English as a Foreign Language Teachers' Assessment Competence at Taibah University

Abdulrahman Olwi
Taibah University, Madinah, Saudi Arabia

Abstract—This research study examines the competence of EFL teachers about language assessment types and strategies. A random sample of 49 teachers at the preparatory year program at Taibah University has taken a questionnaire that tests language assessment competences (i.e., diagnostic, progress, placement, proficiency, and achievement) and its correlation with their gender, native language, qualification, years of experience, and non-academic training. Overall, female teachers had better competence of achievement assessment over male teachers, but not in other types of assessment. Both English native speaking teachers and non-English native speaking teachers have less competence in language proficiency assessment competence. Also, it is seen that higher degrees help better in the achievement assessment competence, but not in general. Regardless of the number of years of experience, less competence occurred with proficiency assessment. Finally, in terms of non-academic training, there was a significant difference in the competence of progress assessment and proficiency assessment.

Index Terms—EFL, teachers, language competence, assessment knowledge, teacher knowledge

I. INTRODUCTION

English language teaching has become an important profession which has an international community that develops standards, publishes journals, establishes organizations, holds events, creates teams, shares discourses, sets principles, … etc. As part of the world, using English in Saudi Arabia has also been growing, and thus the necessity to have more English language teachers in Saudi Arabia has increased (Al-Osaimi, 2013). More research studies investigating teachers' competence are required to start correlating and filling the gap exists between knowledge and performance. Educational decision makers need to have empirical evidence of teachers' strengths and weaknesses based on these standards that represent teacher competence. An absence of this empirical evidence hinders efficient educational decisions.

Teacher knowledge is the actual information, central ideas, arranged principles that scholars know as making up the field (Sanders & Morris, 2000). The results of this research study show teachers themselves the merits and demerits of English language teacher knowledge in Saudi Arabia. Also, this research study is beneficial for English language teachers who work in the field and directly interact with English language students as well as recruitment committees that set the conditions that all applicants should meet such as ~~the~~ higher qualification, native language, non-academic training, …etc.

As Kunnan (2004) stated, the study of language testing did not begin until the 1930s. Even at that time, it did not get enough attention until 1961 when Lado wrote the book of Language Testing and also Carroll published the article "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students". Actually, when an accredited association places a test that is administered to thousands of test takers, mentioning language testing until that time in the literature was still limited (Spolsky, 2000).

II. PURPOSE OF THE STUDY

Because English language teaching has become essential to academic life in Saudi Arabia, there is a large number of research studies conducted to investigate different areas of English language teaching pedagogy to Arabic native speaking students. However, according to Al-Osaimi (2013), student achievement and skills development have received more attention than teacher knowledge and performance in these research studies.

This study aimed to investigate ESL teachers' language assessment and testing competence. Not only this study sought to check the impact of the years of teaching experience and work environment, but also it sought to see how having varied preparation and backgrounds (i.e., academic certificate programs requiring university coursework, professional non-credit bearing certificate programs involving training sessions, and academic degree studies such as master's in TESOL) play an effective role in ESL teachers' knowledge of language assessment and testing.

Given the context of Intensive English Programs (IEP) at language institutions in the United States that have the mission of preparing international English learners to gain entrance to academic programs, this study saw how different ESL teachers from differing levels of preparation differ in knowing language assessment and testing. The main goal was to investigate whether there are differences among ESL teachers in language assessment and testing competence

that can be attributed to academic degrees (related graduate degrees, additional training, or professional development), length of teaching experience, or school policy and environment.

III. RESEARCH QUESTION

Based on the purpose of the study, the following research question was addressed: Are there any statistical significant differences among EFL teachers' assessment competence at Taibah University due to gender, native language, qualification, experience, and non-academic training?

IV. LITERATURE REVIEW

During the recent decade, there have been frequent national attempts to develop educational standards in Saudi Arabia. Huge efforts to promote high teaching standards have been dedicated to increasing teaching quality. Political decisions have been made, strategic plans have been set, budget approvals have been signed, national commissions have been established, immense projects have been launched, private educational companies have been founded, and various work teams have been built at different levels in order to grant better education for next generations of Saudis who face an unknown future.

