Developing Items to Measure the Assessment Literacy of ESL Teachers

Harsharan Kaur Jaswan Singh
Universiti Pendidikan Sultan Idris

Charanjit Kaur Swaran Singh
Universiti Pendidikan Sultan Idris
Infrastructure University Kuala Lumpur (IUKL)
UCSI University
Asia e University (AeU)

Eng Tek Ong
UCSI University

Wong Wei Lun
Universiti Pendidikan Sultan Idris

Tarsame Singh Masa Singh
SEGi University

Revathi Gopal
Universiti Pendidikan Sultan Idris

Mahendran Maniam
Universiti Pendidikan Sultan Idris

Siti Shuhaida Shukor
Universiti Pendidikan Sultan Idris

Dodi Mulyadi
Universitas Muhammadiyah Semarang

This study aims to develop a reliable instrument measuring the assessment literacy of ESL teachers through Exploratory Factor Analysis based on five constructs. A survey was administered to 200 secondary school ESL teachers. The study revealed that only 78 items with a cut-off point of above .60 from the factor loading are retained. Cronbach's Alpha showed that the items retained are reliable. This study developed and validated an instrument to measure ESL teachers' assessment literacy in the Malaysian context. An assessment literacy framework can be developed based on this research which will benefit the education field in Malaysia.

Keywords: assessment literacy, ESL, exploratory factor analysis, principal component analysis

INTRODUCTION

Education plays an important part in the country's development because the nation's future depends on what is happening in its classroom in the present (MEB, 2013-2025). The Ministry of Education has introduced many changes to maintain our status as a developed nation (Varatharaj, Abdullah & Ismail, 2014). The Ministry of Education has launched many initiatives in the education policy to reform the education system (Muhammad, Ali, Zamani, Yamin & Ismail, 2020). The fundamental aspect of it is to introduce creativity, innovation, and higher-order thinking skills in teaching and learning.

The report by OECD (2013) showed that the learning standards in Malaysia have deteriorated over the past 10 years. Malaysian students have performed poorly in international assessments such as The Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). These assessments measure cognitive abilities such as application and reasoning skills (MEB, 2013-2015). The report stated that 35% and 38% of Malaysian students are unsuccessful in meeting a good proficiency level in both Mathematics and Science. According to the Ministry of Education (2013), Malaysian students were only good at duplicating subject content in the past, but now these skills are not practical today. The students must be able to analyze information, think critically, and relate what they have learned in a real-world setting. However, Malaysian students are having problems with higher-order thinking skills.

School-Based Assessment

Consequently, the Malaysian Examination Syndicate (MES) reformed education in Malaysia by implementing School-Based Assessment (SBA) in agreement with Standards-Based Primary School Curriculum (KSSR), which started in 2011, and Standards-Based Secondary School Curriculum (KSSM) in 2012 (Mansor et al., 2019). The Ministry of Education has also revamped the nationwide examination and school-based assessment by increasing the number of questions that assess higher-order thinking skills up to 50% for SPM. It is to ensure that teachers will stop drilling students on the content tested in the examination and instead train students in critical thinking skills and relate their knowledge in various contexts (MEB, 2013-2015). It is hoped that this step will provide an opportunity to move away from only relying on theoretical brilliance and focus more on assessing students holistically. The mid-term and final examinations have been entirely dismissed in 2019 for primary years 1 to 6. SBA has replaced it to create an 'exam-free environment' for younger learners (Mansor et al., 2019).

SBA was introduced to give more priority to formative assessment rather than summative assessment. It is aimed at engaging students in learning (Arumugham, 2019). According to Saad, Jani and Rahmat (2017), SBA is a continuous form of assessment used to assess teaching and learning in the classroom. It does not focus merely on test scores as it considers all aspects of students' intellectual and personality development. So, it is considered holistic because it assesses cognitive, affective and psychomotor domains. The assessment is separated into academic and non-academic. A school-based assessment is an example of an academic assessment. In contrast, the non-academic assessment consists of a psychometric assessment which assesses students in terms of physical, sports and co-curricular activities (Lee, Loo & Chua, 2018).

Talib, Kamsah, Naim and Latif (2014), stated that SBA has changed how assessment is being conducted in Malaysia based on two dimensions, namely, the emphasis on formative assessment rather than

summative assessment and using criterion-referenced rather than only relying on norm-referenced. The researchers also said that SBA focuses more on balancing between centralized exams and classroom assessment. It is hoped that the introduction of classroom assessment will support student-centered learning, which is the objective of the new KSSR and KSSM syllabus. According to Veloo, Ramli, and Khalid (2016), the implementation of SBA allows teachers to have independence in planning the assessment method, constructing assessment instruments, analyzing data, and reporting data. So, the teachers' responsibility has increased and the importance of teachers' assessment literacy has also amplified.

