useful for practical purposes requires primary attention to the conceptualization of fairness. Fairness related theories, models, and definitions suggest that most measurement professionals study test fairness in relation to validity issues since test fairness is sometimes influenced by the interpretations of test scores. Therefore, studying about how fairness and validity are related can provide better insights into the conceptualization of fairness and its practical investigation.

IV. FAIRNESS AND VALIDITY

Fairness has been conceptualized in various ways which result in different approaches of viewing fairness. Though these conceptual approaches may vary with regard to the degree of their emphasis on issues such as the political and social aspects of fairness, a major point on which they differ is how fairness and validity are related (Xi, 2010).

Considering the relationship that may exist between fairness and validity, Xi proposes three views: “whether fairness is independent of validity, subsumes it, or is a facet of it” (p.148).

View 1: Fairness as an independent test quality

This view considers fairness as a test quality facet which is relatively independent and is not consistently and clearly connected to validity. According to this view, “fairness is characterized as a test quality that is separate from validity, although some tenuous and inconsistent references may be made to validity” (Xi, 2010, p. 149). The *Standards for Fairness and Quality* by Educational Testing Service (ETS, 2002) and the *Code of Fair Testing Practices in Education* (1998, 2004) are representative of this approach. The *1999 Standards* claims that “fairness requires that construct-irrelevant personal characteristics of test takers have no appreciable effect on test results or their interpretation” (p. 17). The Code primarily focuses on the partition of responsibilities between test developers and users in ensuring testing practices which are supposed to be fair. Xi (2010) states that “This is also a major contribution of the Code compared to the Standards, as the partition of responsibilities between test developers and users has not always been clear cut (Shohamy, 2001). Since it requires both test developers and users to work in concert to ensure fairness, guidelines as to who is responsible for what help promote fairness” (p. 150). In addition, the ETS Standards for Fairness and Quality presents a broad list of fairness standards; but it does not provide a mechanism for weighing one piece of fairness evidence against another or for prioritizing them either. Furthermore, Xi points to one of the standards, proposed in the ETS Standards for Fairness and Quality, which explains “that if the use of assessment results causes unintended consequences for a studied group, the validity evidence should be investigated to see if the differential impact for the studied group is a result of construct irrelevant factors or construct under-representation” (p. 151). Such elaboration implies the potential existence of a more consistent and coherent linkage between test fairness and validity (Xi, 2010).

View 2: Fairness as an overarching test quality

According to this view, fairness subsumes validity; in other words, fairness is viewed as an all-encompassing test quality with different facets including validity. This view highlights the importance of fairness issues by giving primacy to test fairness and defining it as a test quality, which subsumes validity and goes beyond it. Therefore, for a test to be fair, it must be valid first. Kunnan’s test fairness framework is a manifestation of this view since validity is considered as one of the five qualities that contribute to fairness. Kunnan views fairness as a test quality which includes validity, absence of test bias, access to the test, conditions of administration, and consequences of the test (Kunnan, 2004, 2010). However, this view considers fairness as a broad concept, which consists of several facets that are related to one another. That is why Bachman (2005) criticizes Kunnan’s work in which various fairness components are not necessarily connected to each other. Therefore, he emphasizes on the need for a mechanism to integrate them appropriately to support a comprehensive fairness argument.

In addition, McNamara and Roever (2006), who are proponents of the second view, focus on the social dimensions of language testing evident in item bias investigations. The desire to ensure social justice motivates their test fairness discussion. They argue that factors, which bias some test takers’ educational opportunities and bring about advantages and disadvantages for them, comes from the larger social context. McNamara and Roever put great emphasis on the political and social dimensions of fairness. They believe that the social embeddedness of testing can be dealt with through the procedures of sensitivity review and the promotion of codes of ethics. Sensitivity review or fairness review refers to the formal process of identifying, modifying or eliminating possibly biased items during the construction process of a test. The codes of ethics are “useful for guiding ethical decisions and protecting testers from stakeholder pressures to take actions that contravene professional conduct” (McNamara and Roever, 2006, p. 7). Therefore, adopting this view requires thorough attention to all the aspects of fairness and their relations as well as the social and political context of a testing practice.

