The Impact of Specialty, Sex, Qualification, and Experience on Teachers' Assessment Literacy at Saudi Higher Education

Sabria Salama Jawhar
Applied and Educational Linguistics
King Saud bin Abdulaziz University for Health Sciences
https://orcid.org/0000-0002-1799-8888

Ahmad M. Subahi[*]
Medical Physics
King Saud bin Abdulaziz University for Health Sciences
https://orcid.org/0000-0002-2032-2993

Abstract. The importance of assessment literacy (AL) among teachers at all educational levels has been of interest for a considerable period. Literature has shown the increasing research surrounding the importance of AL knowledge and its impact on students' performance. This paper examines the level of assessment literacy among instructors on a Saudi higher education university preparatory program. It uses a/the Classroom Assessment Literacy Inventory (CALI), developed by Mertler and Campbell (2005), and a demographic questionnaire to investigate the AL level. It also looks at the impact of factors such as specialty (i.e., subject department), sex, academic qualification, and years of teaching experience on the AL level. The subjects of the study included 54 English and Basic Sciences teachers on the preparatory program. The study found that the level of AL among the participants was inadequate (mean=17). However, there was no significant difference found among the participants when it came to sex, qualifications, or years of experience. The subject department, on the other hand, was found to be a factor that impacted the level of AL. The study concludes with some recommendations related to increasing assessment literacy among in-service teachers and, more importantly, it highlights the importance of contextual factors when evaluating assessment literacy.

Keywords: assessment literacy; higher education, gender; teaching experience

---

[*]Both authors have contributed equally in this work.

1. Introduction

Assessment is a critical aspect of the education process, and researchers have placed it at the heart of the teaching and learning process. It is one of the most important tools through which teachers can evaluate their students' yearlong work. Moreover, it is the only scheme through which teachers are able to see the impact and efficiency of their teaching approaches. Consequently, classroom assessment literacy (CAL) is essential for teachers, not only so they can prepare the students to deal with the end of the semester exams but also so they can carry out their classroom-based assessment and make informed decisions regarding their students' achievements. Researchers (e.g., Earl & Katz, 2006) associate assessment with the promotion of learning and improvement of the students' performance during the year.

According to the reviewed literature, the problem is that many teachers get involved in academic activities related to students' assessment and, consequently, their achievement, without possessing the necessary knowledge for it (DeLuca, 2012; Lam, 2015). Given the rising prevalence of CAL, there is a pressing need for educational institutions, which are built on a system that puts formative assessment at the heart of the teaching and learning process, to emphasize assessment literacy (AL) to their staff. Also, they need to understand the consequences of a deficiency in such essential knowledge. Malone (2013) emphasized the mutual relationship between teaching and assessment, adding that they inform each other, and that assessment literacy affects the quality of education. Valencia (2002) argued that though teachers use assessment to gather evidence from their classrooms regarding their students' achievements and their teaching practices, they are still not able to use that information to make an instructional decision. Some teachers cannot understand the implications of the results they obtain through their students' assessment (Rogier, 2014). A closer look at the literature review shows that, despite the importance given to assessment literacy among teachers, it is still under-researched in many educational contexts. Saudi higher education is one of those contexts where there is a scarcity of conducted research papers investigating assessment literacy among teachers.

This study is an addition to the body of work investigating the level of classroom assessment literacy at the higher education level. However, it focuses on the impact of factors such as subject department (referred to as specialty), sex, teachers' qualification level, and years of teaching experience (referred to as experience) with a specific focus on assessment literacy. It also highlights the importance of context when evaluating assessment literacy among teachers. In other words, the paper provides proof that teachers' theoretical knowledge about assessment is not necessarily reflected in their actual practice. The context where assessment is carried out, we argue, is more important than general theoretical knowledge. This is not to undermine theoretical knowledge. On the contrary, it is a call for an expansion of the instrument(s) through which AL is evaluated to include context-specific elements (Inbar-Lourie & Levi, 2020). It is crucial to mention here that the terms 'classroom assessment literacy' and 'assessment literacy' are used interchangeably throughout this paper.

2. The importance of the study

There has been some sustainable research activity in teachers' assessment literacy and its impact on the teaching and learning process. As a result, several instruments have been developed to assess and understand the nature of this knowledge and how it can be enlisted under the construct of AL. The efforts of most research groups have been concentrated on conceptualizing AL and determining what aspects might be considered as crucial knowledge or skills for the teacher to be considered assessment literate. Others focused on the factors that affect the level of AL among pre-and in-service teachers.

To date, the number of studies that have looked at this concept is still limited and context-specific. Saudi Arabia, going through massive reform that has touched upon a variety of different aspects of life, has placed education at the heart of the country's transformation plan, with billions of dollars of funding being invested in the education sector. In addition, the country is counting on education to help reduce its dependence on oil as the primary source of income.

