



## **Problematizing English Language Assessment and Testing Systems in a Public Sector University in Saudi Arabia**

**Muhammad Wasim Latif**

*Jubail University College, Saudi Arabia*

### **Introduction**

Tests are employed as strong disciplinary tools because of their “ability to observe, perform surveillance, quantify, classify, normalize, judge, and punish” (Foucault, 1979 as cited in Shohamy, 1993). Thus, the students’ scores in tests serve as powerful means to pass important judgments, which have serious ramifications in terms of decisions regarding their future. This is reflected in Noam’s (1996) words “How we assess can support careers and make people successful, but it can also destroy peoples’ careers, place unfair burdens on individuals’ self- perception and unnecessary hurdles in the path of their achievement (p. 9).

Since “testing is never a neutral process and always has consequences” (Stobart, 2003, p. 140), the issues such as test impact, test ethicality, test fairness, test bias and issues relating to voice, which reflect power and authority tests possess, (Messick, 1996) manifest assessment as a highly complex phenomenon that has strong connections with multidimensional cultural, political, and institutional variables.

The influence of testing in terms of pedagogical and curricular implications has been extensively debated in educational arenas (Indika et al., 2015). Policymakers, bureaucrats and administrators exploit the enormous power associated with tests using them as strong tools to meet their agendas. For example, Shohamy (1993) investigates the impact of three national language tests introduced by the government of Israel at school level and highlights hidden political agendas of the government to control the educational system. As a consequence of the introduction of all three tests, classroom instruction became test-like. In other words, tests became the de-facto curriculum. Similarly, the socio-political and cultural dynamics of a particular testing situation influences the way testing and assessment practices are carried out (e.g., Cheng et al., 2004; Davidson, 2004; Yang, 2008). These highlight the issues of test fairness and voice, too. According to Shohamy (2001), since tests having a very powerful status in test takers’ lives can be detrimental affecting their self-belief, pride, dignity and opportunities for future success, it is fundamental to explore test takers’ voices based on their testing experiences in a particular testing situation so that those features of the use of tests can be identified which guide testers to fair testing.

In the EFL context of China, which is an examination-driven society, Xiao and Carless (2013) investigate how Chinese high school students view English language assessment systems and how it affects their lives. The student-participants voice their views about assessments in the form of drawing pictures and with an interview with the researcher. The findings suggest that a test can affect test takers emotionally in a positive or a negative way or both. Tests have the power to affect learners’ self-identities, too. In the EFL context of the Gulf, as far as I know, there are only a few studies conducted that are centred on some critical agenda consistent with the principles of critical language testing (e.g., Troudi et



al., 2009; Maedeh & Saeed, 2015). Based on the above literature review, we can conclude that there is a paucity of scholarship that problematizes language assessment and testing systems from the perspective of both teachers and students with a critical agenda intended for questioning current assessment practices.

## The Present Study

Keeping in mind the importance of the way we assess our students and its impact on their lives, the present study informed by critical language testing as theoretical background, not only describes the existing state of affairs, but also problematizes one of the set assumptions in ELT that “the more English language is tested, the better result-oriented it becomes”, which is in a way linked with Phillipson’s (1992) “the maximum exposure fallacy”. Hence, the study bridges a major vacuum in the literature. Based on the proposition of challenging the status quo and seeking a positive change through creating awareness about some of the assessment and testing related issues in one of the public sector universities in Saudi Arabia, this exploratory critical study reports findings centred on mainly interview data collected from 12 tertiary EFL instructors and 12 students through finding answers to the following research questions:

1. What do tertiary EFL instructors and students in KSA think of test fairness?
2. Do tertiary EFL instructors and students in KSA have voice in assessment-related policies?
3. How does the current English language assessment system in one of the public sector universities in KSA impact students and teachers?

## Methodology

### The Participants and the Context

A total of 12 tertiary level EFL practitioners and 12 students participated in the research. As shown in Table 1, the selection of the participants was based on purposive sampling. This technique, according to Patton, (2002 as cited in Perry, 2005), is used “to indicate that the sample is chosen deliberately to supply the most information possible regarding the research question” (p. 57).