For this research study that investigates the English language teachers in Saudi Arabia, it is acceptable to state that highly qualified English language teachers should have received a Bachelor's degree with a major in English language, taken some pedagogical courses in English education, and met English language teacher standards in Saudi Arabia. People around the world established standardized competency tests to measure teacher content knowledge, starting from the 1960s and 1970s. This has become common to judge the competence of teacher candidates by looking at their university transcripts, which is not accurate but somewhat supported by research studies (Goldhaber & Brewer, 2000). Thus, it is understood why teachers sometimes get hired only based on holding a Bachelor's degree related to the teaching position.

Since this research study dealt with the content knowledge of English language teachers, it is reasonable to review language assessment and testing as an approach. Working on language assessment and testing has been seen as an area of an academic major in its own right. This has increased a deepening awareness of the theories and practices of language testing. For example, an annual meeting known as a Language Testing Forum which has been held in the United Kingdom since the 1980s, and participants from different countries come to attend this huge event. Also, the International Language Testing Association (ILTA) and the Association of Language Testers in Europe (ALTE) started developing identities related to systemized codes of assessment ethics and, also, arranging formal language testing events for discussions and debates on different theories and practices. Besides that, many academic journals and scientific periodicals related to language assessment have been widely published. These facts make language testing an area of study that requires professionalism.

Generally speaking, formal testing started around 1,500 years ago in China. In the 16th and 17th centuries when universities were founded, it started in Europe. In the 18th century and the beginning of the 19th century, testing became more popular in many countries, especially in France and Germany. By the end of the 19th century, the United Kingdom and the United States introduced the idea of mass assessment when starting mass education systems.

It is important to give a brief of the development of language testing. At the beginning of the 20th century, around 1913, in the United Kingdom, the University of Cambridge established the Cambridge Proficiency Examination (CPE) to test the language proficiency of foreigners who applied for British schools. Only 12 people took it in the first year. CPE was based on a coherent philosophy of language learning that was developed in 1899 by Henry Sweet who focused on language use over language knowledge. In contrast, in the United States, in 1908, Thorndike developed the first standardized test that focused on students' written performance. He used 200 teachers' samples to come up with one handwriting scale. In 1912, Hillegas developed a scale for written composition, and, in 1914, Courtis combined some scales and added some measurements to prepare portfolios of test takers. In 1915, Kelly developed a multi-choice question (MCQ) format in the Kansas Test of Silent Reading. In 1920, Handschin talked about foreign language tests and his new approach at that time.

The main difference between the United Kingdom and United States' testing techniques was that the United States' technique, with some exceptions like Thorndike's test and the Modern Languages Association of Maryland's declaration, looked at standardization and psychometric excellence before content and validity. In other words, it is the science of measurement in the United States and the art of measurement in the United Kingdom (O'Sullivan, 2011, p. 2). From that time until today, the business of testing has been going on.

By the 1940s, in the United States, high-stakes testing had become almost completely a standardized multiple-choice format, with an exception of the military-inspired Foreign Services Institute Oral Proficiency Interview (OPI) that was the first speaking test. In the 1960s, a well-known first standardized foreign language test was developed: The Test of English as a Foreign Language (TOEFL). It remained as it was in the same format for 40 years until the Internet Based Test (TOEFL iBT) version was developed at the beginning of the current century. After that, the University of Cambridge Local Examinations Syndicate (UCLES), which is now called Cambridge ESOL, was used as a much-revised version of CPE. In 1979, the Test of English for International Communication (TOIEC) was introduced in

response to the Japanese Ministry of Trade's request for a test of English for business purposes. In the 1980s, the International English Language Testing System (IELTS) was established in the United Kingdom. At the beginning, it was known as the English Language Testing Service (ELTS).

It is also important to give a brief of the Modern language testing. Starting from the beginning of this current century, two trends have been noticed. First, test developers tend to support tests of specific purposes more than other general multi-purposes tests. Second, some specialists claim that the style of testing in the United States is to start making a balance between shape and content and no longer looking at the psychometrics over the content. In contrast, test developers in Europe have started giving more attention to the psychometric quality (Council of Europe, 2009).

After reviewing language assessment trends from the 1960s to 2000s, Malone (2013) claimed that earlier language assessment specialists focused on test theories, basic statistics, item development, and practicality. In more recent editions, the concentration has become more toward assessment methods and techniques inside the classroom. As McNamara (2008) indicated, recent attention has been given to accountability and ethics. Also, Davies (2008) found two main trends in language assessment and testing:

- Including all the necessary topics inside the same instructional materials without referring to any additional resources, and
- Considering principles besides knowledge and skills that concern about the appropriate use of assessment. Skills (i.e., items, statistics, …) and knowledge (e.g., validity, reliability, …) need to be supported by principles that give attention to ethics and professionalism.

