

# Investigating Saudi University EFL Teachers' Assessment Literacy: Theory and Practice

Muhammad Umer<sup>1</sup>, Mohamad Hassan Zakaria<sup>1</sup> & Moayad Ahmad Alshara<sup>1</sup>

<sup>1</sup> Foreign Language Department, College of Arts, Taif University, Taif, Saudi Arabia

Correspondence: Muhammad Umer, Foreign Language Department, College of Arts, Taif University, Haweiah, Taif, Saudi Arabia. E-mail: m.khan@tu.edu.sa; m.umer07@gmail.com

Received: January 18, 2018 Accepted: February 12, 2018 Online Published: February 21, 2018

doi:10.5539/ijel.v8n3p345 URL: <http://doi.org/10.5539/ijel.v8n3p345>

## Abstract

Teacher assessment literacy (TAL) is believed to have positive impact on student learning outcomes. Therefore, attempts are made, especially, in advanced educational contexts to increase TAL. In the context of Saudi higher education, available empirical evidence indicates that EFL teacher assessment literacy is replete with loopholes. This mixed-method research investigated Saudi EFL teachers' construction of assessment tasks, the influence the tasks had on students' learning and the extent to which teachers' assessment practices were in alignment with recommended assessment practices. The data were collected through analyzing teachers' summative assessment tasks and a student survey with both close and open-ended questions. Apart from the participants' responses to the open-ended questions of the survey, the data went through quantitative data analysis for frequencies and percentages. The findings revealed a serious incongruity between teachers' assessment tasks and course learning outcomes. For instance, higher order learning outcomes were not assessed at all. Most of the tasks were selected-response questions (SRQs). As confirmed by the survey data, the assessment tasks mainly triggered memorization as a learning strategy. Therefore, suggestions are made that university teachers' professional development with particular focus on their assessment literacy is placed at the center of higher education policies. Without valid assessment in place, the edifice of Saudi (higher) education system may lose its efficacy.

**Keywords:** assessment literacy, learning outcomes, assessment tasks design

## 1. Introduction

Teacher assessment literacy, an adequate understanding and desirable application of the principles of sound assessment, is a fundamental professional requirement in all advanced educational systems (DeLuca, LaPointe-McEwan, & Luhanga, 2016; DeLuca 2012; Popham 2013; Volante & Fazio, 2007). Assessment literacy, in particular, involves teachers' ability to construct and implement high quality assessments instruments (Plake, 2015; Popham, 2004; Stiggins, 2002, 2004). The quality of classroom practices is quite often predicated on the assessment policies in practice in the concerned educational context (Lukin, Bandalos, & Eckhout, 2004, p. 26). Bearing in mind the correlation between TAL and student learning, this study attempted to provide a description of Saudi university EFL teachers' assessment literacy and determine its impact on student learning outcomes. This research was viewed significant as its findings might prove helpful in improving teachers' assessment practices, particularly constructing assessment tasks in the context of this study and similar settings elsewhere.

In Saudi universities, undergraduate programs' teachers are responsible for the assessment of their students' learning outcomes. Course teachers design assessment tasks, administer them to students and grade students' answers without any moderation. The assessment regime comprises both formative and summative components. The formative assessment is mostly given 40% weightage which is further divided into quizzes, presentations, assignments and midterm examination whereas summative assessment is allotted 60% marks. The final examination is in the form of a single paper-and-pen examination for all courses including skill courses like speaking. In addition, a major part of the entire assessment is SRQs. That is, true or false, multiple choice questions, matching items, choosing right words from a list of words for filling in incomplete sentences. Course teachers design all the assessment tasks which mostly do not go through peer review for reliability and validity check. Students' answers are graded by their course teachers and the graded answers are not reviewed for grading reliability. Teachers are not required to share their assessment criteria and marking rubrics with other fellow teachers and even students which often becomes idiosyncratic (Green, 2013). Despite the apparent

freedom the teachers have in assessing their students, there appears to be a strong pressure on the teachers to prepare student-friendly assessment tasks as failure of many students is deemed undesirable. Based on the contextual situation depicted above, it appears very important to examine Saudi EFL teachers' assessment literacy, their assessment task designing practices and determine the kind of impact the nature of questions have on student learning outcomes. The anticipated findings might help teachers give more authentic assessment tasks to help students develop higher order skills.

## 2. Literature Review

Assessment literature indicates that adequate teacher assessment literacy and its fitting application facilitates higher order learning. Assessment literacy, however, is not a straightforward concept. It has many aspects such as teacher knowledge of assessment principles, selection of assessment methods, skills to develop appropriate assessment tasks and using them for instructional purposes, administration, scoring and interpretation of assessment results, using assessment results for decision making about teaching, learning and material development, sharing assessment criteria with students and conveying valid assessment results to all stakeholders, and finally recognizing unethical assessment practices. If one of the assessment components listed above goes off the target it can have hostile effect on the entire teaching and learning processes. For example, evidence indicates that teachers' weak selection of assessment methods or tasks can have a strong negative impact on student learning outcomes (Galluzzo, 2005; Volante & Fazio, 2007; Umer, 2015 & 2016). Such undesirable evidence signposts gaps between teachers' practices and recommended assessment norms (Plake, 2015). However, the quantity of evidence of assessment literacy worldwide is still insufficient to make blind generalizations (DeLuca, 2016; Volante & Fazio, 2007). Thus, further research is asked for even in the developed educational contexts, for instance, North America, the UK, Europe, Australia, and New Zealand to help teachers use assessment for improving students learning outcomes (Birenbaum, DeLuca, Earl, Heritage, Klenowski, Looney, ... Wyatt-Smith, 2015).