With this emphasis on education comes the recognition of the importance of assessment and the teachers' ability to design tests while, at the same time, understand the implications of educational outcomes. As far as the researchers are concerned, little if any research has been done to look at teachers' AL in the Saudi higher education context. This study measures the overall assessment literacy level among university teachers in a Saudi context, considering the impact of factors such as specialty i.e. the subject department, sex, qualification, and years of experience on the level of AL.

3. The research questions

The paper aims to measure the level of assessment literacy among teachers in a Saudi higher education preparatory program. It also tries to determine if factors such as subject department, sex, level and type of qualification, and years of teaching experience impact the teachers' general assessment knowledge. To this aim, we propose the following research questions:
1. What is the assessment literacy level of teachers in Saudi higher education as measured by the CALI?
2. How does the assessment literacy of Basic Science teachers compare to that of English Language teachers?
3. How does the assessment literacy of female teachers compare to that of male teachers?
4. What is the relationship between the teachers' academic qualification and their overall assessment literacy?
5. What is the relationship between the teachers' experience and their overall assessment literacy?

4. Methodology

4.1. Research context
The study was carried out at the Preparatory Program at King Saudi bin Abdulaziz University for Health Sciences (KSAU-HS), Jeddah campus. During this two-year program, sometimes referred to as the pre-professional program, students follow intensive English language courses to improve their English language proficiency to be able to study their future professional courses where English is the medium of instruction. During those two years, the students are also introduced to the basic concepts of general and medical science that are needed for higher-level education in the health sciences. Those concepts are presented in the form of eleven Basic Science courses, such as biology, chemistry, and physics. The program consists of two paths, nursing and unified. The nursing students choose their specialty before joining the university, while the unified students go through a selection process based on their General Point Average (GPA).

Because of the limited number of allocated seats at the School of Medicine (around 150 depending on each campus' capacity), the use of exam scores in this context is crucial in determining the students' future. The students who are not selected for medicine will be directed to the school of Applied Medical Sciences, where they also go through an additional phase of selection based on their GPAs. Those factors make the program highly competitive and place more importance on the accuracy of assessment, and consequently, the teachers' assessment literacy. The study targeted all teachers at the preparatory program, Jeddah campus, as part of the faculty enhancement initiative to assess the need for training to ensure quality teaching and learning. All teachers, i.e., Basic Science and English, were invited to participate in the study. However, it was not mandatory, so only those who were willing joined.

4.2. Participants
The participants of this study consist of 54 English Language and Basic Science teachers in a preparatory program. As the demographic information shows (table 1), the participants varied in their educational background. Most of the participants are English language teachers (n=41) while Basic Science teachers represent less than half of the sample (n=13). The female participants, on the other hand, made 60% of the sample(n=33). When it comes to educational level, most of the participants are master's degree holders (n=45). The instructors' experience varied with the majority having work experience of between 11 and 15 years (See table 1). The paper uses the convenience sample method. More than 70 questionnaires were distributed, but only 54 were completed and returned.

4.3. Procedures
The permission and IRP for this study were obtained. The participants were granted anonymity, and they were told that the result of the questionnaires would not influence their annual appraisal form in any way. They were also informed that the study's primary goal was to help improve the university's assessment practices and, consequently, enhance teaching and learning quality.

The teachers who agreed to participate were provided with a hard copy of the questionnaire containing 35 multiple-choices questions. The participants' responses were calculated for correctness, with a value of one score given to the correct answer and a value of zero given to the wrong response. The total number of correct answers for the five questions represented the score for each standard. The scores of CAL were then divided into three levels, namely: Inadequate or needs improvement (lower than 60%), Fair (60-79%) and High (80% and higher). The survey also included the participants' demographical information (See 4.4). There was no time limit specified for answering the questions, and the teachers were permitted to take the questionnaires home and return them whenever they had completed them.

4.4. Instrumentation

In this paper, we used the Classroom Assessment Literacy Inventory (CALI) approach. The instrument was developed by Mertler and Campbell (2005). It consisted of five scenarios followed by seven questions each, i.e., a total of 35 questions. Each of the seven questions within a single scenario was aligned to one of the following standards:

- Standard (1): related to the teachers' skills in choosing assessment methods appropriate for instructional decisions.
- Standard (2): related to the teachers' ability to develop assessment methods appropriate for instructional decisions.
- Standard (3): about the teachers' ability to administer, score, and interpret the results of both assessment methods.
- Standard (4): about using the assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
- Standard(5): tests the teachers' ability to develop valid students' grading procedures that use student assessments.
- Standard(6): addresses the teachers' skills in communicating assessment results to students, parents, and other educators.
- Standard(7): adresses the teachers' skills in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Each of the 35 items consisted of four options, of which only one was correct. When statistically calculated, the right answer received a score of (1), while the wrong answer received a score of (0). The aim was to assess the general level of the teachers' knowledge regarding some competencies related to assessment. Consequently, a higher total score equated to a higher level of assessment literacy. It is very important to mention that the instrument developers initally aligned the seven standards with the American "Standards for Teacher Competence in Educational Assessment of Students" (The American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association, (1990)). The second part of the questionnaire aimed to collect demographic information related to the participants' subject department, sex, the educational level, and length of years of teaching experience.