TABLE 1

*Background Information about the Participants (n = 12)*

<i>Gender</i>	Male	7
	Female	5
<i>Nationality</i>	American (Patricia, Christopher)	2
	British (Stanley, Benjamin)	2
	Irish (Jessica)	1
	South African (Natalie, Monica)	2
	Algerian (Ahmad)	1
	Filipino (Conrad, Jason, Felix)	3
	Pakistani (Fatimah)	1
<i>Age</i>	35-40	2
	40-45	3
	45-50	4
	50-55	3
<i>Educational Background</i>	Doctoral Degree	5
	Master’s Degree	3
	Bachelor’s Degree (English Major)	2
	Bachelor’s Degree (Non English Major) with a certificate in TESL/TEFL/CELTA	2
<i>Teaching Experience</i>	0-5	2
	11-15	3
	16-20	3
	21-25	2
	26-30	2

*\*All names of the participants given above are pseudonyms.*

The context of the study was one of the public sector universities located in the Eastern province of Saudi Arabia.

## Instrument and Procedures

Believing that interviewing participants can be helpful in acquiring reliable data about the research questions elucidated through the understanding and standpoint of the participants, semi-structured interviews were conducted with 12 tertiary EFL practitioners and 12 Preparatory Year students. In order to generate ideas for the development of interview questions, two pilot interviews were conducted. The rationale behind this was to gain a wide range of ideas from the interview questions. The interview data was analysed using the grounded theory approach with the rationale to explore tertiary EFL practitioners' and students' understanding of the various assessment and testing related issues through comprehending the meanings of the participants' accounts and views in depth.

## Findings and Discussion

A detailed analysis of the interview data and the document review unravelled a number of significant findings about the current assessment and testing system in a public sector university in Saudi Arabia, which are discussed in detail in the following three sections. Every section ends with a brief discussion of the overall findings and their implications with reference to relevant literature and the research question.

### Issues Relating to Test Fairness

The findings of this section centre around two major themes: summative assessments-focused test dominated system and test validity issues. In response to the question about the frequency of tests and assessments, the majority of the teacher participants reported that they were simply too many and had counter-productive implications as was reflected in the words of one participant:

*I think ... to have this many exams is nightmarish for quality teaching and learning as you are not able to get your students up to the standards based on the results of the previous exam because the next exam is up there already...there is not enough time. (Stanley)*

In contrast, four of the teacher participants expressed their satisfaction with the current frequency of exams and assessments, but for three of them, the problem lies in the fact that the current assessment system is predominantly summative assessments-oriented, wherein the formative assessments are hugely ignored. These observations with regard to the summative assessments driven system resulting in the poor implementation of formative assessments are in line with the reflections reported in the literature (Leung & Rea-Dickens, 2007). One teacher participant advocated adding even more assessments to the current number in the form of short quizzes, assignments and unit tests. It is worth noting that almost all the teacher participants emphasized the significance of strengthening formative assessments by increasing their overall weightage and ensuring their better and effective implementation. For example, Ahmad, a teacher-participant, says:

*The summatives give students what you call the surface learning whereas the formatives provide them what we call deep or in-depth learning. (Ahmad)*

In contrast to the teacher participants, more than half of the student participants (7) report that the current number of assessments and tests at the university is appropriate and serves the pedagogical needs of the learners well whereas still for the remaining five, the frequency of tests and assessments is more

than enough. This mixed understanding predominantly in favour of frequent testing indicates the complexity of the nature of assessments, which cannot be seen as detached from the socio-cultural and institutional dynamics of the context (McNamara, 2001; Brown, 2005). When asked about the assessment and testing related challenges and difficulties that students have to face, most of the teacher and student participants identified issues relating to the gaps between classroom pedagogy and assessment, unfairness in the grading of the writing exams, and the poor administration of the formative assessments.

TABLE 2

*Gaps between Classroom Pedagogy and Assessment*

	Number of Teacher Participants who responded		Number of Student Participants who responded	
i. Assessments are aligned with classroom pedagogy	Yes	7	Yes	2
	No	5	No	10
ii. Areas of Gaps	Number of Teacher Participants who responded		Number of Student Participants who responded	
Level of difficulty	1		10	
Length of Reading	1		10	
Comprehension/Listening audios	2		8	
Time factor				
Quality of the speakers in Listening exams	-		8	
Format of Questions	-		6	
Physical environment	1		10	

As indicated in Table 2, almost all the student-participants reported having issues regarding what they get exposed to in the class and what they actually get tested on raising questions about content, construct and ethical aspects of exams' validity. This is reflected in these words:

*... the exam questions and procedures are sometimes very different and somewhat more complicated than what we practise in the classroom. (a student participant, Maha)*

In addition to the exams' validity issues, we notice concerns related to reliability and scorer-reliability as well. According to Hughes (2003), there are two types of scorer reliability: inter- scorer reliability and intra-scorer reliability where the former refers to the consistency of the measurement of student achievement if the same test is marked by different people using the same scoring criteria and the latter refers to the consistency of student scores if the same test is marked by one person on different occasions. The analysis of the student writing samples marked by different teachers and interview data reveals unfairness in the grading of the students' writing exams.