Empirical evidence indicates that assessment task designs influence how learners learn. Tests that are in congruence with learning outcomes mostly result in high-order learning—analysis, synthesis and evaluation etc. (Benedetti, 2006; Ferman, 2004, Saif, 2006; Stecher, Chun, & Sheila, 2004; Manjarrés, 2005; Muñoz & Álvarez, 2010; Cheng, 1997). For example, Benedetti (2006) noted that a video listening test compared to audio listening test to assess students listening skills proved more reliable thanks to its authenticity, i.e., visual impact of the test. Similarly, with communicative learning outcomes and imparting a high level of language proficiency to learners, the Oral Matriculation test in Israel was found having a strong intended influence on students' learning (Ferman, 2004). It caused learners began to focus on oral skills (the intended outcome of the test) instead of reading for the test. The test specifications of Saif's study (2006) in Victoria University in Canada closely resembled the communicative skills required of the international teaching assistants. The experimental group showed far better results than the control group. Stecher et al. (2004) investigated the effects of assessment driven reform of a writing test in Washington State. The results indicated that the changed test specifications positively influenced learning processes. Manjarrés (2005) studied how a newly introduced English language test in the state Examination in Barranquilla, Colombia, positively affected students' learning positively. The student's awareness of the test specifications made them focus more on the target skills rather than on learning isolated language items. The findings of Muñoz & Álvarez (2010) have substantiated the results of previous research that a strong correlation between learning outcomes and assessment tools increases the achievement of students' learning. However, the researchers have strongly recommended that constant guidance should be provided to students in terms of what the assessment design requires of them. Cheng (1997) from the investigation of the washback impact of the revised Hong Kong Certificate of Education Examination in English by the Hong Kong Examinations Authority has concluded that a strong overlap between assessment task design and course learning outcomes does bear a clear beneficial impact on students' learning. However, it is the concerned teachers' assessment and teaching experience, learners' expectations, leadership role that constitute the overall effect of a the assessment.

On the other hand, it has also been reported that assessment tasks that do not accord with the learning outcomes of a given course will cause lower-order learning i.e., memorization and remembrance of knowledge (El-Ebyary, 2009; Gijbles, Segers, & Struyf, 2008; Gijbels & Dochy, 2006; Scouller, 1998). These studies found that because of test specifications, learners had to focus mainly on lexical and grammatical accuracy that was against the intended learning outcomes in the respective courses. For instance, Gijbles et al. (2008) reported that the nature of assessment tasks involved lower level cognitive abilities of the students enrolled in a Psychology course in a Belgium university. Gijbels & Dochy (2006) found that students did not show any preference for those assessment methods that examined higher order cognition as most of the assessment tools targeted surface level

learning. In her study in the University of Sidney, Scouller (1998) examined how SRQs caused surface level learning. Thus, it transpires that for meticulous achievement of learning outcomes, assessing students' learning through valid assessment instruments is indispensable because assessment tasks that are off the learning outcomes are simply sheer wastage of time, efforts, resources and the future of the learners.

Apart from Niveen, Elshawa, Abdullah, & Rashid (2017) and Hakim (2015) who have studied university teachers in Malaysian and Saudi universities respectively, the rest of TAL studies conducted in different parts of the world have investigated school teachers' assessment literacy. For example, Plake & Impara (1992) and Plake et al. (1993) examined assessment knowledge, attitude and practices of in-service school teachers in the United States. Arce-Ferrer, Cab, & Cisneros-Cohernour (2001) studied school teachers' perspectives and familiarity with educational assessment and how they chose assessment methods, their use of assessment results for teaching learning. Susuwele-Banda (2005) investigated teachers' perceptions and practices in Malawi about classroom assessment. In Turkey, Ogan-Bekiroglu (2009) examined 46 teachers' assessment competence and attitudes. DeLuca & Klinger (2010) surveyed 288 Canadian trainee teachers' knowledge of assessment. Koloi-Keaikitse (2012) through a questionnaire investigated 691 teachers of primary and secondary schools from Botswana. Through a quasi-experimental research, Lukin, Bandalos, & Eckhout (2004) studied how positively assessment training impinged on both pre-service and in-service teachers' confidence, skills and knowledge of educational assessment. In the most recent work on assessment literacy by DeLuca et al. (2016) is an account of school teachers' assessment literacy measures taken in educationally advanced contexts. These studies of school teacher assessment literacy suggest that increasing assessment literacy bears positive impact on teachers' performance (Volante & Melahn, 2005; Koloi-Keaikitse, 2012; Lukin et al., 2004). However, University teachers' assessment literacy and how it affects students learning outcomes particularly form educationally developing nations is substantially under-explored. In addition, though assessment literacy incorporates both assessment knowledge and practice, most empirical evidence gathered from different contexts is concerning school teachers' self-reported information about their knowledge and practices. The current study therefore was designed to find out how Saudi EFL teachers' assessment practices looked like rather than examining their (self-reported) knowledge. Empirical evidence from Saudi higher education context, though very limited, shows that university teachers' assessment practices are far from recommended assessment practices (Ezza, 2017; Umer, 2015 & 2016) even though if teachers are theoretically sound in some settings (Hakim, 2015). The present study, therefore, attempted to provide a description of how Saudi university EFL teachers' assessment task designing practices and any observable impact the tasks had on student learning. Thus, the following three questions guided this study:

- 1) To what extent did the teachers' assessment tasks cover all learning outcomes?
- 2) How do the assessment tasks affect students' efforts of learning?
- 3) To what extent are Saudi university EFL teachers' assessment practices in line with recommended assessment principles?