5. Literature review

The term assessment literacy (AL) is used as an umbrella term under which teachers' different knowledge regarding the what, how, and why of classroom assessment can be tested or measured (Stiggins, 1999). Researchers agree on the importance of AL as part of the academic and professional skills that teachers need across the different educational levels and disciplines (Xu & Brown, 2016). However, some of them went as far as considering AL an essential preventive measure against possible problems that might result from the lack of such knowledge, especially if the impact of assessment on the teaching and learning process is taken into consideration (Stiggins, 1995; Brookhart, 2011). Others associated AL with the students' achievement and their ability to learn (Rogier, 2014).

The literature review shows that despite this agreement among researchers and educators regarding AL's importance, there is still a great deal of debate surrounding the generic skills teachers need to possess to be considered assessment literate. For instance, Paterno (2001) argued that the basic knowledge required for assessment literacy included knowledge of assessment terminology and ways of developing and using assessment methodologies and techniques. Fulcher (2012), on the other hand, enlisted the following among the needed skills for assessment literacy: knowledge about how to design and evaluate both standardized and classroom-based test, familiarity with test processes and awareness of principles and concepts that govern the practice (i.e., ethics of the practice).

He also connected assessment literacy to a broader social and political context. Some researchers (e.g., Pill & Harding, 2013 and Volante & Fazio, 2007) define assessment literacy from the stakeholders' perspective, such as students and test developers. Gottheiner and Siegel (2012) added the ability to interpret the assessment and to take action based on the results among the required skills of AL, while Popham (2011) argued that AL should include, among other elements, the use of basic statistics for educational measurement.

Mertler (2003) used the classroom assessment inventory (CALI) to investigate the impact of experience on assessment literacy between two groups. The first group consisted of 197 in-service teachers, while the second group consisted of 67 pre-service teachers. The results show that experience has an impact on the in-service teachers as they performed better in administrating, scoring, and interpreting assessments' results. The pre-service teachers outranked their in-service teachers in the category related to developing valid grading procedures. The paper concluded that in-service teachers outranked their pre-service counterparts in five of the seven competency areas. According to the study, the difference is significant and in favor of in-service teachers, which highlights the importance of experience when it comes to assessment literacy.

Ashraf and Zolfaghari (2018) also assessed EFL teachers' assessment literacy and their reflective teaching. The authors used two questionnaires that were explicitly designed to assess language teachers' assessment literacy and covered three subscales, which were teachers' disposition about assessment, knowledge

about assessment, and performance (Ashraf & Zolfaghari, 2018). The second questionnaire was specifically designed to assess the teachers' reflective teaching. The questionnaire was based on six factors: cognitive, metacognitive, affective, practical, critical, and moral. They concluded that teachers' assessment literacy impacts their reflective teaching and can also be used to predict their teaching practices. Xu and Brown (2017) offered a framework covering the trajectory of professional development that, they claim, encompasses all aspects of teachers' educational growth. This was done through a scoping literature review and synthesis of the previously conducted studies.

The study identified seven competencies related to the base knowledge that teachers need to possess as part of their academic skills. Those competencies are:
1. Choosing assessment methods appropriate for instructional decisions
2. Developing assessment methods appropriate for instructional decisions
3. Administering, scoring and interpreting the results of both externally produced and teacher-produced assessment methods
4. Using assessment results when making decisions about individual students planning, teaching, developing curriculum and school improvement
5. Developing valid pupil grading procedures
6. Communicating assessment results to various stakeholders
7. Recognizing unethical, illegal, and inappropriate assessment methods and uses of assessment information. Xu and Brown (2017:150)

Some researchers added sociopolitical and sociocultural dimensions to the factors that impact assessment literacy in general and language assessment literacy, in particular. Fulcher (2012), for instance, states that "assessment principles and practices should be discussed within a much wider historical and social context" (p.125). Discussing assessment literacy, Scarino (2013) also stressed the importance of experience when it comes to AL as, according to the author, instructors tend to learn "on the job." Crusan, Plakans and Gebril (2016) also looked at the development of language teachers' assessment literacy. Their work is among the recent studies that have urged researchers and educators to look at specific contextual and experiential factors when investigating teachers' assessment literacy.