View 3: Fairness connected directly to validity

This view arises from the *1999 Standards* which supports three prevalent features of fairness in educational and psychological testing fields: test fairness as lack of bias, test fairness as equity in opportunity to learn the content covered in an achievement test, and test fairness as equitable treatment of all test takers in the testing process. The 1999 Standards rejects the popular view that fairness requires the equality of testing results for different groups of examinees, and argues that a more widely accepted view would imply that test takers who belong to different groups and have equal standing regarding the construct being measured should on average get the same test score (Xi, 2010). In addition, the *1999 Standards* emphasizes that gathering various pieces of evidence helps to ensure test fairness. The *Standards* requires the investigation of each type of validity evidence for relevant groups of test takers to determine if the meaning and interpretation of test scores, and the consequences of using the assessment results may differ due to construct irrelevant factors or construct under-representation. Xi, who is in favor of this view, points out that the “connection between discussions of fairness and validity suggests a strong possibility for linking fairness back to validity in a principled way [which could not be observed in Kunnan’s work]. This kind of linkage would allow fairness research and practice to take advantage of a well-defined framework for validity” (p. 152).

In fact, Xi “proposes an approach for studying fairness that links it directly to validity. Fairness is characterized as comparable validity for relevant groups that can be identified” (p. 167). She considers fairness as a facet of validity. Therefore, anything that compromises fairness weakens the validity of a test as well. However, Davies (2010) criticizes Xi’s proposed conceptualization. He believes that validity itself does pertain to all identifiable and relevant groups; then

why do we need to appeal to fairness? Davies argues that “validity guarantees that an ability is being appropriately tested for a relevant population: this population will be made up of various groups but there is sufficient homogeneity across groups for them to be treated as comparable” (p. 175). He believes that a test which is valid for children may not be valid for adults because these two groups of test takers belong to different populations. It is not whether such a test is fair or unfair for adults: the test is just considered to be invalid for the latter group. Davies thinks that pursuing fairness in language testing is chimerical: “First because it is unattainable, and second because it is unnecessary” (p. 171). So Davies’s idea is not in line with any of the three conceptualizations presented above as he rejects the concept of fairness overall.

V. FAIRNESS IN TRADITIONAL AND DYNAMIC ASSESSMENTS

All the three views, discussed regarding the relationship between fairness and validity, have been proposed within the framework of traditional assessment. Traditional testing contexts draw a clear distinction between development and measurement. Assessment usually follows instruction and is not intended to improve test takers’ learning. Since such assessment is usually statistically based and grounded in psychometric principles, it considers any change in the test taker’s performance during the assessment administration as a threat to test reliability, which definitely compromises those principles as well (Poehner and Lantolf, 2005). In such a traditional perspective, test fairness requires providing learners with equal opportunities to learn and subsequently take part in exams. Instruction is planned based on a hierarchy including a sequence of increasingly difficult learning tasks. In fact, teachers provide all the learners with the same material without considering their needs and teach them equally since all learners are supposed to receive the same amount of input and support from the teacher to move through the predetermined hierarchy. In other words, teaching undergoes several distinct stages; therefore, learning can be investigated through traditional assessment instruments, designed equally for all the learners, at a particular point in the teaching sequence. All the learners receive the same test on which they should perform independently. They should not cooperate with their peers or teacher during the exam since the only purpose of assessment is measurement. Therefore, the central focus of fairness within a traditional framework is to provide learners with equal learning opportunities and access to a test which usually takes the form of a summative assessment which evaluates learners’ performance at the end of a program and is often used for the purposes of admission decisions, accountability, selection and promotion (Poehner and Lantolf, 2005).

However, dynamic assessment views teaching and testing from a different perspective in which the pursuit of fairness undergoes a different path. Dynamic assessment (DA) is an approach to instruction and assessment derived from Vygotsky’s socio-cultural theory of mind and his focus on the Zone of Proximal Development (ZPD) (Poehner and Lantolf, 2005). In this approach, teaching and testing are integrated as a single activity that aims to simultaneously understand and promote learners’ abilities through mediated interaction in the Zone of Proximal Development (Poehner, 2008). In other words, dynamic assessment blends instruction and assessment and benefits from tutor mediation to recognize the areas in which students need the most support (Shrestha, et al., 2012). Therefore, DA has two major concerns: first, teaching and testing are dialectically integrated to the extent that one cannot tell the two activities from each other at a particular point. Second, learners receive support from the teacher within their ZPD even when they are performing on a test, because this approach advocates any tools that lead to development; therefore, assessment is not mainly intended to measure learners’ knowledge but to develop it. Thus, development has priority over measurement.