### **3. Methodology**

The data in this study were gathered through three instruments; analysis of teacher designed final exam papers, students' responses to a Likert scale questionnaire and student interviews. The final exam question sheets were analyzed to determine the extent teachers' assessment tasks were congruence with course learning outcomes. In order to determine the extent the assessment tasks were valid, the given tasks of each exam sheet were mapped against the corresponding course learning outcomes. Examining all 36 courses of the BA program was too much for a small-scale study. Therefore, nine courses three from each domain (i.e., literature, linguistics and skills courses) as a representative sample were selected for analysis. The nine courses included situational English, IELTS, paragraph writing (skills), phonetics, semantics, morphology (linguistics), modern English drama, nineteenth century novel, and modern poetry (literature). Thus, the number of courses constitutes 25% of the total number of courses. The data were collected in 2017 based on one of the available programs offered. Constructing valid assessment tasks is an indispensable element of teachers' assessment literacy, particularly where course teachers are responsible for assessing their learners. Authentic and valid assessment tasks, a natural output of strong assessment literacy, culminate into improved achievement of course learning outcomes. Therefore, the other two instruments were employed to determine the impact of the teachers' assessment literacy on learners. The questionnaire was administered to about 600 undergraduate students studying a BA program to get panoramic view of what students studied and how they studied, and the type of assessment tasks they preferred. Of them, 527 responded. To know further about how and what the students studied, 16 students with more than 3 GPA were interviewed.

The examination analysis began with getting familiarized with the data, both the course specifications forms of the selected courses and the examination manuscripts. The data analysis began with categorizing the nine courses according to their domains; skills, linguistics and literature. Afterwards, the questions were put into the two types: selected response questions and constructed response questions. Then, the questions of each course were mapped against its learning outcomes for uniformities and disparities. Based on National Commission for Academic Accreditation and Assessment (NCAAA) course specifications form, the learning outcomes are subdivided into FIVE learning domains: 1) knowledge, 2) cognitive, 3) interpersonal and responsibility, 4) communication, information technology, numerical, and 5) psychomotor. The initial results were given a senior colleague for cross checking and validation to see if the researchers counting and statistics matched the documents analyzed. The questionnaire data were analyzed using SPSS for frequencies and percentages. And finally, the participants' qualitative responses to the open-ended questions of the survey were classified under relevant categories and themes.

#### 4. Findings

Table 1-5 contain the results of question papers analysis. Table one shows that 72% of the assessment tasks were selected response questions (objective type) whereas 28% were constructed response questions (essay questions). The linguistics courses had very few CRQs, i.e., only the literature courses included essay questions. Apart from morphology, the rest of the linguistics and skill courses contained SRQs only. Out of the total marks of the nine courses (=540), one fourth of the marks, i.e., 75% (=405), that is, 72% of the questions, were allotted to the SRQs. Another worth noticing point in Table 1 is the number of questions across questions varying from 2 to 8 that suggest a strong inconsistency.

Table 1. Number of tasks, types of tasks and marks allocation course wise

Course Name	Selected Response Questions		Constructed Response Questions		Total Marks
	No of Tasks	Marks	No of Tasks	Marks	
1 Situational English	6	60	0	00	60
2 IELTS	2	60	0	00	60
3 Paragraph Writing	4	60	0	00	60
4 Phonetics	2	60	0	00	60
5 Semantics	6	60	0	00	60
6 Morphology	6	54	2	06	60
7 Modern English Drama	0	00	4	60	60
8 Nineteenth Century Novel	3	30	1	30	60
9 Modern Poetry	1	21	3	39	60
Percentage of Marks	29 (72%)	405 (75%)	10 (28%)	135 (25%)	540 (100%)

Table 2 gives information about the amount of space the students were expected to use for the CRQs. In "Morphology" examination, task 1 and 2 were CRQs with three marks each and a space of two lines for each task. In the course of "Modern English Drama", all four tasks were CRQs with 15 marks each. However, the space provided on the answer sheet varied. For the first and fourth task, only five double spaced lines were given and for tasks number two and three 23 and 21 lines were available respectively. The course "Nineteenth Century Novel" had one CRQ with four sub-topics, each of which was expected to be answered on four lines. Finally, the course of "Modern Poetry" had ¾ tasks each with 13 marks and 9, 10 and 8 lines space provided for each task. The standard deviation of 7.4 shows how big a difference was noticed among the tasks regarding what students were expected to write, suggesting a high degree of inconsistency.

Table 2. Mixed signals sent to students by constructed response questions

Course	No of tasks	Marks	Space provided (number of lines)
1 Morphology	1	3	2
	2	3	2
2 Modern English Drama	1	15	5
	2	15	23
	3	15	21
	4	15	5
3 Nineteenth Century Novel	1	30 (4 topics)	4
4 Modern Poetry	1	13	9
	2	13	10
	3	13	8
Average No of lines per task	13	10.38	8.9
Standard Deviation			7.4

Table 3 shows the degree of overlap between course-learning outcomes and assessment tasks. It can be seen that more than two-third of the learning outcomes were not assessed. The examinations of four courses, i.e., IELTS, Paragraph Writing, Modern English Drama and Nineteenth Century poetry, did not measure any of the learning outcomes. The exams of Phonetics and Semantics covered single learning outcome. Modern poetry is the only course that assessed more than half of its clearing outcomes.