Finally, Alsomaani (2014) looked at the techniques used by the Saudi novice EFL teachers at 12 public middle schools. Though the study does not use the term 'assessment literacy,' it reported that the teachers' inadequate training in assessment during their undergraduate programs is to be blamed for the mismatch between the teachers' assessment practices and the students' needs. The study also revealed that the teachers obtained most of their assessment knowledge through their "on-the-job experience." This study, and others, point to the importance of continuous assessment training for in-service teachers if we want to ensure we have reliable assessment methods and that students have a better learning experience.

6. Results

In this section, we will report the statistical procedures that were used to analyze the data. A descriptive analysis was conducted for the seven composite scores based on the Standards. Inferential analyses including a t-test, evaluated at an alpha level of .05, was used to compare the English language to the Basic Science teachers' mean scores for each of the seven composite scores, and the total score for the entire instrument. A t-test was also used to compare the performance of the female to the male teachers in each of the seven standards. Finally, ANOVA test was used to look at the multiple variables of obtained qualifications and years of experience in teaching.

Table 1 gives the demographic and basic information regarding the sample's number, subject department, sex, qualifications they have obtained, and the number of years spent in teaching, i.e., experience.

Table 1: Demographic Profile

	Variable	N= 54	%
Department	Basic science	13	24.1
	English	41	75.9
Sex	Male	21	38.9
	Female	33	61.1
Qualification	BA or BS	4	7.4
	MA or MS	35	64.8
	Specialist	4	7.4
	PhD	11	20.4
Experience	1-5	11	20.4
	6-10	6	11.1
	11-15	13	24.1
	16-20	9	16.7
	21-25	3	5.6
	26-30	3	5.6
	31+	9	16.7

Table 1 shows that the majority of teachers were from the English department (n=41), and female faculty constitute 61.1% of the study sample (n=33). Regarding the academic qualifications, the table shows that 64.8% of teachers were MA or MS holders, while 20.4% of them were PhD holders. The participants, however, varied in the number of years they had spent in teaching. As can be seen from the table 1, 13 instructors out of the 54 participating in the study had spent 11-15 years in teaching. Nine of the participants, on the other hand, had spent more than 31 years in teaching. Three teachers reported that they had spent 21-25 years, and another three stated that they had had 26-30 years of experience in education. The second-largest number of instructors (n=11) reported that they had spent between 1 to 5 years in teaching.

Table 2: Descriptive statistics

Standard	Mean of overall assessment	SD
Choosing an assessment method	2.6	1.1
Developing assessment methods	2.7	1.4
Administering, assigning, and interpreting learning outcomes	2.0	1.1
Using assessment outcomes in decision making	3.0	1.0
Using assessment to determine levels of learning outcomes	2.2	1.3
Communicating assessment outcomes	2.6	1.2
Knowing unethical practices	1.7	1.1
Overall	**16.8**	**4.5**

CALI descriptive statistics are illustrated in the table 2. For the overall assessment literacy performance, the minimum obtained score was six, and the maximum obtained score was 25 out of 35. The mean of overall assessment was computed at 16.8, with 4.5 SD. The mean of overall statements was less than 20, which indicates poor classroom assessment literacy among the study sample. Breaking down the results by standards, it is found that the standard related to using assessment outcomes in the decision-making had the highest mean score while knowing unethical practices standard was found to receive the lowest mean score (1.7), indicating poor knowledge in this standard. The next stage of the analysis was dividing the scores of CAL into three levels, namely: Inadequate or needs improvement (lower than 60%), Fair (60-79%) and High (80% and higher). Table 3 summarizes the results based on this division.

Table 3: Teachers overall performance in Classroom Assessment Literacy

Standard	Poor		Fair		High	
	n	%	n	%	n	%
Choosing an assessment method	22	40.7	24	44.4	8	14.8
Developing assessment methods	23	42.6	14	25.9	17	31.5
Administering, assigning, and interpreting learning outcomes	36	66.7	12	22.2	6	11.1
Using assessment outcomes in decision making	17	31.5	21	38.9	16	29.6
Using assessment to determine levels of learning outcomes	32	59.3	13	24.1	9	16.7
Communicating assessment outcomes	31	57.4	9	16.7	14	25.9
Knowing unethical practices	43	79.6	9	16.7	2	3.7
Overall	**42**	**77.8**	**12**	**22.2**	**0**	**0**

The overall scores show that 77.8% of sample cases have a poor CAL level, while 22.2% of them have a fair CAL level. It is essential to mention here that although some participants scored high in some of the listed standards, none of them obtained a high overall score (see table 3). When analyzing the results by standards, it can be seen that standard 4, relating to the use of assessment outcomes in decision-making, obtained the highest percentage. The standard

related to knowing unethical practices, 7th in the table 3, scored the lowest. Almost 80% of the sample showed poor knowledge about ethical practice in assessment. Developing assessment methods and using the assessment scores to make decisions received a relatively high percentage of correct answers, 31.5%, and 29.9 %, respectively, compared to the rest of the standards.