All in all, Weigle (2002) suggested that it is important to determine the elements of real-life language use in order to define the construct of second language use for assessment purposes, which is really related to English language teacher knowledge assessment and testing.

Bachman's model of language ability, which was derived from Hymes' (1972) and Canale and Swain's (1980) work in Second Language Acquisition (SLA), has had a huge impact on today's modern language testing. After the revolution of language testing change in the 1980s and 1990s, some language testers today feel that this period of time is a stagnating time in language testing. However, others suggest an immense language testing change is coming in the near future. As some experts believe, this current period is dismantling the old hegemony of the two dominating language testing centers, the United States and the United Kingdom, and their theories and practices that have been on the surface for decades. This helps language testers take contextual parameters and cognitive processes into consideration from the test taker's perspectives.

Cronbach and Meehl in 1955 (as cited in Manning, 2013), writing about the systematic theories of validity, suggested giving evidence related to one of what has later become known as the three different types of validity: content, criterion, and construct. Also, Messick (1989) developed a new view of validity when arguing that validity does not only have a dichotomy of evidence, yes or no, but validity also has a position of degree regarding the extent that is possible to support test validity. It works in the United States where standardization and measurement have been prominent for many years and in the United Kingdom where the influence has been toward test content as a reflection of the test construct. In contrast, O'Sullivan (2011) believed that the validity provides a coherent outline of the key issues that are influencing language testing today. This approach brings all language testing specialists from all around the world together and helps them (academics, practitioners, sponsors, consumers, … etc.) compromise and work as a team.

Actually, test developers should reflect the effect on the performance of all decisions being made during the process of test development. Since it is difficult to find a significant and accurate understanding of how performance needs to be interpreted, test developers should not marginalize test consequence. Instead, they should recognize test consequence as a main factor of all of the test development decisions.

Nowadays and over the coming years, the area of investigation of language testing will have three main themes: validation, professionalism, and localization. When Messick (1989) extended the validity concept and included test consequence and the unitary nature of validity that makes validity assembled from many resources of validity evidence, a major attempt at that time occurred to switch the focus from validity --academic theorization-- to validation -- operational evidence collecting and reporting. However, this attempt failed because of the complexity and ambiguity of Messick's model of validity.

Localization means "the practices of taking into account those learner-focused factors that can impact on linguistic performance" (O'Sullivan, 2011, p. 6), which is an attempt to support the individualizing of assessment. In the past, tests that were developed locally seemed to have lower quality and less accuracy than international tests. This is no longer correct. Today, many local examinations are likely to be more accurate in a significant way if well prepared. Different populations of test takers have different linguistic and cultural backgrounds, so one standardized international test might not be fair for all of them at the same degree. This reinforces the argument that the conceptualization of test consequence validity should not be accepted in favor of the view that it is a basic aspect of arguments about validation elements. Another aspect of localization is the recognition of the context of test development. Standardized tests that international students from all around the world usually take, such as the TOEFL and IELTS, believe that their scores interpret the test taker's language ability in different language skills and their use of language in real life. This claim is seen as problematic because of assuming that the test takers are similar in their educational contexts and first languages,

which brings the challenge of accuracy or internal consistency. Last, it has been found that successful test developers are those who aim to develop clearer and more transparent theories of validation.

V. PARTICIPANTS

As shown in Table (1), a total of 49 EFL teachers at Taibah University participated in this research study. They are 18 female teachers and 31 male teachers, 10 English native speaking teachers and 39 non-English native speaking teachers. Also, 13 of the participants are teachers with a bachelor's degree in addition to 30 teachers with a master's degree and 6 teachers with a Ph.D. degree. In addition, 14 teachers had five years of experience or less and 35 teachers had more than five years of experience. Finally, 34 teachers reported non-academic training whereas 15 teachers did not have any non-academic training certificates.