The majority of the teacher-participants reported that the marking of the students' writings was not consistent and fair across the markers on account of multiple reasons. The main factors reported as attributable to inconsistent and unfair marking were time constraints, a dearth of appropriate teacher training, and a lack of consistent, uniform and clear marking criteria and rubrics.

Like teacher-participants, the majority of the student-participants also report unfairness in the marking of the writing exams. For instance, five report that it is unfair, 3 report that it is fair whereas 4 are not sure if their marks in the writing exams are fair or not fair. This is expressed by one student participant:

*... some exam results especially those of writing exams have been really surprising...when I expected high marks, I got low whereas when I wasn't that satisfied with my performance in the exam, I got better marks. (Saud)*

This indicates their somewhat lack of trust in the credibility of their scores. As regards formative assessments, i.e., writing portfolio and weekly speaking assessment, the majority of the teacher participants report that these are not administered well resulting in multifaceted pedagogical

complications and unfairness. The main problems identified by the teacher-participants were too many writing assignments invoking too much marking (7); repetitive/archaic nature of the writing tasks (4); inconsistent, unfair marking on account of the lack of clear, uniform marking criteria (10); feedback and its follow-up issues (10); inadequate time and weight given to the speaking assessments (8); and content-based testing of the students' speaking abilities (4). These details indicate undermined and neglected formative assessments in a summative-assessments-dominated setting.

Taken together, the findings of the first section reveal that tertiary EFL practitioners suffer from predicaments as they work in an examination-driven setting wherein summative assessments dominate. Also, whatever formative assessments are administered, they are more or less based on the principles of psychometric testing especially in terms of marking and evaluation. Both these kinds of assessments are based on different paradigms, i.e., positivism/naturalism which informs a traditional psychometric testing culture and interpretivism/constructivism which is the basis of an alternative/formative assessment culture (Inbar-Lourie, 2008).

However, the teacher-participants' emphasis on the strong need to promote and strengthen formative assessments at the university is consistent with the importance of assessment for learning purposes (Brown, 2005; Davison & Leung, 2009).

### **Concerns Pertaining to Tertiary EFL Teachers' and Students' Voice in Assessment-related Policies and Decisions**

This section revolves around three major themes, i.e., a top-down management system, a hierarchical power struggle at different levels, and a very restricted teachers' and students' voice in assessment and testing related policies.

In response to the question about the management of the test committee, the majority of the teacher-participants reported that the committee was very controlling and members worked in a very closed-circuit system. This is manifested in these words:

*I think the mandate of a testing committee in any educational institution is to evaluate/review the exams and take important decisions, which should be communicated department wide rather than being kept in a closed circuit... (Benjamin)*

This power and control, according to most of the teacher-participants was evident inside and outside the class both at the test committee and the departmental level. More than half of the teacher-participants reported that there were very strong power issues in the class-the teacher being the authority and the students being the sub-ordinates. As regards power issues at the test committee as well as the departmental level, Conrad, one of the participants expresses his feelings in these words:

*... in a multicultural academic setting like the one we have here, I think there are factions, there are cliques and if you don't happen to be one of the acknowledged cliques - the ones closed to the chairperson, you feel very powerless. (Conrad)*

The majority of the teacher and student-participants reported having very limited or no voice in assessment-related policies and decisions. The main factors reported as attributable to stifling teachers' voice are: top-down management (3); teacher-personality factor (2); too many barriers in affecting change (3); and lack of trust in the effectiveness of voicing out (4). As regards student voice, teacher-factor still remains the most significant factor in encouraging or stifling it, which is evident in a student participant's words:

*... it really depends on the teacher who teaches you if you have voice or you don't. (Ibrahim)*

Taken together, the results of this section highlight teachers' and students' inadequate and ineffective voice in assessment-related policies and decisions as the power and authority is centralized in a few factions only. However, according to Bourdieu's (1991) notion of symbolic power, it is the powerless and voiceless that grant power and authority to the powerful to let them dominate as the powerless want to be dominated. In order to ensure the healthy progress at the personal and societal level, however, the voice and authority of the powerful must be dialogically challenged by voice which is internally convincing to the individual (Bakhtin, 1981).