Table 4: Differences in CALI performance between departments

Standard	Mean Basic science	Mean English	*P-value
Choosing an assessment method	2.31	2.66	0.332
Developing assessment methods	1.62	3.00	0.001**
Administering, assigning, and interpreting learning outcomes	1.85	2.10	0.463
Using assessment outcome in decision making	2.54	3.12	0.077
Using assessment to determine levels of learning outcomes	1.38	2.44	0.008**
Communicating assessment outcomes	2.08	2.71	0.097
Knowing unethical practices	1.77	1.73	0.913
Overall	**13.54**	**17.76**	**0.003****

* independent t-test; ** p-value is less than 0.05

The t-test for independent samples was conducted to determine whether there was a significant difference between the CAL levels of Basic Science and English department faculty. The results show that the p-value was less than 0.05 (P<0.05), which indicates a significant difference between the two departments in terms of overall CAL scores. The mean scores in the seven standards of English language faculty were found to be higher than the mean scores of Basic Science faculty. This result indicates a higher level of knowledge in CAL among English language department members. In particular, the results show a significant difference in standards two (i.e., developing assessment methods) and six (i.e., using assessment to determine levels of learning outcomes). The difference is in favor of the English faculty. To answer the question related to the participants' sex, a t-test was calculated, as shown in Table 5.

Table 5: The difference in CAL based on sex

Standard	Mean Females	Mean Males	*P-value
Choosing an assessment method	2.64	2.68	0.876
Developing assessment methods	3.14	2.84	0.465
Administering, assigning, and interpreting learning outcomes	2.00	2.21	0.553
Using assessment outcomes in decision making	3.36	2.84	0.107
Using assessment to determine levels of learning outcomes	2.68	2.16	0.120
Communicating assessment outcomes	2.86	2.53	0.354
Knowing unethical practices	1.77	1.68	0.776
Overall	**18.45**	**16.95**	**0.233**

* independent t-test

The summarized results in the table 5, show an insignificant difference in CAL levels between male and female faculty in the English and Basic Science departments (p>0.05). However, a quick look at the table 5 shows that the female staff members performed better than their male counterparts.

Table 6: CAL performance based on academic qualification difference

Standard	Mean BA or BS	Mean MA or MS	Mean Specialist	Mean PhD	*P-value
Choosing an assessment method	2.75	2.60	2.75	2.36	0.903
Developing assessment methods	2.25	2.94	2.25	2.09	0.247
Administering, assigning, and interpreting learning outcomes	1.50	2.14	2.75	1.64	0.194
Using assessment outcomes in decision making	3.25	3.00	2.50	3.00	0.776
Using assessment to determine levels of learning outcomes	2.75	2.40	2.25	1.27	0.055
Communicating assessment outcomes	3.50	2.66	2.00	2.09	0.150
Knowing unethical practices	1.50	1.74	2.25	1.64	0.759
Overall	**17.50**	**17.49**	**16.75**	**14.09**	**0.185**

* F test result from ANOVA

Table 6 shows no significant difference in the participants' performance in the seven CAL standards based on their academic qualifications. Nevertheless, the table shows that the performance of the bachelor's and master's degree holders, whether in Arts or Science subjects, was slightly better than that of the PhD holding staff members.

Table 7 shows the difference among the participants based on their overall scores in CAL and the length of their experience as educators. Table 7 illustrates the relationship between the teachers' general level of assessment literacy, as reflected by CALI, and the number of years they have spent in education, i.e., years of experience.

Table 7: CAL performance based on number of years of experience

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2221.598	6	370.266	1.609	.166
Within Groups	10583.874	46	230.084		
Total	12805.472	52			

In Table 7, it can be seen that there is no significant difference in the classroom assessment literacy level among the participants based on their years of experience (p-value is >0.05).

7. Discussion

Our participants' overall performance shows that the general level of assessment literacy cannot be considered adequate, especially when compared to the importance given to assessment in the program and the role it plays in determining the students' future in this context. The average for the seven standards together is 16.8 out 35, which indicates a lack of AL knowledge. The results show the need for further training in this area, as it seems that the teachers might not be adequately prepared to effectively assess the students' learning. However, these results do not come as a surprise as, despite the different AL evaluation methods and contextual differences, they concur with what is reported in the literature. Most of the previous studies show that teachers' general performance in AL tests is not satisfactory (e.g., Perry, 2013; Brown, 2004; Mertler, 2003; Plake, Impara & Fager, 1993).