TABLE (1)
PARTICIPANTS (N=49)

	Groups		
Gender	Male	Female	
	31	18	
Native language	Native	Non-native	
	10	39	
Qualification	Bachelor's	Master's	Ph.D.
	13	30	6
Years of experience	Five years or less	More than five years	
	14	35	
Non-academic training	With non-academic training	Without non-academic training	
	34	15	

VI. DATA COLLECTION

In this context, the assumption is referred to some understandings that are fundamental to the research design but are not certainly verified through the research methods and procedures. This research study is based on the following assumption that is made before starting the study: It is assumed that the assessment competence level of English language teachers can be efficiently measured by an electronic questionnaire. To collect the data for this research study, an electronic questionnaire was built and tested to examine the different types of language assessment competence. Then, it was sent to the director of the English language institute and the dean of the preparatory year at Taibah University. The questionnaire was sent to all EFL teachers' official emails and circulated in social media groups.

It was sure that the participants have read and understood the purpose of the questionnaire and they willingly consent to participate in this research study. Also, they directly agreed that their data, collected through this questionnaire, may be used for the purpose of the research findings. They were assured that all data gathered in this questionnaire are carried out securely and anonymously. Therefore, no names were required. Finally, if any participants declared that he/she is not currently teaching English at Taibah University, their answers were automatically eliminated from the data analysis of this research study. Regarding risks and benefits, there were no known risks or discomforts associated with this research study. On the other hand, it is believed that this research study provided participants with the strengths and weaknesses of their language assessment and testing knowledge.

VII. DATA ANALYSIS

All questionnaire responses were statistically analyzed using SPSS. At the beginning reliability and validity were checked for the whole questionnaire and for the items of each aspect of assessment competences. Then, basic descriptive analyses were run to explain the means and standard deviations between the types of language assessment competence (i.e., diagnostic, progress, placement, proficiency, and achievement) and their correlation with the teachers' gender, native language, qualification, years of experience, and non-academic training. Finally, the significance of each difference was examined in terms of gender, native language, qualification, years of experience, and non-academic training.

TABLE (2)
RELIABILITY OF THE QUESTIONNAIRE

	No. of Items	Cronbach's Alpha	Sig.
Language assessment questionnaire	25	0.820	Sig.

As shown in Table (2), the Cronbach's Alpha for the whole 25 item questionnaire is highly significant with a value of 0.820, which is appropriate for such nature of study.

TABLE (3)
RELIABILITY AND VALIDITY OF EACH COMPETENCE

	No. of Items	Reliability	Validity
Competence of diagnostic assessment	5	0.961	0.980
Competence of progress assessment	5	0.959	0.979
Competence of placement assessment	5	0.889	0.943
Competence of proficiency assessment	5	0.879	0.892
Competence of achievement assessment	5	0.879	0.853

As shown in Table (3), the reliability and validity are high for each assessment competences. Respectively, the reliability and reliability was the highest in the competence of diagnostic assessment (0.961 and 0.980), in the competence of progress assessment (0.959 and 0.979), in the competence of placement assessment (0.889 and 0.943), in the competence of proficiency assessment (0.879 and 0.892), and in the competence of achievement assessment (0.879 and 0.853); as indicated earlier, these numbers are the values of reliability and validity respectively for each language assessment type.

TABLE (4)
DESCRIPTIVE STATISTICS IN TERMS OF GENDER

	Female teachers (n=18)			Male teachers (n=31)		
	Mean	Std. deviation	Coefficients	Mean	Std. deviation	Coefficients
Competence of diagnostic assessment	4.722	0.826	17.5	4.612	1.115	24.2
Competence of progress assessment	4.888	0.323	6.6	4.354	1.226	28.2
Competence of placement assessment	4.833	0.383	7.9	4.580	0.807	17.6
Competence of proficiency assessment	4.000	1.028	25.7	3.838	1.392	36.3
Competence of achievement assessment	3.111	0.323	10.4	2.774	0.497	17.9

As shown in Table (4), 18 female teachers and 31 male teachers participated in this research study. In terms of female teachers, the place of the competence of progress assessment was the first (M=4.888, SD=0.323), and then the competence of placement assessment was the second (M=4.833, SD=0.383), and then the competence of achievement assessment was the third (M=3.111, SD=0.323), and then the competence of placement assessment was the fourth (M=4.722, SD=0.826), and then the competence of proficiency assessment was the fifth (M=4.000, SD=1.028).

