

# **The Effectiveness of an Online Training Programme in Developing Secondary School EFL Teachers' Alternative**

Assessment Literacy

**Mahmoud Abdelaziz Farag**

English Language Instructor &

Cambridge Certified Teacher Trainer

**Dr. Mohamed Farrag Badawi**

Associate Professor of Curriculum &  
(Instruction (EFL

Faculty of Education - October 6 University

**Dr. Ahmed Hassan Seifeddin**

(Professor of Curriculum & Instruction (EFL  
- Faculty of Education

Menoufia University

**Dr. Eman Mohamed Eldeeb**

(Lecturer of Curriculum & Instruction (EFL  
Faculty of Education - Helwan University

## ملخص البحث

لا شك أن المعلمين هم أهم ركيزه في العملية التعليمية. كما أن ممارساتهم في عملية التقويم تأثر بشكل ملحوظ على تقدم طلابهم، ولكن مازالت الدراسات في هذا المجال نادرة. لذا تهدف هذه الدراسة الى تطوير معلوماتهم ومهاراتهم معلمى اللغة الانجليزية بالمرحلة الثانوية في التقويم البديل من خلال برنامج تدريبي مقترح عبر الانترنت. وتبنى الدراسة المنهج الشبه تجريبي حيث خاضت مجموعة من المعلمين (43) اختبار قبلي بعدى. وقد أظهرت النتائج تواضع مستوى معلومات ومهارات معلمى اللغة الانجليزية بالمرحلة الثانوية في التقويم البديل. وقد تطورت معلوماتهم ومهاراتهم بشكل كبير حيث أن متوسط نتائجهم في الاختبار القبلي كان 11.86 من 30 (39.5%) وقد ارتفع الى 23.09 (76.9%) في الاختبار البعدى. وبناء عليه توصى هذه الدراسة بدمج مثل هذه البرامج التدريبيه فى كل من برامج إعداد المعلم وبرامج تدريب المعلمين لرفع مستوى أدائهم فى تقويم طلابهم.

الكلمات المفتاحية: التقويم البديل - مهارات التقويم البديل - تدريب المعلمين - معلمى اللغة الإنجليزية بالمرحلة الثانوية

## Abstract

Teachers are undoubtedly the most influential agent in the educational process. Their assessment practices heavily impact their students' progress. However, previous studies on teacher alternative assessment literacy are scarce. Therefore, the present study aims to develop secondary school EFL teachers' alternative assessment literacy through an online training programme. The study employs a quasi-experimental one group pre-posttest design. Forty-three teachers participated in the research project. Results indicate that teachers' level of alternative assessment literacy is low. In addition, findings reveal statistically significant treatment effects. Teachers' scores on the pre-posttest have improved significantly from 11.86 (39.5%) to 23.09 (76.9%) out of possible 30 on average. Therefore, such a training programme needs to be incorporated in both pre-service teacher preparation courses and in-service teacher training programmes to help teachers improve their assessment practice.

**Key Words:** alternative assessment, alternative assessment literacy, teacher training, secondary school EFL teachers



## Introduction

Over the past few decades, second and foreign language instruction focused on certain elements of the language such as grammar, vocabulary, pronunciation and punctuation. Consequently, traditional forms of language testing such as selected-response (i.e. true-false and multiple-choice) and fill-in-the-blank tests were appropriate to measure students' abilities to identify the correct answers rather than producing them. Recently, however, the focus has shifted towards communicative and task-based teaching approaches. Therefore, assessment needed to be modified accordingly to meet the new needs. This in turn has paved the way to the emergence of assessment for learning (AfL), as opposed to assessment of learning (AoL). AfL has gained momentum and has come to the fore. Thus, it has caused significant changes in assessment, shifting from standardised testing that focused on measuring memorisation and rote learning, towards more complex and comprehensive assessments of knowledge and higher-order thinking skills.

Alternative assessment is de facto at the heart of AfL as it aims to integrate what happens in the classroom with real-life contexts by replicating and demonstrating purposeful and meaningful manipulation of knowledge and skills of the target language (TL) inside and beyond the classroom. Therefore, when teachers design assessment strategies, they should make sure that their assessment practices would bring about a positive impact on the whole educational process. As a consequence, these significant changes have presented new challenges and implications for teachers and their assessment practices. Such practices have to be in agreement with the most recent conceptions of language assessment. Consequently, teachers' level of alternative assessment literacy has to be systematically

investigated and promoted in order for them to cope with the new paradigms of language assessment.

### **Context of the Problem**

Being a secondary school senior EFL teacher, the researcher has observed that many of his colleagues lack alternative assessment literacy. When the researcher got the feeling that this could be a serious problem, he did the following:

#### **First: Semi-Structured Interviews**

Ten semi-structured interviews were conducted with ten fellow teachers in order for the researcher to gather some preliminary data about their alternative assessment literacy. Findings revealed that eight of the ten teachers had no or very limited knowledge and skills of alternative assessment. However, the other two teachers were just average.

#### **Second: Reviewing Previous Studies**

Upon reviewing some of the key previous related studies in the Egyptian context and beyond, it was found out that teachers lacking assessment literacy is a global issue. The following are some of the key studies undertaken in a number of different countries (Malone, 2013; Popham, 2009) in the US, (Fulcher, 2012) in the UK, (Vogt & Tsagari, 2014) in Europe, (Scarino, 2013) in Australia, (Bozalek et al., 2013) in South Africa, (Xu & Brown, 2017) in China, (Yamtim & Wongwanich, 2014) in Thailand, (Alkharusi, Kazem & Al-Musawai, 2011) in Oman and finally (Abdel Latif, 2018; Gebiril, 2017; Hargreaves, 2001) in Egypt.

### **Statement of the Problem**

The problem of the present study is that secondary school EFL teachers' alternative assessment literacy is limited. Consequently, their assessment practices seem to hinder their students'

achievement. Therefore, the present study attempts to address this problem by suggesting a customised online training programme to promote their alternative assessment literacy.

### **Research Question**

The present study poses the following question:

- How effective is the suggested online training programme in developing secondary school EFL teachers' alternative assessment literacy?

### **Hypothesis of the Study**

This study hypothesises that there is a statistically significant difference at ( $\alpha \leq 0.01$ ) between the mean scores of the study group on the pre and post-administrations of the Alternative Assessment Literacy test, in favour of the post-administration.

### **Variables of the Study**

The present study has the following two variables:

1. The independent variable: An online training programme titled 'Alternative Assessment for Better Learning'
2. The dependent variable: Secondary school EFL teachers' alternative assessment literacy

### **Aim of the Study**

This study aims to investigate the effectiveness of the suggested training programme in developing alternative assessment literacy amongst secondary school EFL teachers.