The Ministry of Education has reformed the education system as an initiative to improve the standard of English education in Malaysia and match the international standard (Alih, Yusoff & Raof, 2020). However, studies done in the past have revealed that primary and secondary school teachers lack assessment skills (Sidhu, Kaur & Lee, 2018). This concern has also been discussed in the media and within the educational system. According to Tajularipin et.al. (2015), teachers lack the knowledge and skills essential for implementing the latest curriculum. Teachers are also not sufficiently ready to apply the new curriculum to their students. Therefore, it is vital to find a way to solve these problems ESL teachers face in schools.

Thus, the research is directed by the research questions as follows:

a) Which set of questions should be incorporated in the final instrument developed from the analyses of the instrument's psychometric properties that measure teachers' assessment literacy?
b) What is the reliability of the instrument to measure teachers' assessment literacy that is developed?

Statement of the Problem

Recently, the Ministry of Education (MOE) has reformed the education policy. One of them is the introduction of School-Based Assessment (SBA) in schools. As a result, there is a higher expectation for school teachers to assess students competently (Lian & Yew, 2016). For the educational policy to be successfully implemented, teachers must be adequately prepared and knowledgeable enough for this responsibility (Nimehchisalem, Foo & Nowrouzi, 2019). According to Lian, Yew and Meng (2014), several supports have been provided by the Ministry of Education in Malaysia for school teachers such as training programmes, workshops, modules, guidebooks, etc. Yet, teachers' assessment literacy is still a major concern.

Singh, Supramaniam, and Teoh (2017) revealed that the understanding of the rationale for implementing SBA is not present in school teachers. It was further elaborated that although some teachers understand the principles of SBA, there were still problems when it comes to putting it into practice. The professional development courses introduced to the teachers did not help the teachers increase their willingness to implement SBA (Singh, et al., 2017). Teachers were still struggling to plan, develop and analyse instruments of SBA.

Some research has been carried out in Malaysia related to the assessment literacy of teachers (Asri, 2007; Rohaya & Mohd Najid, 2008; Suah 2012). However, most of the studies were only about investigating the knowledge that pre-service teachers in assessing students, monitoring how in-service teachers are doing assessment and also the difficulty faced in the application of SBA among in-service teachers. Not much attention was given to proposing a solution to the problems faced by teachers in assessing students' language competence. There is not much literature on developing an assessment literacy framework that ESL teachers could utilise as a reference to comprehend the suitability of assessment methods used to collect reliable evidence about the achievement of students in the classroom (Asri, 2007; Rohaya & Mohd Najid, 2008; Suah 2012).

LITERATURE REVIEW

Assessment Literacy

As defined by Stiggins (1991), assessment literacy is the basic knowledge teachers need concerning assessment in education. Mertler (2004) said that assessment literacy is the understanding of the assessment

principles, for instance, testing, evaluation and measurement. It also involves the use of proper assessment tools to assess students. According to Xu and Brown (2017), teachers must understand the purpose of assessment, suitable tools to measure performance, and how to interpret assessment results. Teachers' assessment literacy determines the accomplishment of educational assessment and increases the value of learning (Xu & Brown, 2017).

According to Mertler and Campbell (2005), it is a serious obligation of classroom teachers to assess how students perform. School teachers spend a large amount of time assessing students yet many teachers still do not feel like they are sufficiently ready for this undertaking (Mertler & Campbell, 2005). Gan, He and Liu (2019) stated that assessment in classrooms consists of teachers and students playing active roles in electing and creating opportunities that involve them to keep track of learning progress and motivate them. This means that assessment is not merely used to report how much students have learnt at the end of the year but also used in classroom settings to gauge students' learning growth and also to stimulate them to learn.

Studies on the Assessment of Literacy

Research related to teachers' assessment literacy has become prevalent, particularly in language teaching around the globe. One of the studies is done by Vogt and Tsagari (2014) who did a study to measure the assessment literacy level of European foreign language teachers and ascertain their training needs. The data were collected using a mixed-method approach from seven European countries. The researchers gave questionnaires to 853 teachers and conducted interviews with 63 teachers in selected countries throughout Europe. By using these methods, the researchers were able to generalize the findings across Europe and also analyse the interview data in-depth. Descriptive analysis in frequency and percentages is used to analyze the data obtained from the questionnaire. The results from teachers' interviews were analyzed using content analysis.