In terms of male teachers, the place of the competence of placement assessment was the first (M=4.580, SD=0.807), and then the competence of achievement assessment was the second (M=2.774, SD=0.497), and then the competence of diagnostic assessment was the third (M=4.612, SD=1.115), and then the competence of progress assessment was the fourth (M=4.354, SD=1.226), and then the competence of proficiency assessment was the fifth (M=3.838, SD=1.392). Thus, it is noticed that both male and female teachers have less competence in language proficiency assessment competence.

TABLE (5)
DESCRIPTIVE STATISTICS IN TERMS OF ENGLISH AS A NATIVE LANGUAGE

	English native speaking teachers (n=10)			Non-English native speaking teachers (n=39)		
	Mean	Std. deviation	Coefficients	Mean	Std. deviation	Coefficients
Competence of diagnostic assessment	4.800	0.632	13.2	4.615	1.091	23.6
Competence of progress assessment	4.600	0.699	15.2	4.538	1.096	24.2
Competence of placement assessment	4.700	0.948	20.2	4.666	0.621	13.3
Competence of proficiency assessment	3.900	1.197	30.7	3.897	1.293	33.2
Competence of achievement assessment	3.100	0.316	10.2	2.846	0.488	17.1

As shown in Table (5), 10 English native speaking teachers and 39 non-English native speaking teachers participated in this research study. In terms of English speaking teachers, the place of the competence of achievement assessment was the first (M=3.100, SD=0.316), and then the competence of diagnostic assessment was the second (M=4.800, SD=0.632), and then the competence of progress assessment was the third (M=4.600, SD=0.699), and then the competence of placement assessment was the fourth (M=4.700, SD=0.948), and then the competence of proficiency assessment was the fifth (M=3.900, SD=1.197).

In terms of non-English speaking teachers, the place of the competence of placement assessment was the first (M=4.666, SD=0.621), and then the competence of achievement assessment was the second (M=2.846, SD=0.488), and then the competence of diagnostic assessment was the third (M=4.615, SD=1.091), and then the competence of progress

assessment was the fourth (M=4.538, SD=1.096), and then the competence of proficiency assessment was the fifth (M=3.897, SD=1.293). Again, as in terms of gender, it is noticed that both English native speaking teachers and non-English native speaking teachers have less competence in language proficiency assessment competence.

TABLE (6)
DESCRIPTIVE STATISTICS IN TERMS OF QUALIFICATION

	Teachers with a bachelor's degree (n=13)			Teachers with a master's degree (n=30)			Teachers with a Ph.D. degree (n=6)		
	Mean	Std. deviation	Coefficients	Mean	Std. deviation	Coefficients	Mean	Std. deviation	Coefficients
Competence of diagnostic assessment	4.692	0.751	16	4.766	0.971	20.3	4	1.549	38.7
Competence of progress assessment	4.692	0.63	13.4	4.566	1.04	22.7	4.166	1.602	38.4
Competence of placement assessment	4.769	0.438	9.1	4.666	0.802	17.1	4.5	0.547	12.1
Competence of proficiency assessment	3.769	1.48	39.2	3.866	1.224	31.6	4.333	1.032	23.8
Competence of achievement assessment	2.692	0.48	17.8	2.966	0.49	16.5	3	0	0

As shown in Table (6), 13 teachers with a bachelor's degree, 30 teachers with a master's degree, and 6 teachers with a Ph.D. degree participated in this research study. In terms of the teachers with a bachelor's degree, the place of the competence of placement assessment was the first (M=4.769, SD=0.438), and then the competence of diagnostic assessment was the second (M=4.692, SD=0.751), and then the competence of progress assessment was the third (M=4.692, SD=0.63), and then the competence of achievement assessment was the fourth (M=2.692, SD=0.48), and then the competence of proficiency assessment was the fifth (M=3.769, SD=1.48).

In terms of the teachers with a master's degree, the place of the competence of achievement assessment was the first (M=2.966, SD=0.49), and then the competence of placement assessment was the second (M=4.666, SD=0.802), and then the competence of diagnostic assessment was the third (M=4.766, SD=0.971), and then the competence of progress assessment was the fourth (M=4.566, SD=1.04), and then the competence of proficiency assessment was the fifth (M=3.866, SD=1.224).