### **Significance of the Study**

The significance of the present study is derived from the fact that it attempts to develop secondary school EFL teachers' alternative assessment literacy. In addition, its findings are expected to contribute to:

1. Highlighting the importance of promoting the level of alternative assessment literacy amongst secondary school EFL teachers.
2. Drawing the attention of the concerned authorities and officials to provide similar online training opportunities for pre and in-service teachers.
3. Laying the groundwork for future research on measuring and improving alternative assessment literacy among teachers of different subject matters and at other educational phases.

### **Delimitations of the Study**

The current study is delimited to the following:

1. A group of secondary school EFL teachers selected randomly from a number of governorates
2. September and October of the first semester in the academic year 2020-2021
3. Developing teacher assessment literacy within the domains of assessment, foundations, assessment use and assessment quality

### **Review of Literature**

#### **Alternative Assessment**

The large body of research on language assessment has provided mounting evidence that traditional methods of teaching and testing negatively impact student learning (Black & Wiliam, 1998; McMillan, 2004; Shepard, 2000; Stiggins, 2002). Besides, the twenty-first-century skills such as skills for learning, creativity, cooperation, collaboration, critical thinking and problem solving have been gaining momentum recently (Binkley et al., 2012), which has resulted in a growing need for constructing diverse methods of assessment (Care, Griffin & McGaw, 2012). Taken together, this dissatisfaction with conventional testing methods has led to the introduction of alternative assessment

to help overcome the shortcomings of standardised traditional tests (Berry, 2008).

It is widely accepted that alternative assessment is an umbrella term to lump together any nontraditional or non-standardised assessment methods. That is, alternative assessment has been interchangeably identified as authentic assessment, informal assessment, productive assessment, portfolio-based assessment and performance-based assessment. Methods of alternative assessment include exhibitions, interviews, journals, teacher observation, oral presentations, portfolios and projects (McMillan, 2007).

### **Assessment Literacy**

Stiggins (1991, 1995) was the first researcher to introduce the term assessment literacy, in the field of educational measurement. The definition of assessment literacy, has evolved significantly from merely the ability to differentiate between sound and unsound assessment practices (Stiggins, 1991, 1995) to the-knowledge, skills and abilities required to design, develop, maintain or evaluate, large scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals” (Fulcher, 2012, p. 125). Despite the fact that -research into assessment literacy is in its infancy” (Fulcher, 2012, p. 4), it has fast become a fundamental professional prerequisite for carrying out sound assessment practice especially in this era of accountability and standards-based education (Popham, 2020). As a result, measuring and developing teacher assessment literacy have become a pressing need.

## Measuring Teacher Assessment Literacy

Due to the increasing accountability framework in education and the recent paradigm shift towards standards-based education, there has been a growing need for teachers to exhibit a higher level of assessment literacy. The premise underlying the research in this realm is that teachers' knowledge and skills in assessment highly affect their own instruction and consequently student learning and achievement. Therefore, it is of great significance to measure teachers' assessment literacy (DeLuca & Klinger, 2010; Earl, 2013; Joughin, 2009). To this end, researchers employed two main approaches. First, self-reported measures to allow teachers to self-evaluate their own assessment competencies. Second, objective measures to equitably identify teachers' strengths and weaknesses in assessment. The self-reported measures are beyond the scope of this study due to the fact that they are more prone to intentional or unintentional distortions by the informant. Abrams (2004, p. 21) warns that "self-reported data rest on the assumption that the participants respond truthfully and accurately; unfortunately, it is almost impossible to test this assumption". The following section synthesises research using objective measures.

### Objective Measures

An online search for available objective measures was conducted. Only eight measures were identified. They are listed in Table 1 below in alphabetical order of the name of the measure, year of publication, author(s), targeted population, type of items, number of items and number of options or distractors in each instrument.

In 1993, Plake and her colleagues developed the Teacher Assessment Literacy Questionnaire (TALQ). It turned out to be the most influential instrument in this area. The seven standards put forward in 1990 by the American Federation of Teachers (AFT), National Council on Measurement in Education (NCME) and National Education Association (NEA) were used as the guiding framework. Each standard was assessed by a set of five MCQs. The sample

consisted of 555 school teachers and 286 administrators. They reported that teachers answered correctly 23.2 items out of the 35 (66%). Their strongest performance was in the areas of administering, scoring and interpreting test results. However, their weakest skill was in communicating test results.

**Table 1:**

**Key Objective Measures of Teacher Assessment Literacy**

No	Measure Name	Year	(Author(s	Population	Type of Items	No. of Items	No. of Options
1	Assessment Literacy Inventory (ALI)	2004	Campbell & Mertler	preservice teachers	MCQs	35	4
2	Assessment Results Interpretation and Use Survey	2012	Jarr	Practitioners	MCQs	11	4
3	Assessment Survey	1994	Drury	inservice teachers	MCQs	35	4
4	Classroom Assessment Literacy Inventory (CALI)	2003	Mertler	pre vs inservice teachers	MCQs	35	4
5	Criterion-Referenced Assessment Questionnaire	2010	King	inservice teachers vs administrators	MCQs	24	4
6	Measurement knowledge scale	2012	Gotch	elementary teachers	MCQs	20	4
7	Measurement Literacy Questionnaire	1998	Daniel & King	elementary & secondary teachers	True / False	30	-
8	Teacher Assessment Literacy Questionnaire (TALQ)	1993	.Plake et al	inservice teachers	MCQs	35	4

The second most cited measure is Assessment Literacy Inventory (ALI). It was introduced as a refined version of the abovementioned

TALQ. Similarly, Campbell and Mertler (2004) developed an inventory of 35 MCQs, aligned to the standards of AFT, NCME and NEA (1990). To make their instrument more user friendly, they created a scenario for each standard, presenting a classroom situation and developed five related questions for each scenario.

Mertler and Campbell (2005) used (ALI) to compare the pre-service and in-service teachers' assessment literacy. Findings indicated that the in-service teachers correctly answered 22 while the pre-service teachers correctly answered 19 out of 35. The in-service teachers scored highest on Standard 3 (Administering, Scoring and Interpreting Results of Assessments) while the pre-service teachers scored highest on Standard 1 (Choosing Appropriate Assessment Methods). Both groups of teachers scored lowest on Standard 5 (Developing Valid Grading Procedures). In addition, the in-service teachers outcompeted their counterparts on five standards plus the total score. However, statistical significant differences were only found in Standard 3, the total score and Standard 4 respectively.

