



## Authenticity and Authentication in Language Testing: an Operational Perspective

**Nematullah Shomoossi**

Sabzevar University of Medical Sciences, University of Isfahan, Iran

**Mansoor Tavakoli, PhD**

Department of English, University of Isfahan, Iran

**Nematullah Shomoossi**, PhD candidate in TEFL at the University of Isfahan, Iran, is the coordinator of English programs at Sabzevar School of Medical Sciences. He is involved in teaching general English as well as research in writing and learner strategies. E-mail: nshomoossi@yahoo.com

**Mansoor Tavakoli**, PhD in TEFL, is the assistant professor of Testing and Research at the Department of English, University of Isfahan, Iran. He has published and presented papers in his field, and is currently teaching and supervising graduate courses. E-mail: mansoor\_tavakoli14@yahoo.com

This paper reviews the concept of authenticity in language testing and intends to find a balance between theoretical and operational levels through the concept of authentication. The argument will be presented by evaluating the traditional and interactive perspectives. Authenticity will be viewed as an interactive and multifaceted notion. Then the concept will be examined in relation to the current views on the nature of test tasks, test takers and the analysis of Target Language Use (TLU), as well as its practical limitations. The ensuing argument will have implications for both achievement and proficiency testing, with a focus on the real-life language use environment. A rough sketch is presented at the end as an operational counterpart to the theoretical aspect of authenticity.



## 1. Introduction

During the past 40 years, the field of language testing has been influenced by many changes in the orientation of interrelated fields. In linguistics, the move from structuralism through Chomskyan generative linguistics to minimalism, with further reactionist moves to cognitive linguistics, has urged testing practitioners to shift their focus. Also, there have been trends in sociolinguistics, with extralinguistic factors being integrated into language testing. In applied linguistics, there have been many shifts from teaching formal structure and explicit instruction to the communicative methods of teaching, as well as other shifts taking them away from the methods' obsession. Studies on L2 acquisition have also revealed the hidden aspects of learning a second language and its impact on the testing enterprise. An influential challenge was articulated by Canale (1984), who believed that the shift of focus into language use would place new demands on language teaching, with bearings on language testing formats, appropriateness rules, administration procedures to emphasize interpersonal interest in authentic situations and new scoring procedures (p. 79). All these have guided the meandering route of language testing from very rudimentary steps to highly sophisticated tests to accommodate for new findings in the interdisciplinary areas.

However, materials writers such as Close (1965) and Broughton (1965) warned that language learners were being exposed to texts that were not representative of the target language they were learning. However, soon there were challenges to the dichotomous definition of authenticity, i.e. texts being either authentic or inauthentic; accordingly other aspects such as the learners, the teacher, and the situation of teaching were taken into consideration (Shomoossi & Ketabi, 2007). It was not until the late 1970s that Widdowson (1978) initiated a debate on the nature of authenticity in relation with genuineness (p. 80). Although his arguments were flawed as it confined the issue to the role of native speakers only, it can be considered as a turning point for the debate in that he believed that genuine texts would only be considered authentic after undergoing a process of *authentication*, a process which, he suggested, may only be accessible to native speakers. But he failed to account for the way *language learners* could process towards being able to authenticate texts. However, he emphasized the importance of the interactions between the audience and the text, and hence the nature of the outcome arising from textual input. Consequently, there was a rush towards authentic input, and towards texts which were not simplified, and tasks felt to be simulating real-life performances. Shomoossi and Ketabi (2007) recommended the development of a pragmatic knowledge and teacher professionalism as the cornerstone of curricular revolution. In short, this



pragmatic phenomenon, influenced by contextual factors, teachers' active and authenticating role, and students' interaction with language, is still in contrast with the validity of bit-by-bit testing knowledge of language, but stresses one of the most important considerations in language testing.

Under the influence of the classical psychometric theory, we were taught to think of basic characteristics of the test quality based on the criteria such as validity, reliability and practicality. However, there was a general agreement among specialists on the centrality of validity as the one most necessary characteristic. This conception has been so strong that the recent Weir publication (2005: 11-16) has prioritized validity in his model as the dominant concept, while other test characteristics are listed as evidence to validity. As mentioned earlier, extralinguistic factors gained a place in the language testing due to the emphasis on the *use* of language, which is concerned with the correspondence between the test and non-test situations. This concern came to be known as *authenticity*, which aimed at achieving a close correlation between the test performance and the criterion performance, and thus became the main focus of this paper.

## **2. Authenticity**

Cumming and Maxwell (1999: 178) attribute the first formal use of the term 'authentic' in the context of language learning and assessment to Archibald and Newman (1988). However, there has long been an agreement over the idea that authenticity is an important quality for the test development (Lynch, 1982: 11). Morrow (1991: 112) points to the overriding importance of authenticity, and Wood (1993: 233) considers it as one of the most important issues in language testing. Also, Bachman and Palmer (1996) see authenticity as a critical quality of language tests (p. 23). Authenticity is pivotal to Douglas' (2000: 16-18) consideration of LSP testing as one of the two features distinguishing LSP from more general tests of language (the other being the interaction between language knowledge and content knowledge).

Other authors have stressed the importance of authenticity as one of the decisive characteristics of a good test (e.g. Carroll, 1980; Bachman, 1990; Bachman & Palmer, 1996; Weir, 2005). They derive the idea from the widely accepted notion of validity and relate it to the validity of testees' *future performance in real-life situations*. Focusing on the naturalness of the test tasks, Carroll (1980: 11) emphasized that all test tasks should sound real-life, interactive and communicative, rather than being typical routine examination responses to the tester's stimuli; that the language of the test should be the day-to-day discourse; and that the rating of a performance

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

should be based on its effectiveness and adequacy as a communicative response and rely on non-verbal as well as verbal criteria. The elements stressed by Carroll are included and developed in Bachman and Palmer (1996: 49-50) in the form of a framework to ensure the authenticity of a test task. This framework elaborates on the characteristics of the setting, the test rubrics, the input, the expected response and relationship between the input and the response to a given task.

Toward the end of the 1970s, there was an inevitable propensity towards *communicative use testing* and scholars felt that authentic stimulus material was a necessary component of any test of communicative ability. For instance, simulations of real-life tasks became part of direct tests of spoken ability and ESP in the British Council ELTS test battery. However, equating authenticity with texts gave rise to some concerns. First, a dichotomy was created between authentic and inauthentic texts (the former considered better, and the latter inferior). Second, authentic tasks were considered to be those mirroring real-life tasks, but they did not give rise to genuine interaction because they were simulations and not real tasks. In addition, real-life holistic tasks did not lend themselves to test situations. Also, the question of task selection in order to make generalizations to non-test performance was not adequately resolved yet. Morrow (1991), however, divided tasks into enabling tasks and tried to generalize the results on the basis of these skills.