In terms of the teachers with a Ph.D.'s degree, the place of the competence of achievement assessment was the first (M=3, SD=0), and then the competence of placement assessment was the second (M=4.5, SD=0.547), and then the competence of proficiency assessment was the third (M=4.333, SD=1.032), and then the competence of progress assessment was the fourth (M=4.166, SD=1.602), and then the competence of diagnostic assessment was the fifth (M=4, SD=1.549). It is noticed that there were only 6 participants with a Ph.D. degree. Overall, it is seen that higher degrees help better in the achievement assessment competence.

TABLE (7)
DESCRIPTIVE STATISTICS IN TERMS OF YEARS OF EXPERIENCE

	Teachers with 5 years of experience or less (n=14)			Teachers with more than 5 years of experience (n=35)		
	Mean	Std. deviation	Coefficients	Mean	Std. deviation	Coefficients
Competence of diagnostic assessment	4.642	0.928	20.0	4.657	1.055	22.7
Competence of progress assessment	4.928	0.267	5.4	4.400	1.168	26.5
Competence of placement assessment	4.857	0.363	7.5	4.600	0.774	16.8
Competence of proficiency assessment	4.000	1.240	31.0	3.857	1.286	33.3
Competence of achievement assessment	2.857	0.534	18.7	2.914	0.445	15.3

As shown in Table (7), 14 teachers with five years of experience or less and 35 with more than five years of experience participated in this research study. In terms of five years of experience or less, the place of the competence of progress assessment was the first (M=4.928, SD=0.267), and then the competence of placement assessment was the second (M=4.857, SD=0.363), and then the competence of achievement assessment was the third (M=2.857, SD=0.534), and then the competence of diagnostic assessment was the fourth (M=4.642, SD=0.928), and then the competence of proficiency assessment was the fifth (M=4.000, SD=1.240).

In terms of more than five years of experience, the place of the competence of achievement assessment was the first (M=2.914, SD=0.445), and then the competence of placement assessment was the second (M=4.600, SD=0.774), and then the competence of diagnostic assessment was the third (M=4.657, SD=1.055), and then the competence of progress assessment was the fourth (M=4.400, SD=1.168), and then the competence of proficiency assessment was the fifth

(M=3.857, SD=1.286). Thus, it is noticed that the less competence happens with proficiency assessment regardless of the number of years of experience.

TABLE (8)
DESCRIPTIVE STATISTICS IN TERMS OF NON-ACADEMIC TRAINING

	Teachers with non-academic training (n=34)			Teachers with no non-academic training (n=15)		
	Mean	Std. deviation	Coefficients	Mean	Std. deviation	Coefficients
Competence of diagnostic assessment	4.823	0.575	11.9	4.266	1.579	37.0
Competence of progress assessment	4.764	0.495	10.4	4.066	1.624	39.9
Competence of placement assessment	4.705	0.629	13.4	4.600	0.828	18.0
Competence of proficiency assessment	4.417	1.131	25.6	3.333	1.397	41.9
Competence of achievement assessment	2.941	0.422	14.3	2.800	0.560	20.0

As shown in Table (8), 34 teachers with non-academic training and 15 teachers without non-academic training participated in this research study. In terms of having non-academic training, the place of the competence of progress assessment was the first (M=4.764, SD=0.495), and then the competence of diagnostic assessment was the second (M=4.823, SD=0.575), and then the competence of placement assessment was the third (M=4.705, SD=0.629), and then the competence of achievement assessment was the fourth (M=2.941, SD=0.422), and then the competence of proficiency assessment was the fifth (M=4.417, SD=1.131).

In terms of not having non-academic training, the place of the competence of placement assessment was the first (M=4.600, SD=0.828), and then the competence of achievement assessment was the second (M=2.800, SD=0.560), and then the competence of diagnostic assessment was the third (M=4.266, SD=1.579), and then the competence of progress assessment was the fourth (M=4.066, SD=1.624), and then the competence of proficiency assessment was the fifth (M=3.333, SD=1.397). Thus, it is noticed that the less competence happens with proficiency assessment regardless of the non-academic training.

TABLE (9)
SIGNIFICANCE OF DIFFERENCES IN TERMS OF GENDER

	Female teachers (n=18)		Male teachers (n=31)		t-value
	Mean	Std. Deviation	Mean	Std. Deviation	
Competence of diagnostic assessment	4.722	0.826	4.612	1.115	0.361
Competence of progress assessment	4.888	0.323	4.354	1.226	1.805
Competence of placement assessment	4.833	0.383	4.580	0.807	1.245
Competence of proficiency assessment	4.000	1.028	3.838	1.392	0.427
Competence of achievement assessment	3.111	0.323	2.774	0.497	2.570

As shown in Table (9), in terms of gender, the t-value was significant for the competence of achievement assessment (especially for female teachers). Regarding other types of assessment, there was no significant difference between male and female teachers.