These findings are in parallel with the results of Mertler (2003) who compared the level of assessment literacy of pre- and in-service teachers using another instrument called Classroom Assessment Literacy Inventory (CALI). It was also adapted from the TALQ. Likewise, the instrument consisted of 35 MCQs, which again paralleled the AFT, NCME and NEA's (1990) standards. He employed 197 in-service teachers and 67 pre-service teachers in the state of Illinois. His results indicated that both groups of teachers lacked assessment knowledge and skills.

Gotch and French (2013) carried out a study using Gotch's (2012) test titled Measurement Knowledge Scale. It was confined to assess teachers' technical knowledge and practical practices with assessment results. The instrument consisted of 20 MCQs with 4

options each. Items of the instrument were developed based on the most common types of information found in the assessment reports in several American states and Canadian provinces. The survey was published online and 650 public elementary school teachers in the State of Washington participated.

Results showed that teachers' average rate was 70% and their correct answers ranged from 6 to 20 correct responses out of possible 20. Gotch and French (2013, p. 52) outlined that the teachers -demonstrated the greatest skill in correctly placing a score into a proficiency category when given a set of cut scores, understanding the concept of the median as a measure of central tendency, and interpreting percentile ranks". Correct answers in these areas were around 90%. Whereas, correct answers in items tackling reliability or analysing qualities of an assessment in order to make sound informed decisions were below 40%.

Jarr (2012) surveyed 220 practitioners in public and private school settings in Iowa, US. Jarr developed a tool called the Assessment Results Interpretation and Use Survey to assess practitioners' abilities to locate, interpret and use assessment results. The instrument was divided into three subscales. Results showed that practitioners did well on the three subscales (Locate Data, Interpret Data, and Use Data). They scored an average total score of 76%. Their highest score was on the first subscale with group percentages ranging from 82 to 92%, followed by the second sub-scale with a percentage of 81% and finally the third subscale with group percentages ranging from 73 to 75%.

King (2010) investigated assessment literacy level among educators in relation to criterion-referenced assessments to identify their needs for professional development. 280 teachers, counselors and administrators in the states of Alabama and Mississippi took part. He devised an instrument called the Criterion-Referenced Assessment

Questionnaire. The relevant section of the questionnaire comprised 24 MCQs which were divided into three subsections. The first subsection (9 items) covered the theoretical differences between criterion and norm-referenced tests. The second subsection (8 items) focused on the fundamental concepts of assessment validity and reliability. The third subsection (7 items) dealt with potential assessment misuses. His findings indicated that counselors scored the highest while teachers scored the lowest. However, no significant differences were found between the three groups. Participants correctly answered 47.36%, 58.68% and 66.68% of the items in the three sub-scores respectively.

### **Final Observation on Research on Objective Measures**

After reviewing the available objective measures of teacher assessment literacy, they turned out to be unsuitable for this study for a number of limitations. First, -the majority of measures continue to be based on early conceptions of assessment literacy” (DeLuca, LaPointe-McEwan & Luhanga, 2016, p. 251). Thus, contemporary assessment concepts such as assessment for learning, alternative assessment, student involvement in assessment, etc. have not been addressed appropriately.

Second, the most widely used measures are the TALQ (Plake et al., 1993), ALI (Campbell & Mertler, 2004) and CALI (Mertler, 2003). They all relied on the Standards for Teacher Competence in Educational Assessment of Students (AFT, NCME & NEA, 1990) as their blueprints. These standards, however, -have become a bit dated” because they do not consider teachers’ knowledge and skills to conduct formative assessment nor to assess students in the current standards-based educational settings (Brookhart, 2011, p. 3).

Third, the only test that touched upon alternative assessment is Drury’s (1994). However, it is also based on the outdated AFT, NCME

and NEA's (1990) standards. Fourth, other tests such as (Daniel & King, 1998; Gotch, 2012; Jarr, 2012; King, 2010) are rather limited in their scope.

### **Design of the Study**

The study deployed a quasi-experimental research design. One-group pretest-posttest design was used to measure the impact of the quasi-independent variable (the suggested online training programme) on the dependent variable (teachers' alternative assessment literacy) by comparing the means of their scores before and after the treatment.

### **Participants of the Study**

A group of secondary school EFL teachers (n = 43) voluntarily participated in the study. The group consisted of 20 female and 23 male teachers from nineteen Egyptian governorates. Teachers at public schools represented a much larger number of participants with 32 teachers while there were 11 AlAzhar teachers. As per school location, 14 participants worked at schools in rural areas, 8 in the outskirts, and the majority (21) in urban locations. In terms of years of experience, 8 participants had less than ten years of experience, the largest number of participants (26) had ten to twenty years, and 9 had more than twenty years of teaching experience. As for their qualifications, graduates of Bachelor of Education represented the largest number of participants with 20 teachers, 16 of Arts, 2 of Alsun, 2 of Languages and Translation, 2 had a master's degree and 1 had a PhD in teaching English as a foreign language (TEFL).

### **Instruments of the Study**

#### **1. The Pre-Post Alternative Assessment Literacy Test**

In order to objectively measure the secondary school EFL teachers' alternative assessment literacy before and after the training

programme, the Teacher Alternative Assessment Literacy Measure, hereafter referred to as TAALM, has been developed.

### **TAALM Development**

The framework of Classroom Assessment Standards: Practices for PK-12 Teachers (Klinger et al., 2015) was utilised as the blueprint during the development of TAALM. This framework has been selected as its standards are -based on latest assessment research findings" (Xu & Brown, 2016, p. 14). Consequently they -more accurately reflect teachers' assessment demands within accountability-driven educational systems" (DeLuca et al., 2018 p. 360). Therefore, -developing survey tests of teacher AL [assessment literacy] needs to build upon more recent professional standards (e.g., Klinger et al., 2015) and contemporary assessment policies" (Xu & Brown, 2017, p. 150).

In addition, the framework is rather generic and not bound to a specific context as it lists -contemporary assessment skills teachers would be universally expected to develop or hold" (Looney et al., 2018, p. 9). Thus, it can be easily adopted in the Egyptian context. Further, the framework focuses on concrete, practical and measurable competencies.

Some items of the TAALM are intended to measure certain aspects of alternative assessment; other items, however, are related to key general concepts of assessment. The TAALM consists of 30 MCQs with four-option each. The four options are used as they are the conventional format of MCQs (Kilgour & Tayyaba, 2016) and they seem to be -a standard in the testing industry" (Haladyna, Downing & Rodriguez, 2002, p. 317). Table 2 below shows the distribution of the TAALM items among the three assessment domains.