Throughout the 1980s, the debate focused on the authenticity of *input* with scant regard to the role of test takers in processing the input. Testers were obsessed with the models of communicative competence (see Canale & Swain, 1980) as an overall aim of instruction. Bachman (1990: 87) added the idea of communicative language ability (CLA) as a measure of mastery of a language. The consequence to both perspectives is that language testing is automatically tailored to these models. This approach, which tried to minimize the gap between instruction and testing, was called a *curricular approach to testing* (Doye, 1986). While it obviously failed to take the test takers' future performance into account, this approach seemed quite plausible and was – or is still – practiced for years. However, a serious threat to this approach is the incongruence or incompatibility of the test situation and the real-life situation where the learner is supposed to master via the curriculum. One of the earliest articulations of this concern is made by Carroll (1961) who distinguished between integrative and discrete-point approaches to testing (p. 37). For Carroll, an approach requiring an integrated facile performance on the part of the examinee was one of priority. He recommends tests in which less attention is paid to specific structure-points or lexicon

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

than to the total communicative effect of an utterance. Various terms have also been developed to describe the concept, e.g. direct testing, performance testing, functional testing, communicative tests and authentic assessment, among others. The concept gained such a significance that an international conference (1984) was totally dedicated to the issue, and one issue of the journal *Language Testing* (2, 1, June 1985) was given the duty of publishing the papers presented in that conference (Bachman, 1990: 301). Further to this, Spolsky (1985) stressed the importance of authenticity by raising important pragmatic and ethical questions in language testing. He warned about the generalizeability of results if authenticity could not be taken into account. Since then, sporadic attempts have raised the issue to the platform of discussion (e.g. Carroll, 1980; Doye, 1986; Bachman, 1990; Bachman & Palmer, 1996; Lewkowicz, 2000; Leung & Lewkowicz, 2006).

**2.1 Materialization of Authenticity**

Options are suggested to operationalize authenticity in language testing. One such suggestion could be putting the test takers in direct testing situations, and observing how well they perform the required task, and assessing their performance quantitatively or qualitatively. However, this method is both time-consuming and impractical. Language testers, therefore, tried to find economical and practical ways of assessing performance. They found that samples of the test takers' future performance could be tested in simulated tests. After analyzing the target situation behavior required from the testees, they designed test tasks which easily lent themselves to testing in classroom or other testing situations. However, the degree to which the two aspects – test task and target situation task – coincide or resemble is of paramount importance in this regard. Doye (1986) provides an example of such tests by referring to the ways that a language learner can demonstrate performing the speech act of “Asking the way in an English speaking environment”. He portrays two possible ways: by taking the learner to an English speaking town (which is hardly ever feasible), letting him/ her find the way and assessing the performance on the basis of the result; and inventing a simulated but realistic situation, letting the learner perform the task, and assessing it on the spot.

**2.2 Authentic tests**



Authentic tests are often regarded as synonymous with 'communicative' tests, 'direct' tests, and 'performance' tests, etc. (Jian-lan, 2007). Lewkowicz (2000) believes that despite the importance of authenticity, there hasn't been a marked body of research to demonstrate the characteristic. While authenticity is important for assessment theorists, this may not apply to other stakeholders in the testing enterprise. For instance, perceptions of testees from the notion, and whether the presence or absence of authenticity affects test takers' performance, are not clear (Lewkowicz, 2000). On the contrary, Bachman and Palmer (1996) stress the potential effect of authenticity on test takers' performance while the mechanism is elaborated neither by explanation nor by research evidence.

Other researchers hold different views on authentic language tests. From the very early references to authenticity (e.g. Carroll, 1961; Close, 1965), real-life performance has been at the heart of the debate. Accordingly, a real-life approach to defining authenticity emerged (Bachman, 1990: 301) which essentially considered the extent to which test performance replicated non-test language performance. Its primary concerns were: (1) the appearance or perception of the test and how things may affect test performance and test use (or the so-called *face* validity), and (2) the accuracy with which test performance predicts future non-test performance (or the predictive validity). This approach was dominant throughout the 1970s and 1980s, particularly in testing oral proficiency in a foreign language, and its proponents have considerably contributed a lot to our understanding of the characteristics and uses of tests which attempt to mirror reality. Two main approaches to authenticity originated from these considerations, and we will discuss them below.

### **3. Approaches to authenticity**

#### **3.1 The Real-Life (RL) Approach**

This approach began with the need to assess oral proficiency of learners in the 1970s. It defines language proficiency as the ability to perform language tasks in non-test situations, and authenticity as the extent to which test tasks replicate real-life language use tasks. Both proficiency and achievement tests are discussed in this approach but the major focus is on the proficiency tests. In fact, proficiency and authenticity are effectively synonymous (Bachman, 1990). Terms most frequently used by the proponents of this approach to characterize authentic tests are *direct* and *performance-based*. Clark (1978, 1987) portrays the distinction between direct and indirect tests of proficiency as follows:

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

In direct proficiency testing, the testing format and procedure attempt to duplicate as closely as possible the setting and operation of real-life situations in which the proficiency is normally demonstrated (Clark, 1978: 10). However, indirect measures are not required to reflect authentic language-use contexts, and in many cases they bear little formal resemblance to linguistic situations that the student would encounter in real life (Clark, 1987: 26).

However, replicating non-test or real-life performance in language tests was recognized as a difficult enterprise even by the proponents (e.g. Clark, 1978, 1987). Therefore, they adopted the concept of a continuum for direct-indirect testing, characterized by the extent to which test performance replicates non-test language use. However, there have been doubts on the possibility of full replicability of real-life non-test performance, or even its coming close to real life. This approach thus seeks to develop tests which strive to resemble as closely as possible the real-life language performance. Bachman (1990) lists three interrelated tenets that characterize the RL approach:

- (1) A view of language ability, or proficiency, as pragmatic ascription,
- (2) The reference to real life performance as a criterion, and
- (3) The belief that face validity, content relevance, and predictive utility are sufficient bases to justify test use (p. 303).