TABLE (10)
SIGNIFICANCE OF DIFFERENCES IN TERMS OF ENGLISH AS A NATIVE LANGUAGE

	English native speaking teachers (n=10)		Non-English native speaking teachers (n=39)		t-value
	Mean	Std. Deviation	Mean	Std. Deviation	
Competence of diagnostic assessment	4.800	0.632	4.615	1.091	0.511
Competence of progress assessment	4.600	0.699	4.538	1.096	0.168
Competence of placement assessment	4.700	0.948	4.666	0.621	0.135
Competence of proficiency assessment	3.900	1.197	3.897	1.293	0.006
Competence of achievement assessment	3.100	0.316	2.846	0.488	1.554

As shown in Table (10), in terms of speaking English as a native language, the t-value was not significant in all types of assessment. Thus, there was no significant difference between native and non-native speaking teachers of English regarding all types of assessment.

TABLE (11)
SIGNIFICANCE OF DIFFERENCES IN TERMS OF QUALIFICATION

		Sum of squares	df	Mean square	F	Sig.
Competence of diagnostic assessment	Between groups	2.966	2	1.483	1.479	Not sig.
	Within groups	46.136	46	1.003		
	Total	49.102	48			
Competence of progress assessment	Between groups	1.153	2	0.577	0.542	Not sig.
	Within groups	48.969	46	1.065		
	Total	50.122	48			
Competence of placement assessment	Between groups	0.301	2	0.151	0.308	Not sig.
	Within groups	22.474	46	0.489		
	Total	22.776	48			
Competence of proficiency assessment	Between groups	1.382	2	0.691	0.423	Not sig.
	Within groups	75.108	46	1.633		
	Total	76.49	48			
Competence of achievement assessment	Between groups	0.754	2	0.377	1.781	Not sig.
	Within groups	9.736	46	0.212		
	Total	10.49	48			

As shown in Table (11), in terms of whether holding a bachelor's, master's or Ph.D. degree, the t-values were not statistically significant in all types of assessment. Thus, pursuing higher academic degrees does not show better competence of assessment.

TABLE (12)
SIGNIFICANCE OF DIFFERENCES IN TERMS OF YEARS OF EXPERIENCE

	Teachers with 5 years of experience and less (n=14)		Teachers with more than 5 years of experience (n=35)		t-value
	Mean	Std. Deviation	Mean	Std. Deviation	
Competence of diagnostic assessment	4.642	0.928	4.657	1.055	0.044
Competence of progress assessment	4.928	0.267	4.400	1.168	1.666
Competence of placement assessment	4.857	0.363	4.600	0.774	1.185
Competence of proficiency assessment	4.000	1.240	3.857	1.286	0.355
Competence of achievement assessment	2.857	0.534	2.914	0.445	0.383

As shown in Table (12), in terms of the years of experience, the t-value was not significant in all types of assessment. Therefore, regarding all types of assessment, there was no statistical significant difference between those whose experience was more than five years and other teachers.

TABLE (13)
SIGNIFICANCE OF DIFFERENCES IN TERMS OF NON-ACADEMIC TRAINING

	Teachers with non-academic training (n=34)		Teachers with no non-academic training (n=15)		t-value
	Mean	Std. Deviation	Mean	Std. Deviation	
Competence of diagnostic assessment	4.823	0.575	4.266	1.579	1.818
Competence of progress assessment	4.764	0.495	4.066	1.624	2.300
Competence of placement assessment	4.705	0.629	4.600	0.828	0.492
Competence of proficiency assessment	4.417	1.131	3.333	1.397	2.157
Competence of achievement assessment	2.941	0.422	2.800	0.560	0.974

As shown in Table (13), in terms of non-academic training, the t-value was statistically significant in the competence of progress assessment and proficiency assessment. In other types of assessment, no different levels of competence were shown between those who have non-academic training and those who do not have non-academic training.