The findings exposed that 68.3% of the participants had received general training in assessment. However, regarding classroom-based assessment training specifically for testing, 42% of participants said they did not receive any training. Teachers' ability to design tests and the concepts and content of assessments were underdeveloped. Half of the participants also mentioned that they would need to receive advanced assessment training. The findings from the interview showed that pre-service teacher education programs did not prepare teachers adequately for assessment. The teachers also said they were not given sufficient training after becoming in-service teachers. The study concluded that sufficient training must be given to teachers across Europe in language assessment and testing (Vogt & Tsagari, 2014).

Ashraf and Zolfaghari (2018) also researched English as a Foreign Language (EFL) teachers' assessment literacy concentrating on reflective teaching. The study intended to explore the association between Iranian teachers' assessment literacy and reflective teaching. The study is a quantitative study where two sets of questionnaires were used, namely the assessment literacy inventory and the reflective teaching questionnaire. Both sets of questionnaires were given to 120 teachers in Iran who teach EFL. The data were collected and then analyzed using SEM. A structural model was proposed to scrutinize the associations between teachers' assessment literacy and reflective teaching. The results showed that there is a positive association between the two variables. It was also stated that reflective teaching teachers can better evaluate students' improvement. The study concluded that teacher training is necessary for assessment and evaluation to provide adequate knowledge and skills to teachers (Ashraf & Zolfaghari, 2018).

Xie and Tan (2019) researched to explore the assessment literacy needs of primary school English language teachers in Hong Kong. The research focuses on both pre-service and in-service teachers. The study employed a mixed-method methodology that started with interviews with in-service teachers and then, the results from the interview were used to design questionnaires for pre-service teachers. The interview was conducted with 11 primary school teachers. The data were analyzed using emerging codes and patterns categorized into themes for the qualitative phase. For the quantitative phase, the data were analyzed descriptively, including mean and standard deviation. The findings revealed that novice teachers

lack skills in interpreting and following marking schemes and confidence in test designing (Xie & Tan, 2019).

Another study on assessment literacy was conducted by Sultana (2019) which examines the nature and also the functionality of language assessment literacy amongst English language teachers. The study's emphasis is on teachers' readiness to implement several assessment tasks and the perception of assessment literacy by teachers. The study data was gathered using semi-structured interviews to thoroughly understand the setting and participants. The participants were 10 secondary English teachers from 5 different schools. The data were analyzed using inductive thematic analysis. The findings revealed that the teachers were not confident and expressed uncertainty when designing assessment tasks. Most of the assessment tasks given by the teachers are written tasks. There is no variety in the assessment task given. Teachers also did not understand the test's purpose, which resulted in teaching the test. The researcher suggests providing teachers with professional training in language assessment to increase their assessment literacy (Sultana, 2019).

The studies discussed on the assessment literacy of teachers show that this problem is present everywhere and not limited to one country. Based on the studies, teachers find it difficult to carry out assessment practice because of their inadequate knowledge and skills.

Assessment Literacy of Malaysian Teachers

There have not been a lot of studies done on assessment literacy in Malaysia. One study conducted by Talib and Naim in 2012 aimed to identify teachers' assessment literacy level due to the implementation of SBA in schools. The Literacy Assessment Test (LAT) which consisted of 45 items, was administered to 465 teachers in Johor. The data were analyzed using a program called Quest 2 which combines Item Test Theory and Classical Test Theory. The findings presented that 86.67% of teachers possess medium to low literacy levels in assessment. The researchers suggested that teachers need more training to prepare themselves for the implementation of SBA in schools.

Another research was conducted by Ghazali, Rabi, Hassan and Wahab (2018), which targeted validating a Classroom Assessment Practices (CAP) instrument. Questionnaires were administered to 320 secondary school teachers to gather data. Then, the CAP was scrutinized using exploratory factor analysis as well as confirmatory factor analysis. The findings showed that all the criteria were fit and the constructs were reliable. So, the instrument was found to be valid. Teachers can use the instrument as a self-assessment tool and to inspect classroom assessment practices among Malaysian teachers.

To execute excellent teaching and learning, teachers must be skilled in classroom assessment (Mohamed, et al., 2016). A study was done by Mohamed, Kamis, and Ali in 2016 to gauge home economics teachers' assessment literacy. The measurement was centered on the Standard for Teachers' Competence in Educational Assessment of Students. The instrument used was questionnaires of 41 multiple-choice items administered to 187 home economics teachers. The study's outcome presented that the teachers have moderate to low levels of assessment literacy. The overall performance of all 41 items is 55% in the average score, which is unsatisfactory. The findings also indicated that the teachers' assessment literacy was low when explaining assessment results to students (Mohamed, et al., 2016).