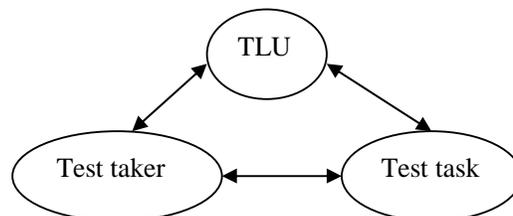
The RL approach has dominated testing especially in testing oral proficiency in a foreign language for more than two decades and is still dominant in most testing practices. Interestingly, validity in the RL approach is nearly identical with authenticity. However, some researchers argue that the RL approach provides an inadequate basis for examining validity and fails to distinguish *ability* from *behavior*. A second criticism is that the RL approach provides an inadequate basis for reexamining validity (Bachman, 1990: 308). Since arguments supporting test appearance, content relevance and predictive utility do not by themselves provide sufficient evidence to justify test use, test validation is seriously threatened in this approach. However, despite serious criticisms against this approach, critics did not provide a viable alternative. But there have been attempts to solve this problem, e.g. attempts to (1) accept real-life as the criterion for authenticity and modify methods of testing (which is difficult to operationalize), and (2) accept that the test language is, by nature, inauthentic and distanced from the real life language. However, what was lacking in the equation was a theoretical framework to provide a coherent rationale in identifying and defining the critical features of



language use, either in test or non-test contexts. Many researchers, however, have contributed to the construction of such a framework known as the *interactive language use* – and this serves as a lead up to the second approach.

### 3.2 The Interactional Ability (IA) Approach

In this model, authenticity is a function of the interaction between the test taker and the test task (Bachman, 1990: 317). One major difference between this approach and the RL approach to authenticity lies in the way we operationalize the concept. Bachman's (1990) theoretical framework of the communicative language ability or CLA (p. 85) and test method facets (p. 119) can be considered as the basis for this approach. In short, the IA approach views authenticity as residing in the *interaction* between the test taker, the test task, and the testing context. Indeed, the primary concern is to construct test tasks that reflect our knowledge of the nature of language abilities and language use.



This approach, advocated by Bachman and Palmer (1996), focuses on the interaction between the language user, the context, and discourse. In this regard, test authenticity is a notion that upholds an interactive dealing between the test taker, the test task and the *Target Language Use* (TLU) domain, and it becomes essentially synonymous with communicative language use. The IA approach is based mainly on measuring language as a *mental ability* and uses a theoretical framework of factors affecting test performance to construct tests (Bachman, 1990: 85). In addition, the IA approach also considers the *abilities of the test taker*. The ability part of this approach originates in the early theories of verbal ability, and is manifested in various forms from Lado (1961), Carroll (1961), and Oller's (1981) *pragmatic expectancy grammar* to Kramsch's (1986) *interactional competence* (all cited in Bachman, 1990: 302).



Rather than relying on non-test language performance as a criterion, the IA approach focuses on what it sees as the distinguishing characteristic of communicative language use – i.e. the interactions between the language user, the context and the test discourse. Therefore, it attempts to design tests that will involve the test taker in the appropriate expression and interpretation of the extent to which the test taker possesses various communicative language abilities, and there is a clear distinction in this approach between the *abilities* to be measured, on the one hand, and the *performance* we observe and the context in which the observations take place, on the other. While the proponents of this approach recognize the importance of the way test takers and test users perceive the test, their primary concern is with demonstrating the extent to which test performance reflects language abilities, i.e. construct validity.

Bachman (1991) divided authenticity into *Situational Authenticity* (the perceived match between the characteristics of test tasks and TLU tasks) and *Interactional Authenticity* (the interaction between the test taker and the test task). In so doing, he acknowledged that authenticity involved more than matching TLU tasks; he saw authenticity also as a quality arising from the test takers' involvement in test tasks. Like Breen (1985), Bachman (1990) recognized the complexities of authenticity and avoided considering it as an absolute quality. For instance, he believed that a test task may be situationally authentic but interactionally less so, or vice versa. Similarly, Douglas (2000: 49) considered these two aspects of authenticity important and vital to LSP tests.

Later on, Bachman and Palmer (1996) separated the notion of authenticity from interactiveness, defining authenticity as the degree of correspondence of the characteristics of a given language task to the features of a TLU task – the same as the situational authenticity in Bachman (1991). They also replaced interactiveness for interactional authenticity – as proposed in Bachman (1991). To approximate the degree of correspondence between the test and TLU tasks – i.e. to determine the authenticity of test tasks – they proposed a framework in Bachman and Palmer (1996: 49-50), as a checklist of task characteristics including the input provided in the test as well as the expected outcome arising from the input by characterizing not only test tasks but also test takers' interaction with these.

#### **4. Reconceptualization of Authenticity**

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

According to Bachman (1990), authenticity is to be viewed in terms of a relationship between features of the test and those of the non-test target-use context. The first approach towards authenticity (the RL approach) tries to develop tests that mirror the 'reality' of non-test language use. This approach appears to be rather naive as test settings often do not exactly resemble the real-life setting (Spolsky, 1985). Also, it does not distinguish between language ability and the context in which this ability is observed, since non-test language performance constitutes the criterion for authenticity and the definition of proficiency (Bachman, 1990: 302). In the second approach (the IA approach), the authenticity of language tests arises from the 'situational' and the 'interactional' authenticity. Thus, the emphasis in this model shifts from 'attempting to sample actual instances of non-test language use to that of determining what combination of test method facets is likely to promote an appropriate interaction of a particular group of test takers with the testing context' (Bachman, 1990: 317). However, both approaches to authenticity are concerned with the context and the manner in which we elicit a sample of performance – with the characteristics of the testing methods we use. Therefore, we need to describe the *facets that characterize the test method*. In other words, just as we must include the features of the language use contexts in a description of non-test communicative language use, so our description of performance on a language test must include test method facets.

The optimal goal in the description of authenticity, therefore, may appear as one that bridges the gap between the test situation and non-test situation so as to ensure the construct validity of tests. An example may help clarify the mutual relationship between *authenticity* and *validity* as the two essential characteristics of tests. For instance, language competence is an invisible mental construct. The only evidence for testing the testees' competence is to test their performance, upon which consequent generalizations are made. The distance between the two invisible-visible components of the evaluation is the area of construct validation, which is far more important than a superficial treatment of the test task. Test developers, therefore, need to consider: (1) the test task (purpose, test methods, test formats, difficulty level, conceptual frameworks for testing, fairness and objectivity considerations), (2) the test taker (age, level, background, expectations, and perceptions), and (3) analysis of TLU (discourse type, formal/informal considerations, and expectations from the learner) (Bachman & Palmer, 1996). However, as the tester's knowledge of the learners' *competence* is based on *performance* samples rendering a partial image, and this is going to be generalized as their ultimate TLU performance in future, the generalization will be valid

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

if and only if the construct validation is successful. Therefore, the idea of introducing authenticity as one of the major characteristics of test uses is another emphasis on the role and centrality of validity over other test characteristics. Authenticity as the degree of correspondence between test task characteristics and those of the TLU tasks signify an intimate association with validity considerations.