Table 3. Overlap between assessment tasks course-learning outcomes

Course Name	No of Tasks	No of CLOs	No of CLOs covered	No of CLOs not covered
Situational English	6	6	2	4
IELTS	2	4	0	4
Paragraph Writing	5	5	0	5
Phonetics	2	5	1	4
Semantics	6	4	1	3
Morphology	6	5	2	3
Modern English Drama	4	2	0	2
Nineteenth Cent. Novel	3	2	0	2
Modern Poetry	10	10	6	4
Percentage of CLOs covered		43 (100%)	12 (28%)	25 (68%)

Chart 1 indicates how much the learning outcomes of the three skill courses were covered. The results show that apart from situational English, the rest of the courses measured something other than what the examinations were supposed to measure. The assessment instrument of situational English was partially valid because it assessed only two learning outcomes out of six. Assessment of skill courses have to be highly valid. Not assessing the required skill sends wrong signals to learners who stop to develop the required skills; rather, they start learning for tests only.

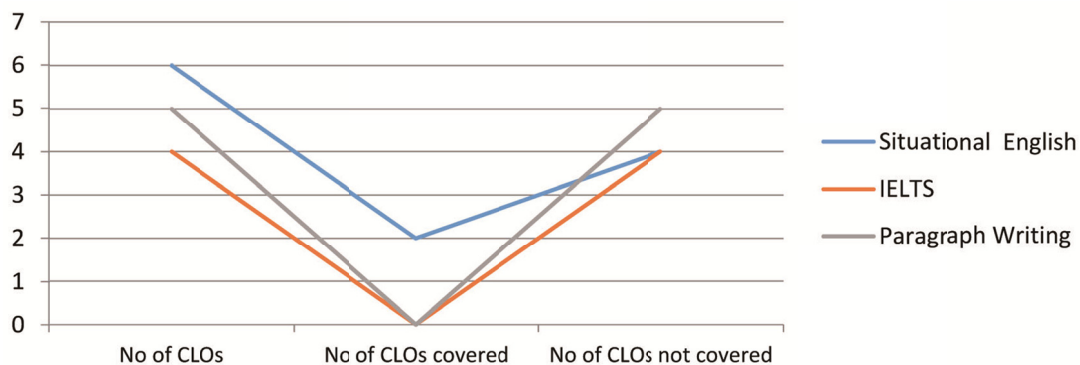


Figure 1. Coverage of skill courses' learning outcomes

Table 4 includes a qualification-based comparison of assessment task designing and coverage of course-learning outcomes by instructors from the three domain of curriculum. First, it can be seen that instructors with PhD as well as MA qualifications share the way they approach assessment task designing. Out of the nine courses, as shown in table 4, six instructors held PhD degrees. Five in linguistics or applied linguistics and one in literature whereas three of them were master degree holders. On the whole the instructors with PhD qualifications had covered 4%, i.e., 4 out of 25, of the learning outcomes while the MA holders covered 41%, i.e., 7 out of 17, learning outcomes. However, this difference though statistically quite big seems practically insignificant because, as displayed in table 1, the majority of the assessment tasks were SRQs, that might have minimized the desirable effect of assessment on learning.

Table 4. Qualification-based comparison of assessment tasks coverage of course-learning outcomes among instructors

Course Name	Instructor Qualification	Number of Tasks	Total No. of CLOs	Number of CLOs covered	Number of CLOs not covered
1 Modern English Drama	PhD	4	2	0	2
2 IELTS	PhD	2	4	0	4
3 Paragraph Writing	PhD	5	5	0	5
4 Phonetics	PhD	2	5	1	4
5 Semantics	PhD	6	4	1	3
6 Morphology	PhD	6	5	2	3
Percentage of CLOs covered		25	25(100%)	4(16%)	21 (84%)
7 Situational English	MA	6	6	2	4
8 Nineteenth Century Novel	MA	3	2	0	2
9 Modern Poetry	MA	10	9	5	5
Percentage of CLOs covered		19	17 (100%)	7(41%)	9 (59%)

Table 5 provides a domain-wise comparison of assessment tasks' coverage of the course-learning outcomes. The skill courses tasks covered 2 out of 15 learning outcomes, the linguistics courses covered 4 out of 14 learning outcomes, and the literature tasks covered 6 out of 14 learning outcomes. Two out of three courses from both skill courses and literature courses did not cover any of the formulated learning outcomes whereas two of the linguistics courses covered only one learning outcome. An interesting aspect of the formulated course learning outcomes is the number of them per course. That is, for linguistics and skill courses they range from 4 to 6. However, the inconsistency is obvious when it comes to the literature courses. The syllabus of modern poetry was based on 10 learning outcomes whereas modern English drama and nineteenth century novel both had only two learning outcomes which is suggestive of inconsistency in approach towards syllabus designing on departmental level.

Table 5. Domain-wise comparison of assessment tasks coverage of course-learning outcomes among instructors

Course Name	Curricular Domain	Total No. of CLOs	No. of CLOs covered	Number of CLOs not covered
1 Situational English		6	2	4
2 IELTS	Skills	4	0	4
3 Paragraph Writing		5	0	5
Percentage of CLOs covered		15 (100%)	2(13%)	13(87%)
4 Phonetics		5	1	4
5 Semantics	Linguistics	4	1	3
6 Morphology		5	2	3
Percentage of CLOs covered		14(100%)	4(29%)	10(71%)
7 Modern English Drama		2	0	2
8 Nineteenth Century Novel	Literature	2	0	2
9 Modern Poetry		10	6	4
Percentage of CLOs covered		14(100%)	6 (43%)	8(57%)

Chart 2 below includes a summary in percentage of domain wise coverage of the nine courses' learning outcomes. The domain wise analysis of the assessment tasks was to know the extent course learning outcomes were covered by the assessment tasks of each domain: skills, literature and linguistics. The chart indicates that

the learning outcomes of skill courses, linguistics and literature were covered 13%, 29% and 39% respectively. Though there is a visible difference among the percentages. The literature exams look slightly more valid than the other domains. In a nut shell, 60% of the learning outcomes of each domain were not assessed that indicates serious validity issues across the curriculum.