**Table 2:**  
**TAALM Items Distribution**

No	Domain	Number of Items	Items
1	Assessment Foundations	10	1- 10
2	Assessment Use	10	11- 20
3	Assessment Quality	10	21- 30

### TAALM Validation

Eight jury members voluntarily reviewed and validated the TAALM. As a result, (1) a few items needed to be rewritten to make them more directly aligned to the guiding framework; (2) the order of the items had to be rearranged to set the context of alternative assessment, activate teachers' schemata and consequently make the instrument more reader-friendly; (3) a number of distractors had to be changed as they were deemed correct to a great extent; (4) a couple of questions needed to be incorporated in order to examine a downplayed construct and finally (5) some questions had to be removed as they were considered redundant.

### TAALM Pilot

After reaching a consensus with the jury members on all the items, weight ratio of the examined concepts, skills and constructs, answer key and distractors, a random group of secondary school EFL teachers ( $n = 19$ ), other than the main study participants, was recruited to answer the piloted instrument. After the pilot, the participants confirmed that the items and options of the measure were straightforward and clear enough to comprehend and they did not feel any kind of confusion. They, however, advised some further grading of the language in order to make the TAALM clearer and easier to the intended participants. Time wise, they stated that the

measure needed around 20 to 25 minutes to complete.

### **TAALM Reliability**

To measure the internal consistency of TAALM, Kuder and Richardson 20 (K-R20) analysis was carried out. The result revealed a value of .8, which is good for such test (Streiner, Norman & Cairney, 2015). Therefore, this indicated that TAALM has yielded reliable raw data.

## **2. The Suggested Online Training Programme**

### **Training Programme Description**

The proposed training programme, titled 'Alternative Assessment for Better Learning', was a customised online course which was made available on <https://aa4elt.com> during September and October in 2020. The programme was tailored specifically for secondary school EFL teachers in the light of their results on TAALM. It was self-study and self-paced to allow teachers to advance through the course content at their own pace and convenience. It was expected to take teachers around twenty hours to complete. However, they varied in their participation, commitment and completion timeframe.

### **Training Programme Objectives**

The training programme aimed to assist secondary school EFL teachers to:

- Evaluate the different assessment methods and their uses.
- Select, adapt and design appropriate assessment methods that best serve different assessment purposes.
- Engage students meaningfully in the assessment process.
- Select, adapt and design effective scoring tools.
- Score student performance accurately and consistently.
- Analyse and interpret assessment results appropriately and make sound educational decisions.
- Provide useful feedback to different assessment users.

- Examine fundamental assessment qualities such as validity, reliability, practicality, fairness, ethics and positive impact throughout the assessment process.

### Training Programme Content

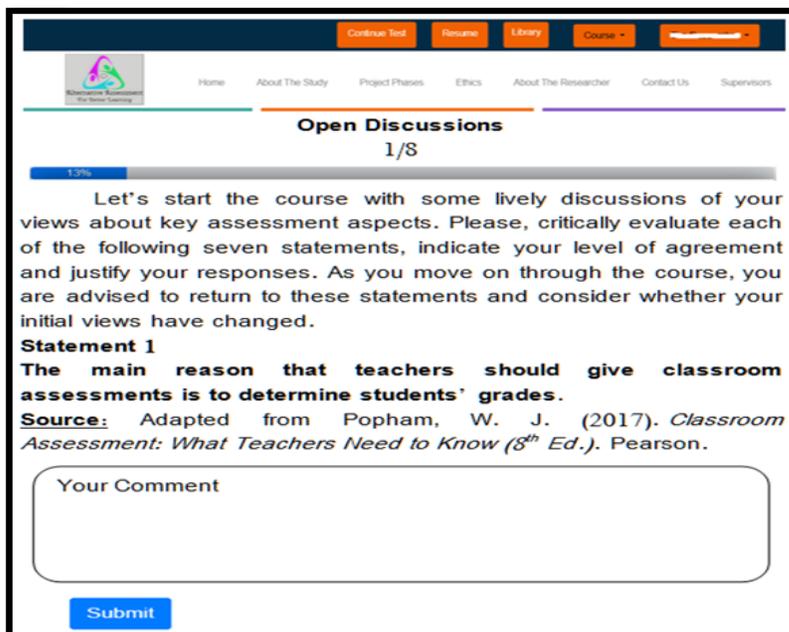
The training programme contents and materials were organised as shown in the Table 3 below:

**Table 3:**  
**Training Programme Table of Contents**

Module No.	Lesson No.	Lesson Title	
Module One	1	What is assessment?	
	2	Alternative Assessment Vs Traditional Testing	
	2.1	Alternative Assessment Vs Traditional Testing- Project 1	
	2.2	Alternative Assessment Vs Traditional Testing- Project 2	
	3	Formative Vs Summative Assessments	
	4	Criterion-Referenced Vs Norm-Referenced Assessments	
	5	Interpretation of Criterion & Norm-Referenced Scores- Scenario1	
	5.1	Criterion Vs Norm-Referenced Assessments- Project 3	
	6	Formal Vs Informal Assessments	
	7	Diagnostic Assessment	
		Module One Assignment	
	Module One Self-reflection		
Module Two	1	Assessment Purpose	
	2	Learning Expectations	
	2.1	Learning Expectations- Scenario 2	
	3	Assessment Design	
	3.1	Assessment Design- Scenario 3	
	3.2	Assessment Design- Scenario 4	
	4	Student Engagement in Assessment	
	5	Self and Peer-Assessment Quiz	
		Module Two Assignment	
		Module Two Self-reflection	

Module Three		Analysis of Student Performance- Warmer	
	1	Analysis of Student Performance	
	1.1	Analysis of Student Performance- Checklists	
	1.2	Analysis of Student Performance- Rating Scales	
	1.3	Analysis of Student Performance- Rubrics	
	1.4	Analysis of Student Performance Quiz	
	1.5	Consistent Scoring- Scenario 5	
	1.6	Analysis of Student Performance- Project 4	
	2	Effective Feedback	
	2.1	Effective Feedback Quiz	
	3	Instructional Follow-Up	
	4	Communicating Results and Instructional Decisions	
		Module Three Assignment	
		Module Three Self-reflection	
	Module Four		Assessment Quality- Warmer
1		Validity	
1.1		Validity Quiz	
2		Types of Validity	
2.1		Content Validity	
2.2		Construct Validity	
2.3		Criterion Validity	
2.4		Face Validity	
2.5		Most Important Type of Validity	
3		Reliability	
4		Types of Reliability	
5		Error of Measurement	
6		Practicality	
7		Fairness	
8		Ethics	
9		Washback/Impact	
10		Reflection	
		Module Four Assignment	
	Module Four Self-reflection		

Here are some screenshots of the training programme materials as they appeared on the website:



Continue Test Resume Library Course -

Home About The Study Project Phases Ethics About The Researcher Contact Us Supervisors

## Open Discussions

1/8

1.00%

Let's start the course with some lively discussions of your views about key assessment aspects. Please, critically evaluate each of the following seven statements, indicate your level of agreement and justify your responses. As you move on through the course, you are advised to return to these statements and consider whether your initial views have changed.