ZPD refers to the ‘difference’ between what learners can do on their own and what they can do when receiving support and assistance on a test. The idea of ZPD is highly associated with Vygotsky’s socio-cultural theory of mind. He believes that engagement in activities mediated by others and by cultural tools allows learners to develop higher levels of consciousness, which are unique to humans (Vygotsky, 1978, 1986). “In Vygotsky’s view, abilities do not simply mature on their own but instead result from individuals’ histories of engaging in activities with others and with cultural artifacts” (Poehner, 2008, p.24). Socio-cultural theory implies that cultural artifacts and other individuals are not only a factor of development, but they are also the source of development. Dynamic assessment is not concerned with attributing development either to the individual or to the environment. According to this approach, the individual and the environment constitute an inseparable dialectical unity that cannot be understood if the unity is distorted or broken. The interaction between learners and their environment helps them develop control over and awareness of their psychological functions, including attention, perception, and memory (Poehner, 2008). Newman *et al.* (1989) also believe that cognitive changes arise from the productive intrusion of other individuals and cultural objects in the developmental process. Kozulin (1998, 2003) considers physical, psychological, and symbolic tools as a way of conceptualizing Vygotsky’s major argument that a learner’s social and cultural environment is the source of the development of higher psychological functions. In a Vygotskian view, humans relate to their world *physically* as well as *psychologically*. Unlike physical tools, symbolic tools that Vygotskian researchers refer to as cultural artifacts are directed both outwardly and inwardly to mediate individuals’ relationship with the world and with themselves (Vygotsky, 1994). Therefore, as Poehner and Lantolf (2005) put it, “the unit of analysis for the study of development is not the individual acting alone, but the interpersonal functional system formed by people and cultural artifacts acting jointly to bring about development” (p. 238). Wertsch (2007) believes that Vygotsky views mediation as the “hallmark of human consciousness because it is through their appropriation of the forms of mediation provided by particular cultural, historical, and institutional forces that their mental functioning is sociohistorically situated” (p. 178).

Highlighting the importance of mediation and intervention, Shrestha *et al.* (2012) believe that both ZPD and mediation are integral to dynamic assessment. While the ZPD is basically about the learner's potential development, mediation paves the way for such development. In other words, mediation can be defined as a process that humans undergo to regulate others, the material world, or their own social and mental activities by using concepts, activities and artifacts which are culturally constructed (Lantolf and Thorne, 2006).

In order to get better insights into the ways traditional and dynamic assessments differ, one can think of them as assessment *while* teaching and assessment *by* teaching respectively. According to Newman *et al.* (1989), assessment *while* teaching requires children to learn base on a hierarchy which is composed of a sequence of increasingly difficult activities. Therefore, determining how successful the children are at moving through the sequence requires the teacher to observe their independent performance on traditional assessment instruments usually in the form of psychometric tools at a particular stage in the instruction sequence. But, assessment *by* teaching, which is in line with dynamic assessment, suggests that teaching is not organized and planned according to "a neat sequence of levels to be mastered in an invariant sequence with a single correct route to mastery. Tasks and knowledge may be organized according to a teacher's assumptions about their relative complexity" (Newman *et al.*, 1989, p. 78). Therefore, once students and teachers get involved in instructional activities, things can move in unanticipated directions and learning occurs at unanticipated rates (Poehner and Lantolf, 2005).

In dynamic assessment, there is a shift of attention from focusing on learners' independent performance on traditional measuring instruments to focusing on the process of development through mediated interaction. Since development has priority over measurement in dynamic assessment, fairness suggests that learners should not be deprived from any tool that promotes their learning. Therefore, even a test, which has traditionally been used only for measurement purposes, should now be in service of development.

In fact, dynamic assessment views test fairness from a perspective different from the one underlying the three views proposed by Xi (2010). Those views discuss fairness within the framework of traditional assessment in which test fairness requires institutions and teachers to provide equal opportunities for all the individuals to learn the same material and consequently provide them with the same testing conditions for measurement purposes. In other words, traditional assessment is more product-oriented and seeks to measure the ultimate level that learners have reached. The views, already discussed regarding the relationship between fairness and validity, are included within this product-oriented approach. However, the third view seems to manifest some traces of dynamic assessment since in this view Xi points to some sort of equitable treatment of individuals but it still follows the main tenets of traditional assessment with regard to instruction and testing procedures. Unlike the traditional approach, dynamic assessment focuses on the developmental process and hence is considered as a process-oriented approach in which a test is a helpful tool that can both measure and promote individuals' knowledge so that they can transfer what they learn to other tasks beyond the test. Having a similar idea, Shrestha *et al.* (2012) state that, "DA is grounded in the notion of assessment as a process rather than a product. In other words, DA is a development-oriented process which reveals a learner's current abilities in order to help them overcome any performance problems and realize their potential" (p. 5).