**5. Authentic Assessment**

Since testing an isolated skill or a retained fact does not effectively measure a student's capabilities, evaluating what a student has learned requires examining their *collective abilities*. The term authentic assessment is synonymous with various forms of assessment that account for student learning, achievement, motivation, and attitudes on instructionally relevant classroom activities. Often, traditional types of assessment (i.e. essays, multiple-choice, fill-in-the-blank, etc.) are heavily language dependent. Surprisingly, *content* assessments have occasionally become English proficiency tests rather than a measure of what students know. In fact, authentic assessment refers to assessment tasks that resemble reading and writing in the real world and in school (Hiebert, Valencia & Afflerbach, 1994). Its aim is to assess different types of abilities that underpin literacy in contexts that are similar to actual situations in which those abilities are used. For example, authentic assessments may ask students to read real texts, to write for authentic purposes about meaningful topics, and to participate in authentic literacy tasks such as discussing books, keeping journals, writing letters, and revising a piece of writing until it works for the reader. Furthermore, authentic assessment values the thinking behind a work as well as the finished product (Pearson & Valencia, 1987; Wiggins, 1989; Wolf, 1989). It is mainly designed to engage the student in a simulation of a real-life problem that they must solve using the knowledge and skills they have gained in the course. A single project can be implemented to assess mastery of course content as well as language-oriented goals such as communication skills, learning and critical thinking skills, as well as social and educational values (Gabriel, 2005).

Working on authentic tasks can be beneficial in that it becomes an edifying experience of learning for the student (Wolf, 1989). From the teacher's standpoint, such tasks can go a long way in supporting students' language learning skills and strategies (Wiggins, 1989). Students learn and practice how to apply important knowledge and skills for authentic purposes. They should not

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

simply practice information recall or encircle isolated vowel sounds in words; they should apply what they know to new tasks. It is not difficult for us to see the pedagogic advantages of asking students to discuss why the author used particular metaphors and what effect they had on the story in a literary text rather than asking them to underline metaphors in it. Such an endeavor will require the students to use their knowledge and skills for a better understanding of how their learning relates to their living.

*Performance assessment* is a term commonly used with, or in place of, authentic assessment. It encourages students to demonstrate their knowledge, skills, and strategies by initiating a response (Rudner & Boston, 1994; Wiggins, 1989). Far from selecting an appropriate item from an array of multiple-choice options, students might indicate their abilities by conducting a research and writing a report, interviewing as well as being interviewed, leading a discussion, telling a favorite story, retelling a story or a lecture, or summarizing information in a writing or speaking prompt, and so on. Simply stated, it requires students to perform a task rather than take a test. It is, therefore, designed to judge students' abilities to use specific knowledge and skills and actively demonstrate what they know rather than recognize or recall answers to questions. It is sometimes called authentic assessment because it involves tasks in a real-life context or a context that simulates a real-life context. The protocols for performance assessment range from relatively short answers to long-term projects that require students to *present and defend* their work. These performances often demand students to do in higher-order thinking and to integrate many language skills. Consequently, some performance assessments become longer and increase in their complexities compared to traditional assessments. As such any complete assessment should have a balanced blend of longer performance assessments and shorter ones. O'Malley and Pierce (1996; pp. 9-32) contend that authentic assessment may include a variety of measures that can be adapted for different situations. Table 1 presents some examples of authentic assessments.

**Table 1 - Examples and descriptions of some authentic assessments***Adapted from O'Malley and Pierce (1996; pp. 9-32)*

Assessment	Description	Advantages
Oral Interviews	Teacher asks students questions about personal background, activities, readings, and interests	<input type="checkbox"/> Informal and relaxed context <input type="checkbox"/> Conducted over successive days with each student <input type="checkbox"/> Observations recorded on an interview guide
Story or Text Retelling	Students retell main ideas or selected details of text	<input type="checkbox"/> Oral report are produced <input type="checkbox"/> Can be scored on content or language components



## Authenticity and Authentication in Language Testing: an Operational Perspective

1. Nematullah Shomoossi & 2. Mansoor Tavakoli

	experienced through listening or reading	<input type="checkbox"/> Scored with a rubric or rating scale <input type="checkbox"/> Reading comprehension and language development can be determined
<b>Writing Samples</b>	Students generate narrative, expository, persuasive written output	<input type="checkbox"/> Written documents are produced <input type="checkbox"/> Can be scored on content or language components <input type="checkbox"/> Scored with a rubric or rating scale <input type="checkbox"/> Writing processes can be assessed
<b>Projects/ Exhibitions</b>	Students complete projects in content area, working individually or in pairs	<input type="checkbox"/> Students make formal presentation, written report, or both <input type="checkbox"/> Oral and written products and thinking skills are obtained <input type="checkbox"/> Scored with a rubric or rating scale
<b>Experiments/ Demonstrations</b>	Students complete experiment or demonstrate use of materials	<input type="checkbox"/> Students make oral presentation, written report, or both <input type="checkbox"/> Oral and written products and thinking skills are obtained <input type="checkbox"/> Scored with rubric or rating scale
<b>Constructed-Response Items</b>	Students respond in writing to open-ended questions	<input type="checkbox"/> Written reports are produced <input type="checkbox"/> Scored on substantive information and thinking skills <input type="checkbox"/> Scored with a rubric or rating scale
<b>Teacher Observations</b>	Teacher observes student attention, response to instructional materials, or interactions with other students	<input type="checkbox"/> Setting is classroom environment <input type="checkbox"/> Takes little time <input type="checkbox"/> Observations are recorded with anecdotal notes or rating scales
<b>Portfolios</b>	Focused collection of student work to show progress over time	<input type="checkbox"/> Integrates information from a number of sources <input type="checkbox"/> Gives overall picture of student performance and learning <input type="checkbox"/> Strong student involvement and commitment <input type="checkbox"/> Calls for student self-assessment

However, Wiggins (1998) suggests that an assessment is said to be authentic if it (1) is realistic, (2) requires judgment and innovation by requiring the student to use knowledge and skills wisely and effectively to solve problems, (3) simulates contexts that mirror the workplace or other real-life contexts, and (4) assesses the student's ability to efficiently and effectively use a repertoire of knowledge and skills to negotiate a complex task (pp. 22 -24).

### 5.1 Limitations to authenticity

Overemphasis on authenticity as a determining factor could also be considered dubious. However useful the postulation of authenticity as one criterion among others may be, it is to be borne in mind that (1) a complete congruence (of the test to target language use) is impossible in practice, and (2) there are other demands that necessarily influence our search for optimal forms of testing and therefore relativize our attempts to construct authentic tests (Doye, 1986). While a language test is a social event that has the intention of examining the competence of language learners, it is a special and formulized event *distanced from real life* and structured for a particular purpose. The very fact that a language test seeks to find out whether the learner is capable of performing a language task distinguishes it considerably from the corresponding performance of this task outside the test situation. Even if we succeed in manipulating the testees to accept the

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

illocutionary point of a test task they are supposed to perform, they will always have in mind the other illocutionary point that is inherent in a test, namely 'to prove that they are capable of doing what is demanded of them'. Even put in direct testing situations, learners will always have another purpose of their verbal activity in mind: 'to show the teacher that they are able to perform the demanded task well enough to satisfy the teacher and to qualify the course requirements'.