**Statement 1**

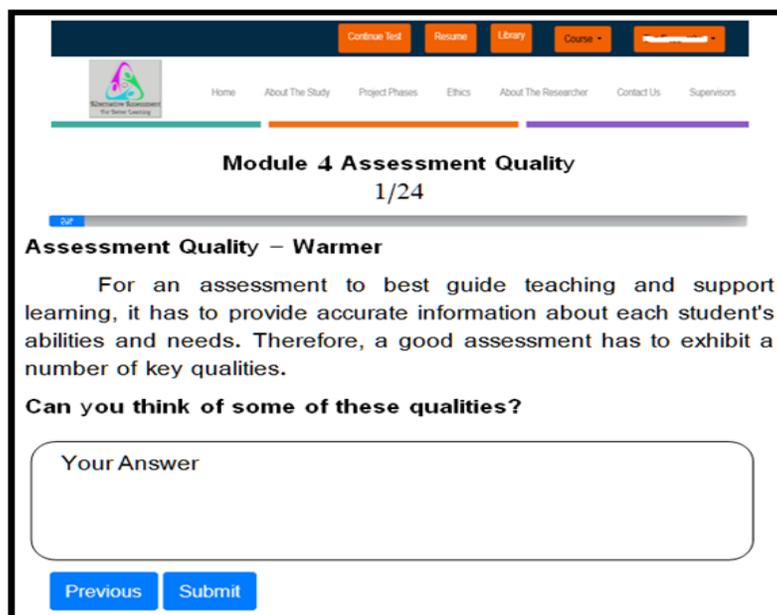
**The main reason that teachers should give classroom assessments is to determine students' grades.**

**Source:** Adapted from Popham, W. J. (2017). *Classroom Assessment: What Teachers Need to Know (8<sup>th</sup> Ed.)*. Pearson.

Your Comment

Submit

Figure 1. Screenshot of Sample Open Discussion



Continue Test Resume Library Course -

Home About The Study Project Phases Ethics About The Researcher Contact Us Supervisors

## Module 4 Assessment Quality

1/24

0.00%

**Assessment Quality – Warmer**

For an assessment to best guide teaching and support learning, it has to provide accurate information about each student's abilities and needs. Therefore, a good assessment has to exhibit a number of key qualities.

**Can you think of some of these qualities?**

Your Answer

Previous Submit

Figure 2. Screenshot of Sample Warmer



Figure 3. Screenshot of Sample Lesson

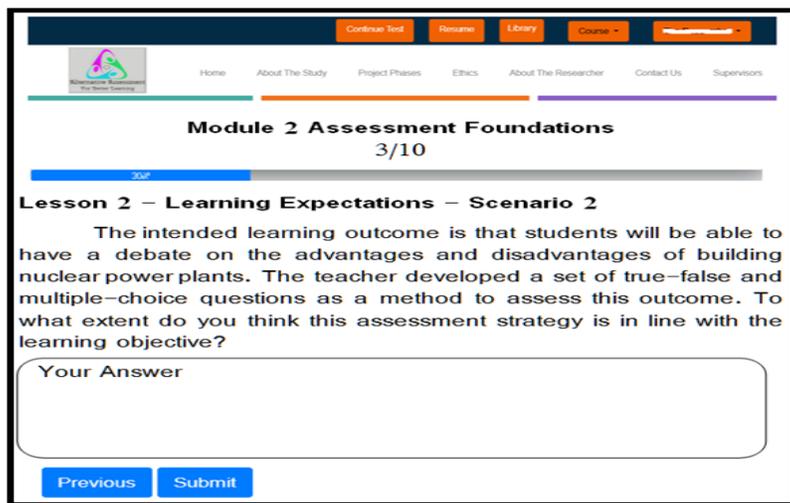


Figure 4. Screenshot of Sample Scenario

### Training Programme Assessment

During the training programme, teachers were evaluated

continuously through a range of alternative assessment methods, which were used formatively to help them promote and consolidate their learning. The assessment methods included open discussions, projects, quizzes, scenarios, assignments, presentations and self-reflection tasks. The screenshots below show some examples of the evaluation methods.

Figure 5. Screenshot of Sample Project

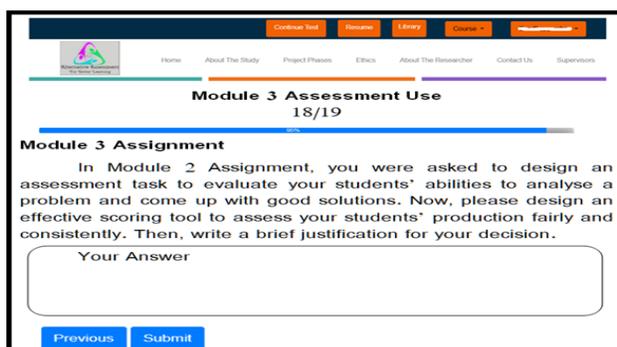


Figure 6. Screenshot of Sample Module Assignment

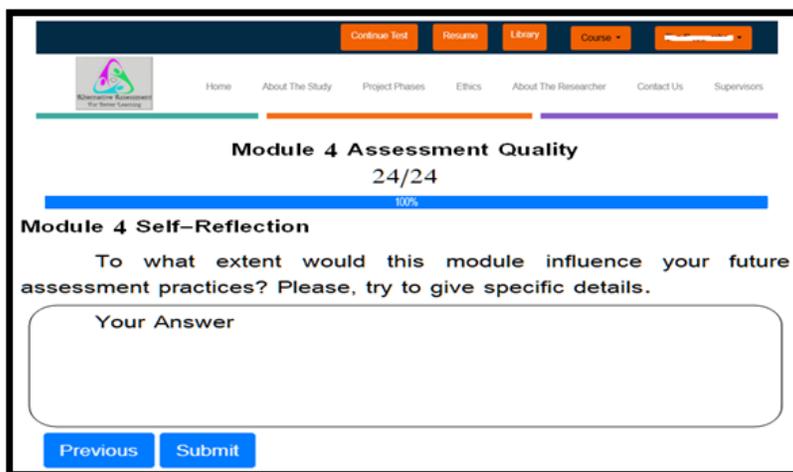


Figure 7. Screenshot of Sample Self-Reflection Assignment

### Results and Discussion

The study hypothesis postulates that there is a statistically significant difference at ( $\alpha \leq 0.01$ ) between the mean scores of the study group in the pre and post-administrations of TAALM, in favour of the post-administration. To examine this hypothesis, a paired-samples t-test analysis was carried out. As shown in Figure 8 and Table 4 below, results revealed that there was a statistically significant increase in TAALM scores from the pre-administration ( $M = 11.86$  &  $SD = 4.34$ ) to the post-administration ( $M = 23.09$  &  $SD = 4.65$ ),  $t(42) = 14.802$ ,  $p < .0001$  (two-tailed).