Although using a different assessment instrument, the same result was reported by Plake, Impara and Fager (1993). In their study, they used the Teacher Assessment Literacy Questionnaire (TALQ) as part of a two-part instrument. The instrument, similar to the one used in this research, consisted of 35 items that measured the seven competencies mentioned in the standards. Their study also reflected the poor assessment literacy level among the participants. Maclellan (2004) also reached the same conclusion, although that study used a different approach to data collection. The researcher tested the teachers' knowledge by analyzing each teacher's written scripts. In a comparable study, Muhammad and Bardakçi (2019), reported a less than satisfactory level of AL among Iraqi EFL teachers. The authors compared the teachers' results with the results reported in the published literature and concluded that Iraqi teachers scored the lowest internationally (mean=16). The average reported score globally, they stated, is 17 to 24 out of 35. As can be seen, the result of our study falls between the Iraqi score and the lowest reported score internationally.

Most of the research studies attribute teachers' lack of knowledge about assessment literacy to inadequate preparation in assessment during their in-service teacher education programs, or perhaps a complete lack of any such training at all (e.g., Alsomaani, 2014; Schafer, 1993). Herrera and Macias (2015) is one of the studies that, although looking specifically at language assessment literacy, emphasized that teachers need to develop their overall assessment literacy and that special attention should be placed on improving the quality of AL courses during pre-service educational programs. They argued that teachers are "expected to have a working knowledge of all aspects of assessment to support their instruction and to effectively respond to students, parents, and the school community" (p.303).

However, when we look closely at our participants' results in the each of the individual standards, we notice that their performance varied across the different standards. For instance, it seems that our participants displayed better knowledge of the standards related to using assessment outcomes in decision-making. That was followed by the standard of administering, assigning, and interpreting the learning outcomes. On the other hand, they scored the lowest in the standard dealing with the knowledge related to unethical practices. The

same result was obtained in the studies conducted by Perry (2013) and Muhammad and Bardakci (2019). The relatively good outcome in standard four can perhaps be attributed to the constant contact new teachers have with their more experienced colleagues, a deliberate placement strategy in the department. It could also be attributed to the standardized system of assessment formatting employed by the university.

To answer the question regarding the relationship between the level of AL and the subject department, the results indicate a significant difference in the overall performance between the English language and the Basic Science teachers. The results showed that English language teachers tend to possess more assessment knowledge than their Basic Science counterparts. Furthermore, a closer look at the detail of the teachers' performance across the standards shows that the difference is evident mainly in the standards related to developing assessment methods and using assessment to determine levels of learning outcomes. This result could be attributed to the nature of the courses being assessed.

Additionally, the fact that the university requires the English language teachers to be graduates of specialized EFL teaching educational programs could also be considered as a contributing factor to this difference. This tends not to be the case with the Basic Science teachers, who are mostly graduates of Basic Science programs and usually not required to have any specific teacher training. Interestingly Alkharusi (2009) found that teachers who specialized in academic areas such as English and Basic Science were better than those who specialized in performance areas such as art and physical education.

The performance of female members of staff was roughly equal to that of their male counterparts indicating that the sex of the participants in our study, i.e., the third research question, did not have any impact on the participants' general performance. This result has led us to rule out sex as a factor when it comes to assessment literacy in Saudi higher education. Contrary to this study, Alkharusi (2009) found that "measurement and test knowledge of pre-service teachers tended to vary as a function of gender and major" (p.15). Males, in his data, have more knowledge in assessment than females.

However, this study shared a similar conclusion to Alkharusi (2009), which showed that the participants' academic qualifications did not influence their general performance. This research reflected no significant difference among the teachers' general level of AL when it came to their educational qualifications. It was found that the average scored mark ranged between 14.09 and 17.50 in the seven standards. In other words, there was no difference in the LA level among teachers with a Bachelor's, Master's, or a Doctorate.

Finally, although the participants' length of experience as teachers varied from 1-5 years to more than 30 years, this difference was not reflected in their overall performance. i.e., they all had low scores. This could be a result of the teachers' tendency to learn from each other's assessment experience, particularly where less experienced teachers look to their more experienced colleagues for guidance. In many cases, this practice has a positive impact. However, the

novice teachers might also copy the formats of older versions of tests without possessing the necessary knowledge regarding the purposes for which those exams were developed. While such practices lead to the sharing of positive knowledge, it can also contribute to the teachers' transference of erroneous practices (Stiggins, 1988). This underlines the importance of ensuring that new teaching staff is adequately mentored and that both old and new faculty are encouraged to engage in continuous professional development.