Also, there has been an emphasis on the specifics of real-life situations which are not included in the so-called inauthentic tests. We have to embed our tests in a realistic setting that contains all such specifics (e.g. background noises, hesitations, interruptions, etc.) so that the test looks plausible enough to the testee. However, the more we include those incidentals, the farther we move away from the reality since not all those peculiarities of the real life exist in all such situations. Therefore, we need not be sorry if we do not succeed in making a test situation *absolutely authentic*. Rather, we should endeavor to employ just the plausible-looking amount of realism in the construction of our tests. In other words, abstraction from those incidentals may seem inevitable for economic and practical reasons or purposes. Also, Weir (2005) contends that full authenticity of setting is obviously not attainable in the classroom or the language test, but the settings selected for testing and teaching should be made as realistic as possible (p. 56). Similarly, Douglas (2000) and O'Sullivan (2006) believe that every attempt should be made within the constraints of the test situation to approximate to situational authenticity.

Now let's take an example in this regard. Suppose you have asked the learners to 'write an answer to a faxed message to the company X on why the delivery of goods has been late'. You do not have to take all test takers to the company X, to show them the message, to let them use the resources there, to consult the manager, to enquire the delivery department, to finally have the reply proofread by the secretary and fax the message under office regulations. All you need is a sample of their behavior (a written reply in this case) which can indicate that the testee is able to put the words together to make a convincing answer to the party who has complained. Therefore, a degree of *abstraction from reality* is inevitable, the amount of which is a matter of controversy. But the degree of abstraction is controversial. In normal communication, a small number of essential features and a great number of incidentals are observed which differ from one context to another. If we want to grasp the essential features of a task, we have to abstract from the incidentals. Abstraction, from this perspective, is a threat and a counterpoint to authenticity in testing. What is needed is the right balance between *authenticity* and *abstraction*, which can be guaranteed with a

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

careful examination of the target situations with the help of studies from pragmatics, discourse analysis, interactive sociolinguists, etc.

Bachman (1990: 112) adds a new dimension to these limitations and contends that the way test method factors are designed and controlled, and the correspondence between these and features of the language use contexts will have a direct effect on the authenticity of the test and test task. While he stresses test methods rather than the existence of such a correspondence, he considers that test methods characteristics are the restricted or controlled versions of those contextual features that determine the nature of the language performance expected for a given test or test task. If tasks are designed carefully, it will be easier to compare the performance of different students and to improve reliability in scoring.

Now let's consider more examples of what will be required of students in future target situations. For students in general English courses, functional tasks can be assigned on the basis of what they have been exposed to in their language courses. For medical doctors, we may assign tasks which require them to write a letter to a local GP about a patient on the basis of a set of printed case notes. For a student in EAP context, it might involve scanning (search reading) of an academic text or preferably texts to extract specified information for use in a written summary, or describing information contained in a diagrammatic form. For those EFL students at a secondary school level, it might involve responding to a letter or writing a paragraph from a wall chart in the class (Weir, 2005).

However, while Weir (2005) believes that 'direct writing tasks' offer a more construct-valid approach and are close to the real-life academic, social and service tasks, Hyland (2002) warns us of the potential problems of contextual validity in direct writing tasks such as TWE and IELTS. These tasks are often based on a brief, timed response to one or two topics. The problem is that these tasks provide little information about the future performance of students to provide a sustained piece of writing for different audiences or purposes. One of the main reasons oral interviews and compositions are not widely used, at least with achievement tests, is that they are very time-consuming, both to administer and to score. Despite the fact that most of us would agree that they can involve authentic language use, considerations of efficiency often take precedence over those of validity and authenticity. However, these types of tasks became popular with IELTS and TOEFL mainly for two reasons: generalizability and consequential validity (Bachman, 1990: 298). The inclusion of such test types are justified by their prediction of the testee's future performance in two

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

major production areas, i.e. speaking and writing, while their scoring and administration are not easy. In fact, what Weigle (2002) proposes is the superiority of portfolios to the timed essays in terms of authenticity. Portfolios can be designed to collect samples of student writing which were written for purposes other than teacher evaluation, e.g. papers written for academic courses. Portfolios are at the high ends of interactiveness too. Especially the act of collecting, selecting and arranging the contents involves the metacognitive strategies to a considerable extent and involves investment on the part of the student, i.e. the portfolio author (Weigle, 2002: 203-4).

**5.2 Reconsideration of Bachman and Palmer's Model**

Authenticity depends, to a great extent, on our view of the language ability since such a model affects our treating the elements of authenticity. For instance, in Bachman and Palmer's (1996) model, authenticity consists of three elements: the test task, the test taker and the TLU domain. Over the past 30 years or so, language ability and language use have been reconceptualized by theorists such as Savignon (1972, 1983), Canale and Swain (1980), Canale (1983, 1984), and Bachman (1990). Thus the foundations for the communicative approach were laid, and the debate on authentic testing was subsequently raised. Initial attempts reflected the real-life approach to designing tests mirroring the real-life resources and situations. Authenticity was, therefore, seen as inherent in the test – either present or absent – and no regard was paid to the interaction that would arise between the test taker and the test input. Built on the earlier models such as Hymes (1972), Canale and Swain (1980) among others, Bachman's (1990) multicomponential model proposed an interactional model of language test performance. It has provided, since then, a principled basis for the development of language tests. However, the major criticism against it is that it is extremely difficult to operationalize (Leung & Lewkowicz, 2006).

Leung and Lewkowicz (2006) contend that the debate on test authenticity and test usefulness has left a number of questions unanswered. It has failed, they claim, to adequately address two persistent problems, which relate to the multifaceted nature of authentic testing. First, *performance tests* are often extremely complex. For example, McNamara (1995) and Alderson and Banerjee (2002) point out that Bachman's (1990) model cannot account for the social aspects of language performance, e.g. the relationship between the test taker and other test takers, between the tester and test taker(s), as well as the personality features of the interlocutors (in oral assessment) and test takers. Indeed, the fact that such oral performances are essentially co-constructed through

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

social interaction, and that such interactions are likely to affect individual performances are often overlooked in the research (Luoma, 2004). The second problem, also acknowledged by Bachman (2002), is associated with the inability of this model to account for *task difficulty*. Currently, there is little agreement about how to control for this aspect of language in test situations, and it is hoped that this unattended area will be enriched by the use of IRT and Rasch model studies. Current models consider test difficulty to be ‘essentially an artifact of test performance’ (Bachman, 2002: 453), seemingly inseparable from test taker and test task characteristics.