It is worth noting that the major difference between the ways that traditional and dynamic assessment view fairness lies in the different teaching and testing relationships within the two approaches. In the traditional sense, instruction and assessment are separate activities carried out at particular stages. All the learners are taught the same material selected based on a predetermined hierarchy and they later receive the same test on which they should perform independently since the only purpose of such a test is measurement. Therefore, in order to observe fairness, educational systems and practitioners are required to treat all the individuals equally regardless of their needs and backgrounds. But dynamic assessment takes on a different perspective in which instruction and assessment are integrated in all the stages so that one cannot distinguish the two activities from one another at a single point. All the individuals do not receive the same instruction. In fact, each learner receives as much assistance as he or she deserves. Development is achieved intentionally rather than incidentally. As Poehner and Lantolf (2005) mention, "dynamic assessment is a pedagogical approach grounded in a specific theory of mind and mental development... [Therefore,] mediation cannot be offered in a haphazard, hit-or-miss fashion, but it must be tuned to those abilities that are maturing" (p. 260). Although one can carry out dynamic assessment either formally or informally, it must be planned and systematic. It mainly insists on the inseparability of instruction and assessment because they constitute a unity which is necessary for individuals' development. In such an approach, fairness requires providing each individual with what they deserve regarding their needs. Learners do not move through a hierarchy of tasks sequenced based on their difficulties rather each individual receives what is required for his or her development based on a theory of mind since mediated interaction and intervention should be systematic in order to be fair and beneficial. Even, their performance on a test is assisted by receiving support from others such as their peers, teachers and whatever exists in the environment. Thus, each individual receives as much assistance as he or she needs. In fact, dynamic assessment attends to individual differences in a practical sense. This is contrary to the views of traditional assessment.

Models following traditional perspectives define fairness in terms of equal treatment of individuals regarding learning opportunities and testing conditions. In such an approach, the focus is on inter-group differences rather than intra-group differences. Within this framework, tests are only used for measurement and any kind of intervention may threaten their reliability since they are often used for purposes such as admission decisions, selection, and promotion.

Therefore, test takers do not usually receive any kind of feedback on their performance to know which areas require more attention and practice. However, dynamic assessment employs a different view toward fairness. In this approach, teaching and testing are dialectically integrated and considered as a single activity since both aim at promoting learners' knowledge; and development has priority over measurement. In fact, this approach requires teachers to assist learners to overcome the difficulty of test tasks and master the intended knowledge being tapped by the test so that they can transfer such knowledge to other tasks beyond the test. Therefore, fairness in dynamic assessment does not only apply to the test itself or the testing process but to the whole teaching and testing activities integrated as a single unity which must ultimately lead to development. In other words, fairness implies that individuals should not be deprived of any opportunity that can promote their learning. From a dynamic perspective, fairness in education does not require teachers to treat all learners as if they were the same, because doing so ignores that they are not. Fairness necessitates doing everything possible to promote learner development, with the understanding that some learners will need more support, time and resources than other individuals (Poehner, 2011).

Reuven Feuerstein, a leading DA researcher, has proposed a "structural cognitive modifiability theory" to suggest that "traditional conceptualizations of the examiner/examinee roles should change in favor of a teacher-student unity that works towards the ultimate success of the student" (Feuerstein et al., 1979, p. 271). Putting this idea another way, Poehner (2011) states that the purpose of assessment is fully realized by actively trying to promote a learner's knowledge. Poehner (2011) claims that "this orientation requires a shift on the part of the assessor, also referred to as a *mediator*, whose responsibility is no longer limited to neutrally observing learner performance but now involves engaging as a co-participant with learners" (p. 100). Feuerstein et al. (2002) do not consider cognitive capabilities to be fixed traits determined by our genetic endowments rather these abilities are supposed to develop in various ways, depending on the presence, and the quality of suitable forms of instruction and interaction. Feuerstein et al. (1979) state that "it is through this shift in roles that we find both the examiner and the examinee bowed over the same task, engaged in a common quest for mastery of the material" (p. 102). Following the same line of thought, Poehner (2008) thinks that teachers' intervention is necessary to help learners stretch beyond current capabilities. In other words, interpretations of learners' abilities and knowledge are extended beyond observations of independent performance to include their responsiveness and contributions during participation and engagement in joint activities with a mediator. Moreover, the instructional quality of the interaction begins and affects the process of helping individuals move toward overcoming their current difficulties (Poehner, 2011). Sternberg and Grigorenko (2002) also state that dynamic assessment is a procedure with results that take into account the outcomes of an intervention in which the examiner teaches the examinee how to perform more efficiently on individual items or on the whole test. Therefore, the final score can be considered as a learning score which represents the difference between pretest (before an individual's learning) and posttest (after his/ her learning) scores, or it may be the score on the posttest considered alone. It seems that the major difference between the traditional and dynamic approaches has to do with whether or not the administration of the assessment should have the expressed purpose of modifying learners' performance during the assessment itself (Poehner and Iantolf, 2005).