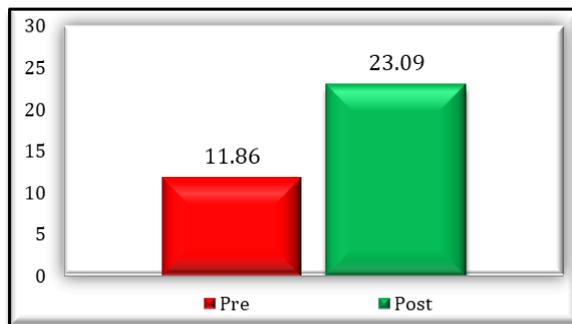


Figure 8. Teachers' Pre & Post Means on TAALM

In addition, Table 4 below illustrates that the mean increase in TAALM scores was 11.23 and a standard deviation of 4.98 with a 95% confidence interval ranging from 12.76 to 9.70. More importantly, in order to investigate the effect size of the increase in TAALM scores, the eta squared statistic analysis was performed.

Table 4: Paired-Samples T-Test Statistics on TAALM

Analogy	Mean	SD	Increase Mean	Increase SD	confidence interval of 95% the differences		t	df	Sig (2-tailed)
					Lower	Upper			
Pretest	11.86	4.34	11.23	4.98	Lower	Upper	14.802	42	.000.>
Posttest	23.09	4.65			12.76	9.70			

Pallant (2016, p 240) explains that the formula for calculating the eta squared value is as follows:

$$t^2$$

$$\text{Eta squared} =$$

$$t^2 + N - 1$$

Results revealed that eta squared statistic was .84, indicating a large effect size of the training programme (Cohen, 2013). Hence, it can be inferred that there was a large effect, with a substantial difference in TAALM scores achieved prior and post the experiment.

As per the first domain (Assessment Foundations), a paired-samples t-test analysis was performed in order to examine the impact of the treatment on the study sample's scores. Results revealed that there was a statistically significant increase in the scores of the first section of the TAALM, covering the first domain from the pre-administration (M = 5.12 & SD = 2.26) to the post-administration (M = 7.91 & SD = 1.92),  $t(42) = 7.150$ ,  $p < .0001$  (two-tailed). The mean increase in the first domain scores was 2.79 and a standard deviation of 2.56 with a 95% confidence interval ranging

from 3.58 to 2.00. To further examine the effect size of the posttest improvement, the eta squared statistic was calculated. The result showed that eta squared statistic was .55, producing a large effect size of the training programme.

In the same vein, to investigate effectiveness of the treatment in improving the scores of the study subjects in the second domain (Assessment Use), a paired-samples t-test analysis was done. Results indicated that there was a statistically significant increase in the scores of the TAALM section, covering the second domain from the pre-administration ( $M = 3.84$  &  $SD = 1.80$ ) to the post-administration ( $M = 7.80$  &  $SD = 2.03$ ),  $t(42) = 10.847$ ,  $p < .0001$  (two-tailed). The mean increase in the second domain scores was 3.95 and a standard deviation of 2.39 with a 95% confidence interval ranging from 4.69 to 3.22. In order to find out the effect size of the posttest improvement, the eta squared statistic was calculated. The result indicated that eta squared statistic was .74, producing a large effect size of the training programme.

Similarly, a paired-samples t-test analysis was completed to analyse the efficacy of the treatment on the scores of the study participants in the third domain (Assessment Quality). It was also found that there was a statistically significant increase in the scores of the third section of TAALM, covering the third domain from the pre-administration ( $M = 2.91$  &  $SD = 1.62$ ) to the post-administration ( $M = 7.40$  &  $SD = 2.19$ ),  $t(42) = 14.952$ ,  $p < .0001$  (two-tailed). The mean increase in the third domain (Assessment Quality) scores was 4.49 and a standard deviation of 1.97 with a 95% confidence interval ranging from 5.09 to 3.88. The eta squared statistic was calculated to examine the effect size of the posttest improvement. The result indicated that eta squared was .84, showing a large effect size.

## Discussion

Results of the current study indicated that there was a large improvement in the participants' alternative assessment literacy. This manifested itself in the significant progress in their performance on TAALM from the pre- (11.86) to the post-administration (23.09 out of 30). Such large improvements could be attributable to the proposed training programme.

During the pre-assessment, it was evident that the teachers lacked alternative assessment literacy, scoring 11.86 out of 30 (39.5%) on average. This score is significantly lower than other scores documented in a number of key studies conducted in different parts of the world. For instance, in the US, Plake et al. (1993) reported 66% for school teachers. Similarly, Mertler (2003) reported 63% for in-service teachers and 54% for pre-service teachers. Likewise, Mertler and Campbell (2005) reported 67% for student teachers. Moreover, King (2010) reported a result of 47.36% for teachers in the states of Alabama and Mississippi. In addition, Jarr (2012) reported an average total score of 76% for practitioners in public and private school settings in Iowa. Finally, Gotch and French (2013) reported 70% for public elementary school teachers in the State of Washington.

In completely different educational contexts, Zheng (2010) reported 46% for primary school teachers in China. In a similar manner, Alkharusi, Kazem and Al-Musawai (2011) reported 44% for pre-service teachers and 36% for primary and preparatory in-service teachers in Oman. In addition, Yamtim and Wongwanich (2014) reported 49% for primary teachers in Thailand.

However, in the post-evaluation of the current study, teachers' alternative assessment literacy improved significantly, scoring 23.09 out of 30 (76.9%) on average. This outcome seems to make the Egyptian secondary school EFL teachers slightly better than Jarr's

(2012) participants. It is noteworthy that Jarr (2012) documented the highest score in the literature of teacher assessment literacy as yet.

Furthermore, the average percentage reported in the present study (39.5%) in the pre-administration is way less than the level that Egyptian EFL teachers and student teachers self-reported on their own level of assessment literacy in Gebril's (2017) study. He reported that pre-service teachers rated their own assessment literacy just under 60% while the in-service group rated themselves around 63%. It is widely recognised that overestimation is a typical limitation of self-reported studies.