**6. Authentication versus Authenticity**

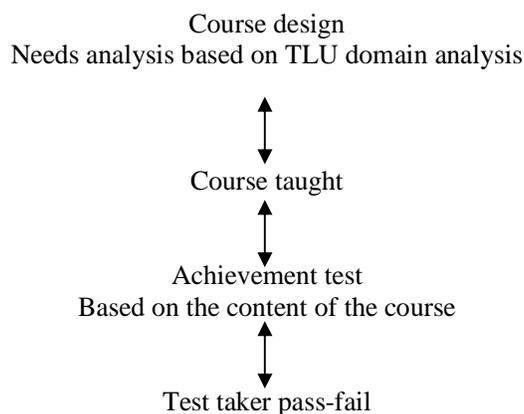
Having reviewed the history of, and research in, authenticity in language testing, the most sought-after model seems to be Bachman and Palmer’s (1996) model, which is associated with some problems mentioned earlier in this article. Despite the widespread use of this model, it appears to have much less impact on *achievement testing* within the classroom than on large-scale proficiency testing.

The idea of viewing the testee either as a *learner* (as in classroom testing) or a mere *test taker* (as in proficiency testing) is the area where the highest possible discrimination occurs. In the former, authenticity focuses more often on the correspondence between the course content and the test content, which addresses the fairness demands of the test takers, rather than the one-to-one correspondence between the test task and TLU features. As far as the classroom testing is concerned, matching TLU with the course content is expected to be done at the stage of *course design* through *needs analysis*. Course designers and curriculum developers analyze the target language use needs of the learners first, and then attempt to apply the needs into the teaching modules. Therefore, the authenticity of testing is indirectly treated. The reason for this is that the knowledge of language can be assessed through *separate* modules, skills, and tasks but the final outcome is collectively considered as an estimate of one's language competence. Also, what may account for the test takers in achievement tests is not how much the test task and the TLU features match; rather, what is important is the correspondence between the course content and the test tasks. It is likely that test takers trust the teacher – simultaneously the tester too – in that the students’ concern is merely on taking the test successfully. It is also possible that they do not even have a realistic conception of what they are going to do, or what they will be required to do since they are merely students and a clear picture of their career expectations is not conceivable. Also, Sook



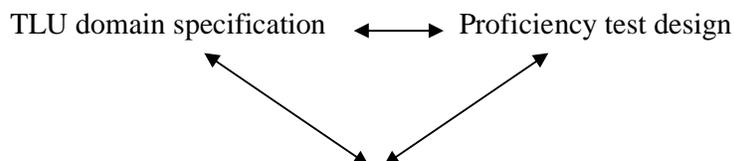
(2003) contends that assessment can be used to improve instruction and to help students take control of their learning. Therefore, it is necessary to focus on its backwash effect too. Figure 1 diagrammatically presents this argumentation.

**Figure 1 – A tentative diagram of Authentication in Achievement Testing**



However, in the proficiency testing, the relationship is more direct in the sense that the testee is expected to fulfill the TLU requirements through test tasks which are, in the main, simulations of the TLU tasks. Although the test task might not include all the incidentals of the TLU domain, the idea is that if the testee can pass the test, it means that she can manage the expected tasks in future TLU assignments. Almost in all cases, proficiency test takers know that they are required to take the test as a *sign* of guarantee for their future performance, and this may emphasize the significance of consequential validity (Weir, 2005: 37). For instance, in IELTS, two modules are designed: the General Training Module for those who apply for technical professions in the English speaking countries; and an Academic Module for those who apply for university seats or other academic vacancies. Therefore, addressing the TLU domain will be of paramount importance for the test takers; if the correspondence is violated, the test takers would often consider the test to be unfair and not valid. Figure 2 below presents the relationship in diagram.

**Figure 2 – A tentative diagram of Authentication in Proficiency Testing**





Test task Expectations

Test taker  
Qualified-Disqualified

This takes us back to the consideration of the relationship between language ability and how that ability is assessed, or in Alderson and Banerjee's (2002) words, 'what language is and what it takes to learn a language, which then becomes the basis for establishing ways of assessing people's ability' (p. 80). Therefore, what is important is to view authenticity as an integral aspect of test usefulness. Perhaps test developers should foreground the notion of construct validity, that is, what testers are trying to measure (Leung & Lewkowicz, 2006).

Another aspect to the issue of congruence is that in many cases, test designers often ignore the correspondence in favor of practicality, face validity and cost-effectiveness. Authentic assessment is a difficult task by itself since there are many factors affecting the outcome – the test task. A desirable balance between all these factors would be really what is wanted. Lewkowicz (2000) believes that correspondence is to be adjusted at the moderation stage in the sense that the relationship between the test taker and test task might need to be strengthened instead of over-reliance on the total congruence between the test task and the TLU domain specifications.

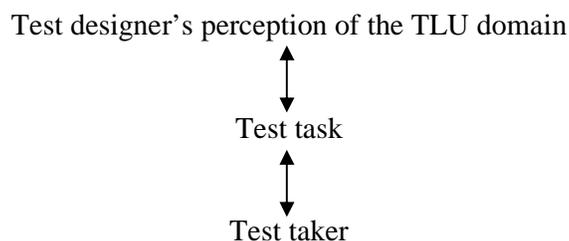
Moving a step farther from the purpose of testing (either classroom or proficiency testing), the characteristics of the test taker could be considered more critical than those of the other two. In one case, for instance, Lewkowicz (2000) found that test takers tended to be very pragmatic and more concerned with the test's difficulty rather than its authenticity. In other words, despite the alluring acceptability of Bachman's model, the connecting links (arrows in Figures 1 and 2) do not, in fact, plausibly demonstrate what the nature of such relations could be. The test taker characteristics require a more serious treatment especially in the era of globalization where English is becoming the global lingua franca – if not the international language (Shomoossi & Ketabi, 2008). EIL speakers are not a homogenous population and one single test cannot be designed to measure their proficiency (Iyldyz, 2007). Nunn (2007), for instance, portrays a different competence framework for the EIL speakers and users. Accordingly, the TLU domain will be



different from the ESL / EFL situations, and users will be required to fulfill their tasks quite differently.