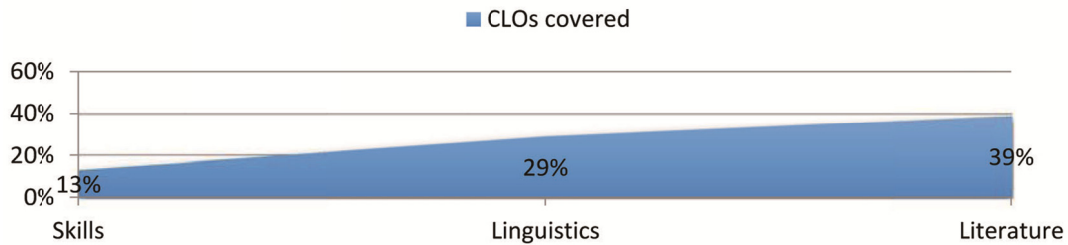


Figure 2. Parentage of domain-wise coverage of learning outcomes

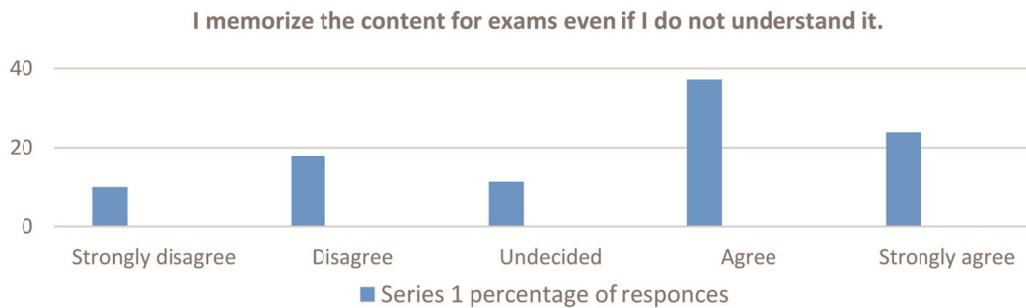


Figure 3. Learning strategy used by the learners

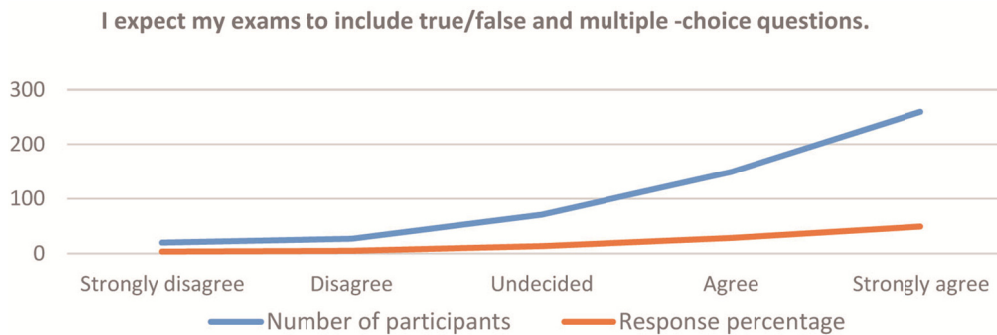


Figure 4. The kind of assessment tasks the learners expected

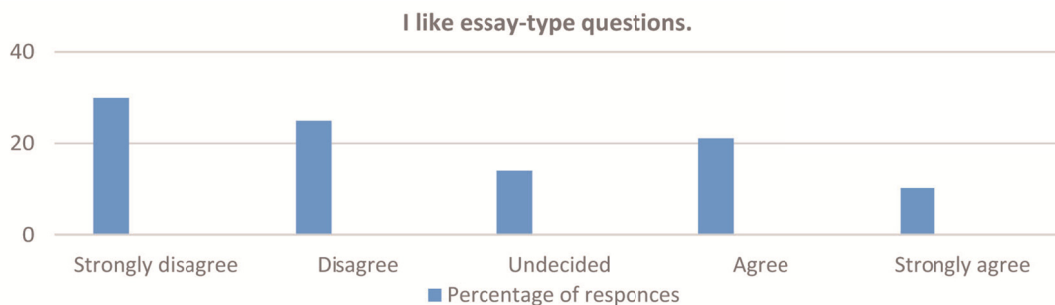


Figure 5. The learners' preference for constructed-response questions

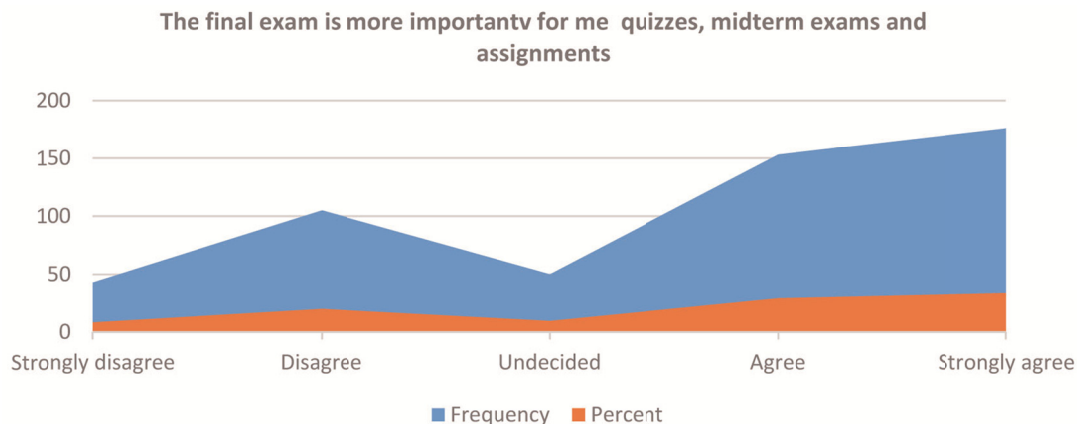


Figure 6. Importance given to the final exams compared to its formative counterpart

Having the assumption in mind that the students gave more importance to final-terms exams compared to the formative bits of assessment, the participants were invited in the survey to express their perceptions of the degree of importance they attached to both the summative and formative elements of assessment. In response, about two-third of the participants agreed that they considered the summative element more important than its formative component. In addition the results indicated that students preferred and expected easy assessment tasks (questions) in their quizzes and mid-term examinations in order to get good grades. They wanted true/false and multiple choice questions to be given in the tasks. They believed that such questions are easy to answer and allow them to get good grades. In addition, a majority of the respondents disagreed with the idea of including essay-type questions because they feared that they would fail such tasks due to their weak spellings and grammar, and lack of content they knew. One of the respondents in the open-ended question of the survey said “I am serious. The questions should be of true-false type to help us get good grades. Most of the teachers don’t do that. They give us essay-type questions in exams. Students forget everything when they see essay questions”. Another respondent explained it further why they liked and expected multiple choice questions: “Multiple choice questions are more helpful for students in getting higher grades compared to essay type questions. We make mistakes in essay type questions and thus we obtain low grades. All teachers should include SRQs in examinations.”

## 5. Discussion

The aim if this article was to seek answers for the research questions that included finding out the extent Saudi university EFL teachers’ assessment tasks covered student learning outcomes and the kind of impact teachers assessment practices had on students’ learning. In addition, the study was expected to show the extent the teachers assessment practices were in line with recommended assessment principles. In this section, each of the three research questions is discussed in light of the findings presented.

The first question targeted the teachers’ assessment tasks’ coverage of student learning outcomes. The results presented indicate serious validity concerns. It can be seen that the examination questions of skill courses, linguistics and literature covered 13%, 29% and 39% of the concerned learning outcomes respectively. Shockingly, it was noticed that four of the examinations—IELTS, Paragraph Writing, Modern English Drama and Nineteenth Century Novel—did not cover any of the given learning outcomes. In addition, the results revealed that most of the teachers’ questions assessed learners’ memorization or remembrance of knowledge. For instance, out of the total marks of the nine courses (=540), one fourth of the marks, i.e., 75% (=405) were allotted to the SRQs of which majority were true and false items. The survey (statistical) data and the participants’ responses to the