Another intriguing result from the pre-administration is that teachers' scores ranged from a low of 5 to a high of 23 out of 30. This highest mark was achieved by one teacher only. However, in the post-administration, 24 teachers (55.8%) scored as high as or higher than 23, with five teachers achieving the full mark. Moreover, given the fact that the final score on TAALM is 30, the passing score would be 18, which is 60%. Out of the 43 participants, only 4 teachers (0.09%) scored 18 or above the cutoff point in the pre-administration while 37 teachers (86%) were able to do the same thing in the post-administration.

Additionally, Figure 9 below indicates that the participants' weakest domain in the pre-assessment was Assessment Quality, on which their average score was 2.91 out of 10. Their second weakest domain was Assessment Use with an average score of 3.84 out of 10. Finally, their strongest assessment domain was Assessment Foundations with an average score of 5.12 out of 10. However, their weakest domain in the pre-assessment (Assessment Quality) became their second strongest one in the post-assessment.

Figure 9. Teachers' Pre & Post Means of TAALM's Domains

As per the first domain, teachers' knowledge and skills of assessment foundations significantly improved. That is, their average score developed from 5.12 in the pre-administration to 7.19 afterwards. In the pre-assessment, they exhibited a low level of literacy as only 37% of the informants correctly answered 60% of the questions in this domain, albeit this percentage increased significantly in the post-administration (91%), with 10 teachers securing the full mark. In the pre-administration, question 9 was found to be the most difficult item in the first domain, with only 12 correct answers (28%) while 35 teachers (81%) answered it correctly in the post-administration.

Regarding the second domain, teachers' competencies in assessment use developed significantly. Their scores progressed from 3.84 in the pre-administration to 7.80 in the post-administration. The highest score in the pre-administration was 8 out of 10, which was achieved by just one teacher, whereas 29 teachers (67%) scored 8 or more in the post-administration. The percentage of the teachers who correctly answered 60% of the questions in this domain raised from 18% in the pre-administration to 84% in the post-administration, with 10 teachers achieving the full mark. In the pre-administration, the most difficult question in this domain was question 14, with only 8 (19%) correct answers. However, in the post-administration, 35 teachers (81%) were capable of answering it right.

Finally, teachers' competence in the third domain (assessment quality) witnessed the largest improvement. Their scores increased from 2.91 in the pre-administration to 7.40 afterwards. In the pre-administration, only 2 teachers correctly answered 60% of the questions in this domain, albeit 30 teachers (70%) succeeded to do so in the post-administration, with 13 teachers (30%) achieving the full mark. In the pre-administration, two of the most difficult questions in the domain were items 25 and 27, with only 5 (12%) and 7 (16%)

correct answers respectively. However, in the post-administration, 26 teachers (60%) and 34 teachers (79%) got them right respectively.

In conclusion, these findings resonate with other studies, which documented the effectiveness of in-service assessment literacy training courses (e.g. Barcelos & Kalaja, 2011; Koh & Velayutham, 2009). Therefore, such results are testimony of the possibility of modifying teachers' conceptions, practices and attitudes through research-driven instruction (Borg, 2011, 2015; DeLuca, Chavez, Bellara & Cao, 2013).

### **Conclusion**

The present study has provided evidence that secondary school EFL teachers lack alternative assessment literacy. It manifests itself in their poor performance on TAALM, scoring an average of 11.86 out of possible 30 in the pre-administration. Besides, results of the post-administration augur well for the effectiveness of the suggested training programme. It has helped improve teachers' scores significantly, lifting them from 11.86 to 23.09 out of 30 on average. In addition, the study shows preliminary support for TAALM. Its items seem to function reasonably well. However, further review and refinement of the instrument are appreciated in order for TAALM to offer the most possible robust evidence of teacher assessment literacy.

### **Recommendations**

Based on the findings of the present study, the following recommendations can be made:

1. Providing sufficient training on alternative assessment needs to be an essential component in teacher preparation programmes.
2. Developing teachers' alternative assessment literacy should be an integral part of teacher professional development training programmes.

3. Promoting alternative assessment literacy among teachers of different subject matters and various educational phases ought to be prioritised. Thus, teachers would be more assessment literate and subsequently more able to execute assessment effectively and confidently.

### **Suggestions for Further Research**

On the basis of the outcomes of the present study, the following studies are recommended for future research:

1. Measuring and developing the level of alternative assessment literacy among teachers of different subject matters and other educational phases. The suggested training programme as well as TAALM can be adopted or adapted to serve such purposes.
2. Examining and promoting alternative assessment literacy amongst other relevant stakeholders such as administrators, curriculum developers and policymakers. That is because they are involved in making crucial educational decisions based on the available assessment data (Popham, 2020).
3. Evaluating and developing assessment literacy among students themselves. That is due to the fact that recent research suggests that students' awareness and understanding of the assessment process could promote their learning (Mercer & Dörnyei, 2020; Moss & Brookhart, 2019; Taras, 2013).

### **References**

- Abdel Latif, M. M. (2018). English language teaching research in Egypt: trends and challenges. *Journal of Multilingual and Multicultural Development*, 1-12. DOI:10.1080/01434632.2018.1445259
- Abrams, L. (2004). *Teachers' Views on High-Stakes Testing: Implications for the Classroom*. Boulder, CO: National Education Policy Center. Retrieved on 14th October 2019 from <https://scholar.colorado.edu/nepc/409/>

- Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, 39(2), 113-123. DOI: 10.1080/1359866X.2011.560649
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for Teacher Competence in Educational Assessment of Students*. Washington, D.C: National Council on Measurement in Education. Retrieved on 1st, September 2018 from <https://buros.org/standards-teacher-competence-educational-assessment-students>
- Barcelos, A. M. F., & Kalaja, P. (2011). Introduction to Beliefs about SLA Revisited. *System*, 39(3), 281-289.
- Berry, R. (2008). *Assessment for Learning*. Hong Kong: Hong Kong University Press.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 17-66). New York: Springer.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Borg, S. (2011). The impact of in-service teacher education on language teachers' beliefs. *System*, 39(3), 370-380. DOI:10.1016/j.system.2011.07.009
- Borg, S. (2015). *Teacher Cognition and Language Education: Research and Practice*. Bloomsbury Publishing.
- Bozalek, V., Gachago, D., Alexander, L., Watters, K., Wood, D., Ivala, E., & Herrington, J. (2013). The use of emerging technologies for

authentic learning: A South African study in higher education. British Journal of Educational Technology, 44(4), 629-638.