Therefore, fairness is conceptualized differently in traditional and dynamic assessment. In traditional approaches, teaching and testing are considered as distinct activities with different objectives. Hence, fairness, in the former approach, requires institutions to provide equal opportunities and conditions for all learners to learn the same material and to perform on the same traditional measuring instrument independently. However, the latter approach which blends instruction and assessment views fairness in terms of providing each individual with what he or she deserves based on need analysis and ongoing assessments used for diagnostic purposes.

VI. CONCLUSION

This paper investigated test fairness conceptualizations within the perspectives of traditional and dynamic assessment. It started with presenting available definitions for fairness to provide a portrait of the concept at hand. In order to get better insights into the very nature of this test quality in a more concrete sense, it investigated Kunnan's (2004) test fairness framework as the most comprehensive available fairness model. However, a closer look at the framework and its components resulted in a number of criticisms. First, it is impractical due to the lack of guidelines on how to ensure validity. Contrary to what Kunnan claims about the comprehensiveness of his framework, it does not apply to the whole system of a testing practice since it does not indicate any concern for accurate reporting of test results and informing test takers as well as providing them with feedback on their performance with regard to their strengths and weaknesses. In addition, this model understates the important roles of test developers and test users by not clarifying their responsibilities in the testing process. Furthermore, it ignores the intra-group differences and only attends to inter-group differences. Therefore, the test fairness framework does not seem to be practical and comprehensive enough to be applied to the whole system of a testing practice appropriately. Xi (2010) suggests that establishing a useful framework for practical research requires measurement professionals to have concerns for the conceptualizations of fairness. She proposes three views regarding the relationship between fairness and validity. The first view considers fairness as an independent test quality, the second view, to which Kunnan's test fairness framework belongs, sees fairness as an all-encompassing test quality which is composed of several facets including validity. The third view considers fairness as being directly related to validity. It is worth noting that, all the views proposed by Xi are discussed from a traditional

perspective toward assessment. However, fairness is viewed quite differently within the framework of dynamic assessment in which instruction and assessment are integrated in order to simultaneously measure and promote learner development. In DA, development has priority over measurement. As the paper discusses, the traditional and dynamic assessments conceptualize fairness from different perspectives. While, in the former approach, fairness requires instructors and measurement professionals to teach all the learners based on a predetermined schedule and provide them with equal opportunities to learn and take tests on which they are not assisted, the latter perspective defines fairness in terms of providing each individual with what he or she deserves. Therefore, individual differences as well as their needs and interests are taken into consideration. Although, dynamic assessment seems to view fairness from a more humanistic perspective, it requires careful attention and programming on the part of educational institutions and practitioners so that all the learners get what they really need and deserve. Employing needs analysis before beginning a course and having small classes help teachers implement the tenets of dynamic assessment and reach fairness as much as possible.