open-ended question provide clear explanation in this regard and answer for the second research question. Two third of the survey respondents expected SRQs. It seems that the teachers unconsciously, if not intentionally, responded to learners’ expectations. These results confirm the finding of Umer (2016) that Saudi learners preferred multiple choice questions. Research elsewhere has also shown that teachers quite often tailor their examination questions in accordance with learners’ expectations (Saville & Hawkey, 2004; Cheng, 1997) which causes impact on learning. Therefore, Stecher et al. (2004) have explicitly warned that unless multiple choice questions are replaced by performance-based assessment, learners learning cannot be improved.



The results also indicate that important cognitive learning outcomes are not assessed because of assessment tasks not including CRQs. The assessment tasks are constructed to test students' memory or knowledge that causes surface level learning, producing graduates with no skills who cannot satisfy job market requirements. This goes against the country's 2030 vision which is asking for improved human capital. This finding lends support to the hypothesis that assessment will influence how students learn and the depth of their learning (Alderson & Wall, 1993). Previous research shows that changing test format can bring positive changes in teaching and learning strategies (Saif, 2006; Ferman, 2004). Therefore, to produce learners with higher order learning and life skills, the teachers should revisit their approach to assessment tasks as SRQs mostly result in lower order learning (see for example, El-Ebyary, 2009; Gijbels, Segers, & Struyf, 2008; Gijbels & Dochy, 2006; Scouller, 1998). Paul (2008) has recommended that assessment should incorporate tasks that replicate the language use outside-classroom environment, i.e., if students would be able to make use of the target skills effectively when asked for in real life. Therefore, there is a dire need for using valid and authentic questions in the context of this study and other similar settings that reflect real-life-like skills and application for causing positive washback (Green, 2006; Messick, 1996; Archbald & Newmann, 1988; Benedetti, 2006).

Another worth noticing point is the number of questions across questions varying from 2 to 8. In addition, the skill and linguistics courses only included SRQs. Showing inconsistency across courses or lack of centralized criteria provided by the department. How can 2 questions cover the content of a whole course? Or how can SRQs assess cognitive skills such as explanation and evaluation which are an integral part of each course specifications. In technical terms, this situation brings the assessment tasks' validity into question. How can valid and reliable inferences be drawn from such apparently invalid tools (Messick, 1993; Coombe & Evans, 2005)? Rather, such assessment concedes false inferences, students holding certificates with grades but have little or no relevant knowledge and skills (Green, 2007). Thus, for desirable influence on learners' achievements there has to be a strong overlap between task construct and course objectives (Green, 2007). This shows that blind reliance on teachers' assessment task designing skills (no matter how renown or expert they are in their fields or subjects) should stop. University teachers like school teachers are also in need of adequate assessment literacy. Being a great researcher or university professor does not necessarily guarantee the knowledge and application of sound assessment principles.