VIII. CONCLUSION

In conclusion, female teachers showed much progress assessment competence while male teachers showed much placement assessment competence. At the same time, both male and female teachers showed less proficiency assessment competence. Also, Female teachers showed better competence of achievement assessment over male teachers while no significant differences were found across the other types of assessment based on gender. In terms of English as a native language for teachers, it is also noticed that both English native speaking teachers and non-English native speaking teachers have less competence in language proficiency assessment. Native English speaking teachers had a better competence of achievement assessment whereas non-English speaking teachers had a better competence of placement assessment.

In terms of the teachers with a bachelor's degree, they showed better competence of placement assessment and less competence of proficiency assessment. With a master's degree, the competence of achievement assessment was the first, and the competence of proficiency assessment was the last. Last, the teachers with a Ph.D.'s degree showed better competence of achievement assessment and limited competence of diagnostic assessment. Overall, it is seen that higher

degrees help better in the achievement assessment competence, but pursuing higher academic degrees does not show better competence of assessment regarding the assessment competence in general.

On the other hand, teachers with five years of experience or less showed better competence of progress assessment. In comparison, teachers with more than five years of experience showed better competence of achievement assessment. Regardless of the number of years of experience, less competence happened with proficiency assessment. In comparison between all types of assessment, there was no statistical significant difference between the teachers who had various years of experience. Finally, teachers with non-academic training showed high competence of progress assessment while teachers without non-academic training showed high competence of placement assessment. In terms of non-academic training, there was a significant difference in the competence of progress assessment and proficiency assessment, but not in the other types of assessment. Overall, it is noticed that the less competence happened with proficiency assessment regardless of the non-academic training.

REFERENCES

[1]   Al-Osaimi, M. (2013). A standard-based evaluation of Al-Madinah Al-Munawarah in-service EFL teachers' knowledge, and its correlation to their performance (Unpublished master's thesis). Taibah University. Medina, Saudi Arabia.
[2]   Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1, 1-47. http://dx.doi.org/10.1093/applin/I.1.1.
[3]   Council of Europe. (2009). Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A manual. Strasbourg: Council of Europe. Available at www.coe.int/portfolio, section CEFR and Related Documents.
[4]   Davies, A. (2008) Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347
[5]   Goldhaber, D & Brewer, D. (2000). Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement. Educational Evaluation and Policy Analysis - EDUC EVAL POLICY ANAL. 22. 129-145. 10.3102/01623737022002129.
[6]   Hymes, D.H. (1972). "On Communicative Competence" In: J.B. Pride and J. Holmes (eds) Sociolinguistics. Selected Readings. Harmondsworth: Penguin, pp. 269-293.
[7]   Kunnan, A. J. (2004). Test fairness. In Milanovic, M. & Weir C., (Eds.), *European language testing in a global context: Proceedings of the ALTE Barcelona Conference* (pp. 27–48). Cambridge, UK: Cambridge University Press.
[8]   Malone, M.E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing,* 30(3), 329–344
[9]   Manning, A. (2013). EAP teacher assessment literacy (Doctoral dissertation, University of Leicester).
[10] McNamara, T. (2008). The social-political and power dimensions of tests. In E. Shohamy and N.H. Hornberger (Eds.), *Encyclopedia of Language and Education*, Vol. 7: Language testing and assessment (2nd ed., pp. 415–427). Dordrecht, The Netherlands: Springer.
[11] Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13– 103). New York, NY: American Council on Education and Macmillan.
[12] O'Sullivan, B. (2011). Language testing: theories and practices. Palgrave Macmillan. Basingstoke, UK: ISBN 9780230230637.
[13] Sanders, S. E. & Morris, H. (2000). Exposing student teachers' content knowledge: Empowerment or debilitation? *Educational Studies, 26(4)*, 397-408.
[14] Spolsky, B. (2009). Language management. Cambridge, New York: Cambridge University Press.
[15] Weigle, S. (2002). Assessing writing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff. (Eds.). *The Cambridge guide to second language assessment.* (pp. 236-246). New York: Cambridge University Press. Pp.10-18.

Abdulrahman Olwi is an assistant professor of language education and the dean of e-learning at Taibah University, Madinah, Saudi Arabia. He got his PhD from Florida State University and his MA from the University of San Francisco in TESOL. His research interest is in language assessment and language teacher preparation. He has taught methods of teaching English, English language practicum, research methods in foreign language studies, English language skills, language assessment, ESOL in content area, … etc. He can be reached at this email: amaolwi@gmail.com.