A more serious issue in this model is the relationship between the test taker and the test task, which is not dealt with attentively in the research, except for Lewkowicz (2000) and Leung and Lewkowicz (2006) who have stressed the test takers' *perception* as an indication of authenticity of the test task. Other ways could be conceived for data elicitation; for instance, content analysis, expert judgment, think-aloud protocols, online introspective and retrospective techniques can be used to delve into the hidden aspects of their perception. On the other hand, the test taker and her conception of the TLU demands are often overlooked in favor of the simplicity of the research technique. In fact, this relationship is indirectly, and in a linear order, treated in the real world. In other words, the test designer's perception of the TLU domain – through experience, needs analysis, contemplation or whatsoever – is the basis for the test takers' conception of the TLU demands (See Figure 3).

**Figure 3 – Authenticity Perceptions of Test-takers and Test designers**



In fact, test takers are always in direct contact with the test task. The other major component in the Bachman's (1990) model, i.e. the TLU domain, and its relationship with the test task and the test taker only indirectly relate to the test taker. However, it does not mean that the latter is not a significant element in the model. Rather, it is precisely the starting point for other stakeholders to enter the scene.

The major flaw of the model could be its reliance on the test takers' perception of authenticity. In other words, if authenticity is a matter of perception – and accordingly a subjective issue – how can we ensure that it is realized in one test rather than in the other? How can we ensure that what degree of authenticity is realized in the test? And how is it possible to determine the degree of authenticity which is appropriate for a specific group of testees and a specific situation? Do test takers care about such things? Do they have the right knowledge to judge that? Are they

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

honest with test designers? Isn't it the case that what they are given as the test task, they have to succeed in doing it? Otherwise they will fail the test (which can be either less authentic or more authentic)? Aren't the testers more powerful than the testees to exert their power on them by not observing authenticity? If they perceive the test to be inauthentic or less authentic, what would the result be? Will the testers be willing to make a change to the test usefulness?

Another aspect of the test takers' perception may be the effect of course elements on their achievement in the end-of-course tests. For instance, the effect of teacher's personality, gender and teaching, the ease or difficulty of the course, the relevance of the course to the learner's ultimate goal, the motivation to perform the test task, and many other factors can influence their perception of authenticity in one way or another. Also, there might be a mismatch between the test designer's, the teacher's and the policy maker's perception of authenticity with that of the test takers. While the former three types of perceptions can result from scholarly expertise, the latter can only be a superficial impression simultaneously affected by factors external to the test task. What would be required is the 'authenticator'. And the role of that authenticator would not be then finding the right balance between all these elements – the test task, the test taker, and the TLU domain specification. Although the operationalization of this simple equation, (i.e. Bachman and Palmer's model), is not simple, the role of an *authenticator* may be assigned to multiple users provided that their data fit into the other's model. In other words, *authentication* is not a single-user business and all stakeholders may try having a hand in it.

## **7. Conclusion**

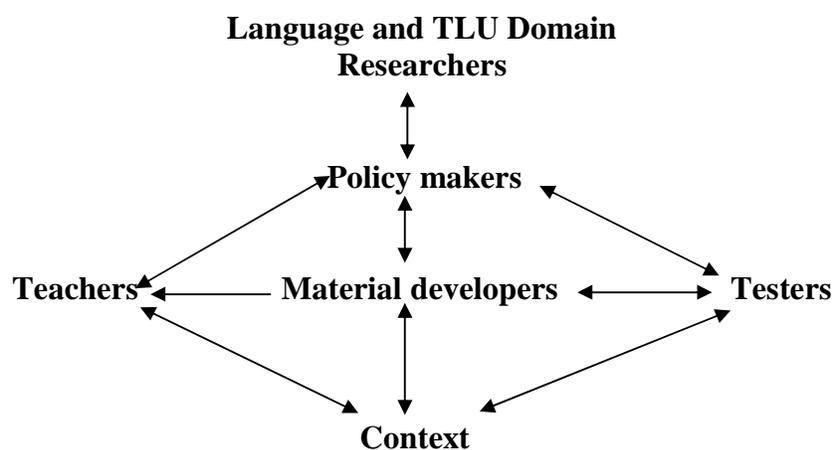
Having reviewed the story of authenticity, we might come to the conclusion that the authenticity as a test feature may be sidelined by a more important notion in test validity, i.e. the authentication process. In other words, the theoretical model of test authenticity (Bachman & Palmer, 1996) may require a dynamic and operational model to include the human elements and all stages of designing authentic test tasks, which may not be considered as the end point of the testing process.

The first stage may involve the participation of the researchers of the interdisciplinary fields, i.e. applied linguists, sociolinguists, discourse analysts, in order to contribute to the process via their analysis of the TLU domain and of what a prospective testee may require in order to fulfill the demanded tasks in the real world. The second stage may involve the role of material developers,



curriculum developers and syllabus designers to invest on the findings from the first stage, and to design courses and material on the basis of models of language ability which suit the target situation as well as the characteristics of the test takers (taking either achievement tests or proficiency tests). This stage is important due to its contribution to the construction of the relevant competence in the minds of testees as well as the model of competence to be tested. The third stage may be the practitioners' stage, which can be considered in two divisions regarding their roles: teachers (who will be testers too at the end of the achievement courses) and testers (whose main concern is assessing testees' proficiency) (See Figure 4).

**Figure 4 –A rough Sketch for the Authentication Process**



Teachers will be more concerned with the course material. Although successful teachers are informed by the literature of their field, their knowledge of the earlier stages can be considered subsidiary and optional. But their awareness of the contextual factors, i.e. the learners' characteristics and background (who will be the ultimate testees), the nature of the teaching material (which will be the basis of testing at the end of the term), teaching and testing techniques and their congruence, test design procedures, scoring and interpretation and use of the test results, will be crucial to the success of their pedagogical endeavor.

Testers, whose main concern is testing testees' proficiency, on the other hand, should be aware of contextual factors of a different nature. In other words, these factors may include their comprehensive knowledge of test construction stages, as well as their cooperation with policy

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

makers and test users (who are, in the main, a body of powerful experts or non-experts deciding on the fate of powerless testees).

Overall, this rough sketch may be suggested as the operational side to Bachman and Palmer's (1996) model of authenticity. However, further details and elaboration might require extended articles and research backup. Also, this might raise criticisms and scholarly comments to enrich the discussion.