REFERENCES

- [1] American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999). Standards for educational and psychological testing. Washington, DC: Author.
- [2] Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland and H. Wainer (Eds.), *Differential item functioning* (pp. 3–24). Hillsdale, NJ: Lawrence Erlbaum Associates.
- [3] Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.
- [4] Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1–34.
- [5] Cole, N.S. and Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38(4), 369–382.
- [6] Davies, A. (2010). Test fairness: A response. *Language Testing*, 27(2), 171–176.
- [7] Educational Testing Service. (2002). ETS standards for quality and fairness. Princeton, NJ: Author.
- [8] Elder, C. (1997). What does test bias have to do with fairness? *Language Testing*, 14(3), 261–277.
- [9] Feuerstein, R., Rand, Y. and Hoffman, M. B. (1979). The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. Baltimore: University Park Press.
- [10] Feuerstein, R., Falik, L., Rand, Y. and Feuerstein R.S. (2003). Dynamic assessment of cognitive modifiability. Jerusalem: ICELP Press.
- [11] Jensen, H. R. (1980). Bias in mental testing. New York: Free Press.
- [12] Joint Committee on Testing Practices. (1988). Code of fair testing practices in education. Washington, DC: Author.
- [13] Joint Committee on Testing Practices. (2004). Code of fair testing practices in education. Washington, DC: Author.
- [14] Kozulin, A. (1998). Psychological tools: A sociocultural approach to education. Cambridge, MA: Harvard University Press.
- [15] Kozulin, A. (2003). Psychological tools and mediated learning. In A. Kozulin, B. Gindis, V.S. Ageyev, and S.M. Miller (Eds.), *Vygotsky's Educational Theory in Cultural Context* (pp. 15–38). Cambridge: Cambridge University Press.
- [16] Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge: Cambridge University Press.
- [17] Kunnan, A. J. (2004). Test fairness. In M. Milanovic and C. Weir (Eds.), *European language testing in a global context* (pp. 27–48). Cambridge: Cambridge University Press.
- [18] Kunnan, A. J. (2010). Test fairness and Toulmin's argument structure. *Language Testing*, 27(2), 183–189.
- [19] Lantolf, J. P., and Thorne, S. L. (2007). Sociocultural theory and second language acquisition. In B. V. Patten and J. Williams (Eds.), *Explaining second language acquisition* (pp. 201–224). Cambridge: Cambridge University Press.
- [20] McNamara, T. F. and Roever, C. (2006). Language testing: The social dimension. Oxford: Blackwell.
- [21] Newman, D., Griffin, P., and Cole, M. (1989). The construction zone: working for cognitive change in school. Cambridge: Cambridge University Press.
- [22] Poehner, M. E., and Lantolf, J. (2005). Dynamic assessment in the language classroom. *Language Teaching Research* 9(3), 233–265.
- [23] Poehner, M. E. (2008). Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development. Berlin: Springer.
- [24] Poehner, M. E. (2011). Dynamic assessment: Fairness through the prism of mediation. *Assessment in Education, Principles, Policy and Practice*, 18(2), 99–112
- [25] Rawls, J. (2001). Justice as fairness: A restatement. Cambridge, MA: Harvard University Press.
- [26] Saville, N. (2003). The process of test development and revision within UCLES EFL. In C. Weir and M. Milanovic (Eds.), *Continuity and innovation: Revising the Cambridge Proficiency in English examination 1913–2002* (pp. 57–120). Cambridge: Cambridge University Press.
- [27] Saville, N. (2005). Setting and monitoring professional standards: A QMS approach. *Cambridge ESOL Research Notes*, 22, 2–5.
- [28] Shohamy, E. (2000). Fairness in language testing. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 15–19). Cambridge: Cambridge University Press.
- [29] Shohamy, E. (2001). The power of tests: A critical perspective of the uses of language tests. London: Longman.
- [30] Shrestha, P., and Coffin, C. (2012). Dynamic assessment, tutor mediation and academic writing development. *Assessing Writing*, 17(1), 55–70.
- [31] Sternberg, R.J. and Grigorenko, E.L. (2002). Dynamic testing. The nature and measurement of learning potential. Cambridge: Cambridge University Press.
- [32] Velasquez, M., Andre, C., Shanks, T., and Meyer, M. J. (2008). Justice and fairness. California: Santa Clara University, Markkula Center for Applied Ethics.

- [33] Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Massachusetts: Harvard University Press.
- [34] Vygotsky, L.S. (1986). *Thought and language*. Cambridge, MA: MIT.
- [35] Vygotsky, L.S. (1994). The problem of the cultural development of the child. In R. Van der Veer and J. Valsiner (Eds.), *The Vygotsky reader* (pp. 338-354). Cambridge, MA: Blackwell.
- [36] Webster's Ninth New Collegiate Dictionary. (1988). Springfield, MA: Merriam-Webster.
- [37] Wertsch, J.V. (2007). Mediation. In H. Daniels, M. Cole, and J.V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 178–92). Cambridge: Cambridge University Press.
- [38] Willingham, W. W. and Cole, N. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum.
- [39] Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2) 147–170.

Afsaneh Baharloo is a Ph.D. student of TEFL, Department of Foreign Languages and Linguistics at Shiraz University, Shiraz, Iran. She obtained her BA and MA degrees in English literature and ELT respectively from Shiraz University. She teaches English at both private and state universities. Her areas of interest are language skills, research and translation.