### Effects of Assessment and the Testing System on Teachers and Students

This section pivots around two major themes, i.e., negative test washback (teaching to the test) and detrimental test effects on students' and teachers' lives. In response to the question about how the current assessment system affects teaching, the majority of the teacher-participants report that the system is so restrictive that it forces teachers to teach to the tests. For instance, a teacher-participant says:

*I think we are teaching to the test because everything that we do in the class is what is going to come in the test ... the system is too restrictive; it doesn't test their ability to use the language freely.*  
(Jessica)

Another participant, Felix, says that "...here in our context, students' results are reflective of your success as a teacher-you are considered a good or a bad teacher by the students based on your class results, so under these pressures, the teacher gets tempted to mainly teach to the test".

The main factors reported as leading teachers to teach to the tests are: excessive testing (6); high stakes associated with the tests' results (3); teacher training, educational background and teaching experience factor (2); and a somewhat mechanical and formulaic nature of the course (1).

This general trend of tests serving as de-facto curriculum, wherein instruction becomes test-like making learning limited, formulaic and superficial is aligned with the literature (Shohamy, 1993). In addition to the impact of an examination-driven system on classroom pedagogy, the data analysis also reveals that the assessment system at the university has both positive and negative effects on teachers' and students' lives.

TABLE 3

#### *Impact of Tests and Assessments on Students*

	Number of Student-participants who responded
i. Positive Effects	
Helpful in learning time-management skills	3
Effective in improving ability to work under pressure	2
Good for learning the use of English language skills	5
Good for getting familiar with the dynamics of the testing system	2
ii. Negative Effects	
Physical tiredness	2
Emotional and psychological complications	4
Use of unfair means to pass	
Impact on learning and achievement on account of being under constant pressure	2
Disbelief/ lack of trust in the credibility of test results	3
	1

As revealed in Table 5, almost half of the student-participants report that the current testing system provides them good opportunity to learn how to use their English language skills. As regards negative effects of the testing system, most of the students report emotional, psychological, and pedagogical complications they suffer in the hands of the powerful social institution of tests. This gets reflected in one student-participant's words:

*... the system here puts a heavy pressure on the students...you know, before every exam I feel so nervous and think a lot about it. During the exam time, I feel so afraid, anxious and nervous, however, mostly I feel relieved after I finish, but this good feeling comes after a difficult journey of nervousness and anxiety. (Maryam)*

These detrimental effects associated with the use of tests indicate their power and significance in the test-takers' lives as they determine "winners and losers, successes and failures, rejections and acceptances" (Shohamy, 2001, p. 15).

In summary, the findings of the study based on teachers' and students' judgments and the review of some exam-related documents suggest that the current assessment system under investigation, which is predominantly summative-assessments driven, has some issues relating to content, construct, systemic, ethical and consequential aspects of exams' validity. Here, the content and construct validity refer to the content and principles on which a test is based whereas systemic, ethical and consequential aspects relate to the various dimensions of test impact. Also, the findings reveal that the whole assessment system is controlled and centrally managed by the university organizing body whereas at the departmental level, it is the testing committee that manages and controls the assessment policies and develops assessment tools. Both teachers and students have very limited voice in this regard. The results of the study are somewhat consistent with the findings of Farah (2007) and Troudi et al., (2009).

Based on the issues identified in the current assessment and testing system and suggestions put forward by the teacher and student informants, it is essential to consider the following to change the status quo for fair, valid and reliable assessment practices:

There is an urgent need to \_\_\_\_\_

1. reduce the number of summative assessments and enhance formative assessments. A move from a customary psychometric testing culture which supports assessment of learning (AOL) to an assessment culture centred on assessment for learning (AFL) embedded in the principles of constructivism and interpretivism need to be encouraged. Even the summative assessments must be carried out formatively.
2. introduce and increase the creative ways of testing, i.e., the inclusion of authentic/real-life tasks in the assessments.
3. reconsider the composition and the role of the testing committee by sharing power and giving appropriate consideration to the role of the stake holders such as teachers and students in the assessment-related decision-making process.
4. promote research at the departmental level in order to have a better understanding and insight into the various pedagogical and assessment-related issues at the institutional level.
5. encourage the development of teacher language assessment literacy. Also, teachers need to be trained in the effective implementation of a variety of post-modern pedagogies such as critical pedagogy and participatory approaches to learning in order to promote collaborative, democratic system-based learning.
6. empower learners by giving them voice through their involvement in various assessment-related policies and decisions.
7. promote critical language testing activities at the department level by encouraging democratic methodologies to assessment, wherein power is shared and knowledge is constructed collaboratively.