## References

- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35: 79-113.
- Archibald, D. A., & Newman, F. M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4): 671-704.
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4): 453-476.
- Bachman, L. F., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Breen, M. P. (1985). Authenticity in the language classroom. *Applied Linguistics*, 6: 60-70.
- Broughton, G. (1965). *A technical reader for advanced students*. London: Macmillan.
- Canale, M. (1983). On some dimensions of language proficiency. In Oller J. W. Jr. (ed.). *Issues in language testing research*. Rowley, Mass.: Newbury House.
- Canale, M. (1984). Testing in a communicative approach. In Gilbert A. Jarvis (ed.). *The challenge for excellence in foreign language education*. (pp. 79-92). Middlebury, Vt.: The Northeast Conference Organization.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1): 1-47.
- Carroll, B. J. (1961). *Fundamental considerations in testing for English language proficiency of foreign students. Testing the English proficiency of foreign students*. Washington, DC: Center for Applied Linguistics. (pp. 30-40).
- Carroll, B. J. (1980). *Testing communicative performance*. Oxford: Pergamon Press.
- Clark, J. L. D. (1978). Theoretical and technical considerations in oral proficiency testing. In R. L. Jones & B. Spolsky. *Testing language proficiency*. (pp. 10-28). Arlington, VA: Center for Applied Linguistics.

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

Clark, J. L. D. (1987). *Direct testing of speaking proficiency: theory and application*. Princeton, NJ: Educational Testing Service.

Close, R. A. (1965). *The English we use for science*. London: Longman.

Cumming, J. J., & Maxwell, G. S. (1999). Contextualizing authentic assessment. *Assessment in Education*, 6(2): 177-94.

Douglas, D. (2000). *Assessing language for specific purposes: theory and practice*. Cambridge: Cambridge University Press.

Doye, P. (1986). *Authenticity in foreign language testing*. Online document retrieved January 15<sup>th</sup>, 2008 from [http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?\\_nfpb=true&\\_ERICExtSearch\\_SearchValue\\_0=ED350819&ERICExtSearch\\_SearchType\\_0=no&accno=ED350819](http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED350819&ERICExtSearch_SearchType_0=no&accno=ED350819)

Gabriel, G. (2005). *Performance assessment, authentic assessment, and primary trait analysis*. Online document retrieved January 15<sup>th</sup>, 2008 from <http://www.nvcc.edu/about-nova/directories--offices/administrative-offices/assessment/resources/performance/index.html>

Hiebert, E. H., Valencia, S. W., & Afflerbach, P. P. (1994). Authentic reading assessment: Definitions and perspectives. In S. W. Valencia, E. H. Hiebert, & P. P. Afflerbach (Eds.), *Authentic reading assignment: Practices and possibilities* (pp. 6-21). Newark, DE: International Reading Association.

Hyland, K. (2002). *Teaching and researching writing*. Harlow, UK: Longman/Pearson Education.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (eds.), *Sociolinguistics*. (pp. 269-93). Harmondsworth: Penguin.

Iyldyz, L. B. (2007). Rethinking validity of the L2 proficiency concept: Lessons for EIL. *EIL Journal*, 1 (1): 88-101.

Jian-lan, B. (2007). An analysis of authenticity in CET-4 and TEM-8. *Sino-US English Teaching*, 4(2): 28-33. Retrieved 20<sup>th</sup> January 2008 from <http://www.linguist.org.cn/doc/su200702/su20070206.pdf>.

Kramsch, C. (1986). From language proficiency to interactional competence? *Modern Language Journal*, 70(4): 366-71.

Lado, R. (1961). *Language testing*. New York: McGraw-Hill.

Leung, C. & Lewkowicz J. A. (2006). Expanding horizons and unresolved conundrums: language testing and assessment. *TESOL Quarterly*, 40(1): 211-234.

Lewkowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. *Language Testing*, 17: 43-64

Luoma, S. (2004). *Assessing speaking*. Cambridge, UK: Cambridge University Press.

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

- Lynch, A. J. (1982). Authenticity in language teaching: some implications for the design of listening materials. *British Journal of Language Teaching*, 20: 9-16.
- McNamara, T. (1995). Modeling performance: Opening Pandora's Box. *Applied Linguistics*, 16: 159-179.
- Morrow, K. (1991). Evaluating communicative tests. In S. Anivan (ed.). *Current developments in language testing*. (pp. 111-118). Singapore: SEAMO Regional Language Center.
- Nunn, R. (2007). Competence and teaching English as an international language. *EIL Journal*, 1(1): 11-24.
- Oller, J. (1981). Research on the measurement of affective variables: some remaining questions. In Andersen, R (ed.) *New dimensions in second language acquisition research*. (pp. 1-27). Newbury House, Rowley, Mass.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English learners: Practical approaches for teachers*. Reading, MA: Addison-Wesley.
- O'Sullivan, B. (2006). *Issues in testing business English: the BEC revision project*. Cambridge: Cambridge University Press
- Pearson, P. D., & Valencia, S. W. (1987). Assessment, accountability, and professional prerogative. In J. E. Readence & R. S. Baldwin (Eds.), *Research in literacy: Merging perspectives, Thirty-Sixth Yearbook of the National Reading Conference* (pp. 3-16). Rochester, NY: National Reading Conference.
- Rudner, L. M., & Boston, C. (1994). Performance assessment. *ERIC Review*, 3(1): 2-12.
- Savignon, S. J. (1972). *Communicative competence: an experiment in foreign language teaching*. Philadelphia, PA: Center for Curriculum Development.
- Savignon, S. J. (1983). *Communicative competence: Theory and classroom practice*. Reading, Mass.: Addison-Wesley Publishing Company.
- Shomoossi, N., & Ketabi, S. (2007). A Critical look at the concept of authenticity. *Electronic Journal of Foreign Language Teaching*, 4(1): 149-155. Retrieved online from <http://e-flt.nus.edu.sg/v4n12007/shomoossi.pdf>
- Shomoossi, N., & Ketabi, S. (2008). Authenticity within the EIL Paradigm. *Iranian Journal of Language Studies*, 2(2): 165-166. Retrieved online from <http://ijols.znu.ac.ir/volumes/volume2issue2/shomoosi3.pdf>
- Sook, K. H. (2003). The types of speaking assessment tasks used by Korean Junior Secondary school English teachers. *Asian EFL Journal: L2 Language Assessment in the Korean Classroom*. Retrieved January 2008 from [http://www.asian-e-fl-journal.com/dec\\_03\\_gl.pdf](http://www.asian-e-fl-journal.com/dec_03_gl.pdf)
- Spolsky, B. (1985). The limits of authenticity in language testing. *Language Testing*, 2(1): 31-40.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weir, C. (2005). *Language testing and validation: an evidence-based approach*. UK: Palgrave Macmillan.

**Authenticity and Authentication in Language Testing: an Operational Perspective****1. Nematullah Shomoossi & 2. Mansoor Tavakoli**

Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70 (9): 703-713.

Wiggins, G. (1998). *Educative assessment: designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.

Wolf, D. (1989). Portfolio assessment: sampling student work. *Educational Leadership*, 46 (7): 35-39.

Wood, R. (1993). *Assessment and testing*. Cambridge: Cambridge University Press.