- Brookhart, S. (2011). Educational Assessment Knowledge and Skills for Teachers. Educational Measurement: Issues and Practice, 30(1), 3-12.
- Campbell, C., & Mertler, C. (2004). Assessment Literacy Inventory. Unpublished Instrument.
- Care, E., Griffin, P., & McGaw, B. (2012). Assessment and Teaching of 21st Century Skills. New York: Springer.
- Cohen, J. (2013). Statistical Power Analysis for the Behavioral Sciences (2nd Ed.). New York: Lawrence Erlbaum Associates.
- Daniel, L. G., & King, D. A. (1998). Knowledge and Use of Testing and Measurement Literacy of Elementary and Secondary Teachers. The Journal of Educational Research, 91(6), 331-344.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. Assessment in Education: Principles, Policy & Practice, 17(4), 419-438. DOI: 10.1080/0969594X.2010.516643
- DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for Preservice Assessment Education: Supporting Teacher Candidates' Assessment Literacy Development. The Teacher Educator, 48(2), 128-142. DOI:10.1080/08878730.2012.760024
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher Assessment Literacy: A Review of International Standards and Measures. Educational Assessment, Evaluation and Accountability, 28(3), 251-272. DOI.org/10.1007/s11092-015-9233-6
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D.,

- & Luhanga, U. (2018). Teachers' Approaches to Classroom Assessment: A Large-Scale Survey. *Assessment in Education: Principles, Policy & Practice*, 25(4), 355-375. DOI.org/10.1080/0969594X.2016.1244514
- Drury, J. H. (1994). A Survey to Investigate Teacher Awareness of Alternative Assessment of Students in Mathematics (Doctoral Thesis). The Ohio State University, USA.
  - Earl, L. M. (2013). *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. California: Corwin Press.
  - Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
  - Gebril, A. (2017). Language teachers' conceptions of assessment: an Egyptian perspective. *Teacher Development*, 21(1), 81-100. DOI: 10.1080/13664530.2016.1218364
  - Gotch, C. M. (2012). An investigation of Teacher Educational Measurement Literacy (Doctoral Thesis). Washington State University, USA.
  - Gotch, C. M., & French, B. F. (2013). Elementary teachers' knowledge and self-efficacy for measurement concepts. *The Teacher Educator*, 48(1), 46-57. DOI: 10.1080/08878730.2012.740150
  - Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-333.
  - Hargreaves, E. (2001). Assessment in Egypt, *Assessment in Education: Principles, Policy & Practice*, 8(2), 247-260.
  - Jarr, K. A. (2012). Education Practitioners' Interpretation and Use of Assessment Results (Doctoral Thesis). University of Iowa, Iowa City, USA.

- Joughin, G. (2009). Assessment, Learning and Judgement in Higher Education: A Critical Review. In G. Joughin (Ed.), Assessment, Learning and Judgement In Higher Education (pp. 1-15). New York: Springer.
- Kilgour, J. M., & Tayyaba, S. (2016). An investigation into the Optimal Number of Distractors in Single-Best Answer Exams. *Advances in Health Sciences Education*, 21(3), 571-585.
- King, J. D. (2010). Criterion-Referenced Assessment Literacy of Educators (Doctoral Thesis). University of Southern Mississippi, USA.
- Klinger, D.A., McDivitt, P.R., Howard, B.B., Munoz, M.A., Rogers, W.T., & Wylie, E.C. (2015). *The Classroom Assessment Standards for PreK-12 Teachers*. Kindle Direct Press. Retrieved on 3rd March 2019 from <https://jcsee.org/classroom/>
- Koh, K. H., & Velayutham, R. L. (2009). Improving Teachers' Assessment Literacy in Singapore Schools: An Analysis of Teachers' Assessment Tasks and Student Work. *Research Brief*, 9(2), 1-4.
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the Role of Teachers as Assessors: Teacher Assessment Identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442-467. DOI: 10.1080/0969594X.2016.1268090
- Malone, M. E. (2013). The Essentials of Assessment Literacy: Contrasts Between Testers and Users. *Language Testing*, 30(3), 329-344. DOI: 10.1177/0265532213480129
- McMillan, J. H. (2004). *Classroom Assessment. Principles and Practices for Effective Instruction* (3rd ed.). Boston, MA: Pearson.
- McMillan, J. H. (2007). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction* (4th ed.). Boston, MA: Allyn and Bacon.
- Mercer, S., & Dörnyei, Z. (2020). *Engaging Language Learners in*

Contemporary Classrooms. Cambridge University Press.

- Mertler, C. A. (2003, October). Preservice Versus Inservice Teachers' Assessment Literacy: Does Classroom Experience Make a Difference? [Paper presentation]. The Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Mertler, C. A., & Campbell, C. (2005). Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory". Online Submission.
- Moss, C. M., & Brookhart, S. M. (2019). Advancing Formative Assessment in Every Classroom: A Guide for Instructional Leaders. Alexandria, Virginia: ASCD.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment Competencies of Teachers: A National Survey. Educational Measurement: Issues and Practice, 12(4), 10-12.
- Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental?. Theory into Practice, 48(1), 4-11. DOI: 10.1080/00405840802577536
- Popham, W. J. (2020). Classroom Assessment: What Teachers Need to Know (9th ed.). Boston: Pearson.
- 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. DOI: 10.1177/0265532213480128
- Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. Educational Researcher, 29(7), 4-14.
- Stiggins, R. J. (1991). Assessment Literacy. Phi Delta Kappan, 72(7), 534-539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. Phi

Delta Kappan, 77(3), 238-245.

- Stiggins, R.J. (2002). Assessment Crisis: The Absence of Assessment for Learning. Phi Delta Kappan, 83(10), 758-765.
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). Health Measurement Scales: A Practical Guide to their Development and Use. Oxford: Oxford University Press.
- Taras, M. (2013). Feedback on Feedback: Uncrossing Wires across Sectors. In S. Merry, M. Price, D. Carless, & M. Taras (Eds.), Reconceptualising Feedback in Higher Education: Developing Dialogue with Students (pp. 30-40). New York, NY: Routledge.
- Vogt, K., & Tzagari, D. (2014). Assessment Literacy of Foreign Language Teachers: Findings of a European Study. Language Assessment Quarterly, 11(4), 374-402. DOI: 10.1080/15434303.2014.960046
- Xu, Y., & Brown, G. T. (2016). Teacher Assessment Literacy in Practice: A Reconceptualization. Teaching and Teacher Education, 58, 149-162. DOI:10.1016/j.tate.2016.05.010
- Xu, Y., & Brown, G. T. (2017). University English Teacher Assessment Literacy: A Survey-Test Report from China. Papers in Language Testing and Assessment, 6(1), 133-158.
- Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. Procedia-Social and Behavioral Sciences, 116, 2998-3004. DOI: 10.1016/j.sbspro.2014.01.696
- Zheng, D. (2010). An Investigation into the Assessment Literacy of Secondary and Primary School Teachers: A Report from Z Province. Global Education Review 39(2), 31-36.