The third research question sought to evaluate if Saudi university EFL teachers assessment practices were in alignment with internationally recommended sound assessment principles. Based on the findings of this small-scale study, it can be safely said that by and large university teachers' assessment practices in the context of this research were far from satisfactory. The assessment tasks reviewed were invalid. Most of the assessment tasks assessed lower order learning. No agreement was noticed on any rubrics for the constructed response questions. Even teachers' approach to grading students' answers (the findings are not a part of this article) indicated tempering and inflation. Further research should look into teachers' assessment perceptions and beliefs that mostly have strong effect on their practices (Cheng, 2002). Their perceptions also affect their goals, values, beliefs in relation to the content and the process of teaching. However, the most crucial factor in determining teachers' approach towards teaching is the teachers' awareness of the formats, skills and contents to be tested in an examination (Alderson & Wall, 1993). The more is the awareness, the greater is the impact. Thus, it transpires that teachers' perceptions of an educational environment have to be in line with the aims of the curriculum at hand.

## 6. Conclusion

The findings of this study confirm the gap between assessment theory (recommended sound assessment principles) and teachers' actual assessment practices. For assessment tools to be truly effective, they have to measure the learning outcomes the tools (questions) are meant for. In addition, good assessment is supposed to influence students learning positively. However, results from this study indicate minimal desirable overlap between CLOs and examination questions culminating in students' surface-level learning. Therefore, to help students use deeper learning strategies changes in teacher assessment practices are essential. For this purpose, greater investment from government, universities and departments is required to improve teacher assessment literacy and its judicious application. In order to come up with an immediate solution, every department should develop their own quality assurance mechanism. Question papers and marked students' manuscripts should go through moderation process for validity and reliability check. It does not seem to suffice to rely on university teachers' high level of academic qualifications. Assessment is a standalone field (to master), yet inseparable from teaching and learning processes.

## Acknowledgments

This research was funded by the Research Department of Taif University (Research No.5832/438/1).

## References

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129. <https://doi.org/10.1093/applin/14.2.115>
- Arce-Ferrer, A. J., Cab, V. P., & Cisneros-Cohernour, E. J. (2001). *Teachers' assessment competencies*. Paper presented at the meeting of the American Educational Research Association, Seattle, OH.
- Archbald, D. A., & Newmann, F. M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Madison, WI: National Center on Effective Secondary Schools. ERIC database. (ED 301587). <http://www.eric.ed.gov/>
- Benedetti, K. D. (2006). Language testing: Some problems and solutions. *Mextesol*, 30(1). Retrieved from <http://mextesol.net/journal/public/files/2bbd95c84cfda1581bedde28ef1333ae.pdf>
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., ... Wyatt-Smith, C. (2015). International trends in the implementation of assessment: implications for policy and practice. *Policy Futures in Education*, 13(1), 117-140. <https://doi.org/10.1177/1478210314566733>
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-75. <http://dx.doi.org/10.1080/0969595980050102>
- Black, P., & William, D. (1998). Inside the black box. *Phi Delta Kappan*, 80(2), 139-148. Retrieved from <https://www.rdc.udel.edu/wp-content/uploads/2015/04/InsideBlackBox.pdf>
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11, 38-54. <https://doi.org/10.1080/09500789708666717>
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perception toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing; Research context and methods* (pp. 147-170). London: Lawrence Erlbaum Associates.
- Cheng, L., & Curtis, A. (2004). Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. In L. Cheng, Y. Watanabe, & A. Curtis, (Eds.), *Washback in language testing; Research context and methods* (pp. 3-17). London: Lawrence Erlbaum Associates.
- Coombe, C., & Evans, J. (2005). Writing assessment scales: Making the right choice. In D. Lloyd, P. Davidson, & C. Coombe (Eds.), *Fundamentals of Language Assessment: A Practical Guide for Teachers* (pp. 99-104). Dubai: TESOL Arabia Publications.
- DeLuca, C. (2012). Preparing teachers for the age of accountability: toward a framework for assessment education. *Teacher Education Yearbook XXI: A Special Issue of Action in Teacher Education*, 34(5/6), 576-591. <https://doi.org/10.1080/01626620.2012.730347>
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates learning. *Assessment in Education: Principles, Policy & Practice*, 17, 419-438. <http://dx.doi.org/10.1080/0969594X.2010.516643>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272. Retrieved from <https://eric.ed.gov/?id=EJ1112727>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy. *Educational Assessment*, 21(4), 246-268. Retrieved from <https://eric.ed.gov/?id=EJ1118826>
- El-Ebyary, K. (2009). Deconstructing the complexity of washback relation to formative assessment in Egypt. *Cambridge ESOL: Research Notes*, 35, 2-5. Retrieved from [www.cambridgeenglish.org/images/23154-research-notes-35.pdf](http://www.cambridgeenglish.org/images/23154-research-notes-35.pdf)
- Ezza, E. S. Y. (2017). Criteria for assessing writing at Majma'ah University. In H. Sahbi & C. Christine (Eds.), *Evaluation in Foreign Language Education in the Middle East and North Africa, Second Language Learning and Teaching*. Switzerland: Springer. [https://doi.org/10.1007/978-3-319-43234-2\\_11](https://doi.org/10.1007/978-3-319-43234-2_11)
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language Testing; Research context and methods* (pp.