An additional factor that may explain the lack of differentiation in results according to years of experience may be that the survey asked about teaching experience and did not ask specifically about what kind of experience each teacher had had. A teacher may have had 20 years teaching part-time, perhaps as little as 2 hours a week, or in an institution where little or no standardized assessment was used. In such cases, the 20 years of teaching would give them limited opportunities to learn about assessment. To conclude, it is worth mentioning that this study has its limitations as the number of participants is relatively small. The use of a single instrument to assess AL is yet another possible source of limitation. Fulcher (2012) reported that the use of questionnaires to assess LAL might elicit skewed responses. Nevertheless, the study is a good indicator and a step in the right direction towards improving the teaching and learning quality in the Saudi context.

The result of the overall performance in AL in this study indicates poor knowledge of the principles underlying classroom assessment, which is a wake-up call to educators in the higher education sector. The results also bring to our attention the need to emphasize the importance of attending continuous assessment training for in-service teachers. Furthermore, policies and ethics that govern assessment should also be placed at the heart of any training related to continuous professional development in the Saudi context. The importance of continuous assessment literacy training during the teachers' years of service was also emphasized by Popham (2006). It is also important to note that the instrument used for this study is not context-specific. It is an instrument designed according to the American education context, which is significantly different from the Saudi one. For that reason, we would like to call for more studies investigating assessment literacy in the Saudi context, but would suggest the use of an instrument that is context-sensitive, and that takes the Saudi education policies and practices into consideration (Yan, Zhang & Fan, 2018). The use of CALI in this study is a good start and an indicator- of the teachers' basic knowledge. However, it should not be used on its own and should also be modified to include more context-specific elements.

8. Conclusion
Assessment is one of the most critical aspects of education; therefore, assessment literacy among teachers is considered one of the most crucial skills to be mastered at all educational levels. This study evaluated the AL level of teachers at a Saudi higher education institution preparatory program. It used the Classroom Assessment Literacy Inventory developed by Mertler and Campbell (2005). The study found that the general level of AL among teachers in the Saudi higher education program is less than satisfactory. It also reported that factors

such as sex, the qualification that the participants obtained, or length of years of teaching experience have no impact on their assessment literacy. Nevertheless, the English language teachers scored higher than their Basic Science colleagues, which highlighted the role that the subject department played in the level of AL. This might be attributed to different factors such as pre-service preparation programs or the nature of the courses, as explained in the discussion section. The overall result placed the Saudi teachers within a level of performance close, if not similar to, most of the reported results in different contexts around the world. This study is additional empirical proof, adding to the already published work, calling for continuous assessment training for in-service teachers. Significantly, it also adds emphasis to the importance of including contextual elements to AL (Crusan, Plakans & Gebril, 2016; Yan, Zhang & Fan, 2018.).

9. Contextualization of the result

The inadequate level of assessment literacy among our participants does not mean inadequate assessment practices in the institution. It is imperative to mention here that at KSAU-HS assessment is strictly regulated, and the university uses a variety of methods to ensure that all assessments adhere to international best practice standards, to ensure validity and fairness. As per institutional policy, instructors are required to align their test items with the learning outcomes of each course. This alignment practice, as stated by Yon, Zhang and Fan (2018), can help teachers to "develop sensitivity to the content validity of their assessments because of the mandate to assess students' learning outcome against the standardized curriculum" (p. 165). The values of reliability coefficients of the exam samples that were taken from the different preparatory level courses indicate that all items are highly reliable. The accepted overall reliability values of exams in KSAU-HS is (>0.7), which provides additional evidence of the importance and the influence that other factors such as context, belief, and experience have on teachers' assessment literacy Deluca, LaPointe-McEwan and Luhanga (2016).

10. Recommendations for educators

1. Assessment education should be part of the teachers' continuous development practices.
2. Assessment related workshops should be introduced throughout the academic year, particularly every time a change in the assessment scheme is introduced (Deluca, LaPointe-McEwan & Luhanga, 2016).
3. Teachers' assessment literacy requires further recognition in the Saudi higher education system and should be placed at the heart of any research tackling higher education issues.
4. Assessment literacy evaluation instruments should be context-specific and reflect the teachers' beliefs, practices, and educational policies.

11. References

Alkharusi, H. A. (2009). Correlates of Teacher Education Students' Academic Performance in an Educational Measurement Course. *International Journal of Learning: Annual Review, 16*(2), 1-16. https://doi.org/10.18848/1447-9494/CGP/v16i02/46111

Alsamaani, A. (2014). Evaluating classroom assessment techniques of novice Saudi EFL teachers. *Journal of Arabic and Human Sciences, 7*(2), 63-81. https://doi.org/10.12816/0009602

Ashraf, H., & Zolfaghari, S. (2018). EFL Teachers' Assessment Literacy and their Reflective Teaching. *International Journal of Instructions, 11*(1), 425-436.