## **Conclusion**

The present qualitative study based on mainly interview data collected from EFL practitioners and students has revealed some of the concerns relating to test fairness, test impact and voice prevalent in the

assessment and testing system in the context of one of the public sector universities in Saudi Arabia. Overall, the findings suggest that the system has a number of issues regarding the consequential and ethical aspects of exams' validity, which have detrimental effects on students' and teachers' lives. Also, the analysis of data reveals that both the teachers' and students' voice is very limited in assessment-related decisions and policies owing to some of the asymmetrical power relations existing both in and outside the classroom. Despite being almost voiceless in decision making for assessments and tests, the participants do voice out the strong need to enhance and promote a formative assessment culture at the university because, to them, an over-emphasis on summative assessments and excessive testing is counter-productive. The findings of the study that EFL instructors have a restrictive voice in assessment-related decisions and that how high-stakes tests impact students' and teachers' lives are partially consistent with the results of existing empirical studies (e.g., Farah, 2007; Troudi et al., 2009).

Overall, the study despite being limited in its scope, qualitative in nature, and having a small sample size has important implications in the field of assessment and testing. Further research in this area that involves multiple data collection instruments is recommended in different EFL/ESL contexts so that the results of this study can be verified.

### The Author

*Muhammad Waseem* has a doctoral degree (EdD TESOL) from the University of Exeter, United Kingdom. He has a wide range of teaching experience at tertiary level in both EFL & ESL contexts. His present work place is English Language & Preparatory Year Institute, Jubail University College, Royal Commission for Jubail, Saudi Arabia. His research interests include second language acquisition, teacher education, language pedagogy & language testing and assessment.

Royal Commission Colleges & Institutes  
Royal Commission for Jubail  
Jubail Industrial City, Saudi Arabia  
P.O. Box 10099  
Email: mwaseemlatif21@gmail.com

### References

- Bakhtin, M. (1981). Discourse in the novel. In M. Holquist (Ed.), *The dialogic imagination: Four essays by M. M. Bakhtin* (translated by C. Emerson & M. Holquist, pp. 259-422). Austin: University of Texas Press
- Bourdieu, P. (1991). *Language and symbolic power*. Cambridge, MA: Harvard University Press
- Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education, 1*, 81-89.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom practices: Purposes, methods, and procedures. *Language Testing, 21*(3), 360-389.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australia and Hong Kong. *Language Testing, 21*(3), 305-334.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393-415
- Farah, B. (2007). Students' voices on their experiences with a high stakes language test. In S. Midraj, A. Jendli, & A. Sellami (Eds.), *Research in ELT contexts* (pp. 252-270). Dubai: TESOL Arabia Publications.
- Foucault, M. (1979). *Discipline and punishment*. New York: Vintage.
- Hughes, A. (2003). *Testing for language teachers*. New York: Cambridge University Press.



- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Indika, L., Brendan, B., Tony, W. & Xuhong, G. (2015). Assessment policies, curricular directives, and teacher agency: Quandaries of EFL teachers in Inner Mongolia. *Innovation in Language Learning & Teaching*, 9(3), 251-264.
- Leung, C. & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Language Assessment Quarterly*, 4(1), 6-36.
- Maedeh, G. & Saeed, K. (2015). Voices from the voiceless: Iranian EFL students' attitudes. *Innovation in Language Learning & Teaching*, 9(2), 102-114.
- McNamara, T. K. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18, 333-350.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-257.
- Noam, G. (1996). (Moderator). Assessment at a crossroads: Conversation. *Harvard Educational Review*, 66, 631-657.
- Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage Publications.
- Perry F. L., (2005). *Research in applied linguistics: Becoming a discerning consumer*. London: Lawrence Erlbaum Associates.
- Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.
- Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. *NFLC Occasional Papers*, ERIC, The National Foreign Language Centre Washington, DC.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. United Kingdom: Pearson Education.
- Stobart, G. (2003). The impact of assessment: Intended and unintended consequences. *Assessment in Education*, 16, 139-140.
- Troudi, S., Coombe, C., & Al-Hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555. Doi: 10.1002/j.1545-7249.2009.tb00252.x
- Xiao, Y., & Carless, R. D. (2013). Illustrating students' perceptions of English language assessment: Voices from China. *RELC Journal*, 44(3), 319-340. SAGE DOI: 10.1177/0033688213500595
- Yang, T. L. (2008). EFL teachers' multiple classroom assessment practices of young language learners. *Studies in the Humanities and Social Sciences*, 9, 139-176.

(Received January 06, 2021; Revised February 20, 2021; Accepted March 10, 2021)