- 191-210). London: Lawrence Erlbaum Associates.
- Galluzzo, G. R. (2005). Performance assessment and renewing teacher education. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(4), 142-145. <https://doi.org/10.3200/TCHS.78.4.142-145>
- Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: Can formative assessment make a difference? *Educational Studies*, 32(4), 399-409. <https://doi.org/10.1080/03055690600850354>
- Gijbels, D., Segers, M., & Struyf, E. (2008). Constructivist learning environments and the (im)possibility to change students' perceptions of assessment demands and approaches to learning. *Instr. Sci.*, 36, 431-443. Retrieved from <https://link.springer.com/article/10.1007/s11251-008-9064-7>
- Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly*, 3(4), 333-367. <https://doi.org/10.1080/15434300701333152>
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97. <https://doi.org/10.1080/09695940701272880>
- Hakim, B. (2015) English Language Teachers' Ideology of ELT Assessment Literacy. *International Journal of Education and Literacy Studies*, 3(4), 42-48. <http://dx.doi.org/10.7575/aiac.ijels.v.3n.4p.42>
- Koloi-Keaikitse, S. (2012). *Classroom assessment practices: A survey of Botswana primary and secondary school teachers* (Unpublished doctoral dissertation, Ball State University, Indiana).
- Lukin, L. E., Bandalos, D. L., & Eckhout, T. J. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23(2), 26-32. <https://doi.org/10.1111/j.1745-3992.2004.tb00156.x>
- Manjarrés, N. B. (2005). Washback of the foreign language test of the state examinations in Colombia: A case study. *Arizona Working Papers in SLAT*, 12, 1-19. Retrieved from <https://journals.uair.arizona.edu/index.php/AZSLAT/article/view/21283>
- Messick, S. (1993). Foundations of validity: meaning and consequences in psychological assessment. *ETS Research Report*, 2, i-18. <https://doi.org/10.1002/j.2333-8504.1993.tb01562.x>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256. <https://doi.org/10.1177/026553229601300302>
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49. <https://doi.org/10.1177/0265532209347148>
- National Commission for Academic Accreditation and Assessment. (2017). Handbook for quality assurance and accreditation in Saudi Arabia. Retrieved from <http://www.ncaa.org.sa/en/Releases/HandBookDocuments/Handbook%20Part%201.pdf>
- Niveen, R. M., Elshawa, A. N. A., & Sabariah, M. R. (2017). Malaysian Instructors' Assessment Beliefs in Tertiary ESL Classrooms *International Journal of Education & Literacy Studies*, 5(2), 29-46.
- Ogan-Bekiroglu, F. (2009). Assessing assessment: Examination of pre-service physics teachers' attitudes towards assessment and factors affecting their attitudes. *International Journal of Science Education*, 31, 1-29. <http://dx.doi.org/10.1080/09500690701630448>
- Paul, P. V. (Ed.) (2008). *Language and Deafness* (4th ed.). London: Jones and Barlett.
- Plack, B. (2015). *Assessment in the Service of Learning*. Presented in the Second International Conference for assessment and evaluation: Learning outcomes assessment. Retrieved from <http://ica.qiyas.sa/Presentations/ResearchSummary/En.pdf>
- Plake, B. S., & Impara, J. C. (1992). *Teacher competencies questionnaire description*. Lincoln, NE: University of Nebraska.
- Plake, B., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: a national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-39. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>

- Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*, 62, 82-83. Retrieved from <http://www.ascd.org/publications/educational-leadership/sept04/vol62/num01/Why-Assessment-Illiteracy-Is-Professional-Suicide.aspx>
- Popham, W. J. (2009). Assessment illiteracy for teachers: faddish or fundamental? *Theory into Practice*, 48, 4-11. <https://doi.org/10.1080/00405840802577536>
- Popham, W. J. (2013). *Classroom assessment: What teachers need to know* (7th ed.). Boston: Pearson.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1-34. <https://doi.org/10.1191/0265532206lt322oa>
- Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language Testing: Research context and methods* (pp. 73-96). London: Lawrence Erlbaum Associates.
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453-472. <https://doi.org/10.1023/A:1003196224280>
- Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington State. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language Testing: Research context and methods* (pp. 53-71). London: Lawrence Erlbaum Associates.
- Stiggins, R. (2002). Assessment crisis: the absence of assessment for learning. *Phi Delta Kappan*, 83, 758-765. <https://doi.org/10.1177/003172170208301010>
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86, 22-27. <https://doi.org/10.1177/003172170408600106>
- Susuwele-Banda, W. J. (2005). *Classroom assessment in Malawi: Teachers' perceptions and practices in mathematics* (Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Virginia).
- Umer, M. (2015). *Formative assessment and consequential validity: A practice yet to be truly implemented in Saudi higher education*. Presented in the Second International Conference for assessment and evaluation: Learning outcomes assessment. Riyadh: Saudi Arabia.
- Umer, M. (2016). *Formative assessment has to be made formative: It does not work itself*. Presented in National Conference on Education (IER 2016), Education for Tomorrow: Issues, Challenges and Prospects. Peshawar: Pakistan.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770. <https://doi.org/10.2307/20466661>
- Volante, L., & Melahn, C. (2005). Promoting assessment literacy in teachers: Lessons from the Hawaii school assessment liaison program. *Pacific Educational Journal*, 13, 19-34.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).