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318. https://doi.org/10.1080/0969594042000304609

Brookhart, S. M. (2011). Starting the Con versation About Grading. *Educational Leadership*. Retrieved from https://www.greatschoolspartnership.org/wp-content/uploads/2016/11/Starting-the-Conversation-about-Grading-2.pdf

Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs and practices. *Assessing Writing, 28*, 43-56. https://doi.org/10.1016/j.asw.2016.03.001

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education, 34*(5/6), 576-591. https://doi.org/10.1080/01626620.2012.730347

Deluca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of International standards and measures. *Education Assessment, Evaluation and Accountability, 28*(3), 251–272.

Earl, L. M., & Katz, S. (2006). *Leading school in a data-rich world: Harnessing data for school improvement.* Corwin Press.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language AssessmentQuarterly, 9*(2), 113-132.
https://doi.org/10.1080/15434303.2011.642041

Gottheiner, D., & Siegel, M. A. (2012). Experienced Middle School Science Teachers' Assessment Literacy: Investigating Knowledge of Students' Conceptions in Genetics and Ways to Shape Instruction. *Journal of Science Teacher Education, 23*(5), 531-557. https://doi.org/10.1007/s10972-012-9278-z

Herrera, L., & Macias, D. (2015). A call for language assessment literacy in the education and development of Teachers of English as a foreign language. *Colombian Applied Linguistics Journal, 17*(2), 302-312. http://dx.doi.org/10.14483/udistrital.jour.calj.2015.2.a09

Inbar-Lourie, O., & Levi, T. (2020). Assessment Literacy as Praxis: Mediating Teacher Knowledge of Assessment-for-Learning Practices. In M. Poehner & O. Inbar-Lourie (Eds.) *Toward a Reconceptualization of Second Language Classroom Assessment*. Educational Linguistics, Vol. 41, (pp. 241-259). Springer, Cham. https://doi.org/10.1007/978-3-030-35081-9_12

Lam, R. (2015). Language assessment training in Hong Kong: Implication for language assessment literacy. *Language Teaching, 32*(2), 169-197. https://doi.org/10.1177/0265532214554321

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualization. *Teaching and Teacher Education, 20*(5), 523-535. https://doi.org/10.1016/j.tate.2004.04.008

Malone, M. (2013). The Essentials of Assessment Literacy: Contrasts between Testers and Users. *Language Testing, 30*(3), 329–344. https://doi.org/10.1177/0265532213480129

Mertler, C. (2003). Preservice Versus Inservice Teachers' Assessment Literacy: Does Classroom Experience Make a Difference. *Annual meeting of the Mid-Western Educational Research Association.* Columbus: University of Maryland.

Mertler, C. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education, 33*(1), 49-64.

Mertler, C. A., & Campbell, C. (2005, April). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory. *American Educational Research Association.* Montreal, Canada.

Muhammad, N., Hama, F., & Bardakçı, M. (2019). Iraqi EFL Teachers' Assessment Literacy: Perceptions and Practices. *Arab World English Journal (AWEJ), 10*(2). http://dx.doi.org/10.2139/ssrn.3418574

Paterno, J. (2001). *Measuring success: A glossary of assessment terms*. Building cathedrals: Compassion for the 21st century.

Perry, M. L. (2013). Teacher and principal assessment literacy. Retrieved from https://scholarworks.umt.edu/etd/1391

Pill, J., & Harding, L. (2013). Defining the Language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(2), 381-402. https://doi.org/10.1177/0265532213480337

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*, 10-12. https://doi.org/10.1111/j.1745-3992.1993.tb00548.x

Popham, W. J. (2006). Assessment for Learning: An Endangered Species?. Educational Leadership, 63(5), 82-83.

Popham, W. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator, 46*(4), 265-273. https://doi.org/10.1080/08878730.2011.605048

Rogier, D. (2014). Assessment Literacy: Building a Base for Better Teaching and Learning. *English Teaching Forum, 52*(3), 2-13.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, *30*(3), 309-327. https://doi.org/10.1177/0265532213480128

Schafer, W. D. (1993). Assessment in teachers' education. *Theory into practice, 32*(2), 118-126. https://doi.org/10.1080/00405849309543585

Stiggins, R. J. (1988). Revitalizing Classroom Assessment: The Highest Instructional Priority. *The Phi Delta Kappan, 69*(5), 363-368.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77*(3), 238.

Stiggins, R. J. (1999). Assessment, Student Confidence, and School Success. *The Phi Delta Kappan, 81*(3), 191-198.

Valencia, S. (2002). *Understanding Assessment: Putting Together the Puzzke(1st ed.).* Boston: Houghton Mifflin. Retrieved from https://www.eduplace.com/state/author/valencia.pdf

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*, 749-770.

Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but…": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, *74*, 158-168. https://doi.org/10.1016/j.system.2018.03.003

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 149-162. https://doi.org/10.1016/j.tate.2016.05.010

Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China. *Language Testing and Assessment*, *6*(1), 133-158.