Kalajahi and Abdullah (2016) also studied assessment literacy in Malaysia, focusing on lecturers rather than teachers. The researchers said that the assessment literacy of lecturers is very important because higher institutions depend merely on lecturers in assessing students' understanding and skills. The study evaluates the assessment literacy of lecturers and also looks into common assessment practices. 65 lecturers from a public university in Malaysia are given questionnaires. The outcome of the research showed that the assessment literacy level of lecturers was unsatisfactory. The study showed a weak association between assessment practices and assessment beliefs. The lecturers mostly rely on summative assessments to assess students (Kalajahi & Abdullah, 2016).

Based on previous studies, it is evident that Malaysian teachers' assessment literacy level is unsatisfactory. It is imperative to keep the teachers updated on the knowledge and skills of assessing students formatively to implement SBA successfully.

METHOD

Research Design

Research design is the first step in planning and organizing a research blueprint (Creswell, 2012). This study aims to develop an instrument to measure the assessment literacy of ESL teachers in Malaysia. To accomplish the objective of this study, a mixed method approach, specifically a sequential exploratory design, is employed because it aims to identify important themes and then develop an instrument and test it (Fraenkel, Wallen, Hyun, 2016). A sequential exploratory design is used where the researcher starts with gathering qualitative data first, and then quantitative data is collected and used to test data empirically (Shorten & Smith, 2017).

Sampling

The sampling technique used for this study's second phase is cluster random sampling. According to Kothari (2004), cluster random sampling is used when a list of all population members is unavailable. So, the researcher selects clusters or groups of teachers from randomly chosen schools from the state of Perak as the research sample. 200 teachers from secondary schools in Perak were chosen as the respondents for this research. Boomsma (1982) stated that a sample size of a minimum of 200 is enough to attain reliable results when conducting a factor analysis.

Instruments for Data Collection

Instrumentation is the process of preparing to collect data. According to Fraenkel, Wallen and Hyun (2016), there are two ways to obtain an instrument to collect data in research. The first method is to use a previously existing instrument and administer it. The second method is to personally develop an instrument to collect the desired data. The researcher developed an instrument for this study to collect data from secondary school English teachers in Perak. Therefore, the instrumentation used in this study will be a survey questionnaire.

The questionnaire from this study has been developed based on the constructs attained from the analysis of past literature and documents. The survey consists of two parts. The first part of the instrument included demographic information on gender, age, teaching experience, and highest educational level. The second part of the survey consists of the constructs acquired from the literature review which are teachers' belief in assessment practices, teachers' knowledge of assessing students, knowledge of validity and reliability in assessment, interpretability of results and washback, knowledge of assessment methods and teachers' training on assessment.

The instrument used a five-point Likert scale. The scale ranges from 1: Strongly Disagree to 5: Strongly Agree. The questionnaire items gather secondary school ESL teachers' opinions of the importance of each item for assessment practices in schools. The respondents must indicate whether they agree or disagree with the items by selecting the appropriate response. Since a survey questionnaire allows researchers to gather information from a large sample of respondents, it is a suitable technique to collect data from secondary school ESL teachers all over Perak. The survey questionnaire is sent to selected schools using email, which is a Google Form link.

Data Analysis

Exploratory Factor Analysis (EFA) is used to analyze the data from the study. EFA is used to explore the newly developed factors underlying the target population. The constructs for this study were developed based on document analysis, and have been validated by selected experts using the Delphi method. Williams et al. (2010), explained that the researcher can explore the main dimensions to create a theory or model from a fairly large set of latent constructs frequently symbolized by a set of items using EFA. For this study, EFA is used to identify the domain substrata, remove cross-loaded variables and check the reliability of the items (Hair, William and Babin, 2011). Kaiser criterion (i.e., eigenvalue > 1) was used to decide the number of factors in this study.

The Statistical Package for Social Science (SPSS) software version 25 is used to conduct the factor analysis for this study. A 92-item questionnaire was distributed to 200 in-service ESL teachers from secondary schools in Perak. The analysis was conducted to ensure that the items are suitable for the constructs developed in the questionnaire (Tabachnick & Fidell, 2001). To identify the factors specific to the construct, the Principle Component Analysis (PCA) with varimax rotation as well as Kaiser normalisation was conducted.

The analysis starts by using Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy Test (KMO) to check the appropriateness of the respondents' data for conducting factor analysis in the following step. The KMO index ranges from 0 to 1, with 0.50 considered appropriate for factor analysis. Bartlett's Test of Sphericity should be significant ($p<.05$) for factor analysis to be adequate (Williams, Onsman and Brown, 2010). In this study, the result of Bartlett's Test is ($p<0.05$) which means that the correlation between the items is adequate. The result of KMO is between 0.8 and 1. So the data is factorable for the next stage.

According to Watkins (2018), there are many ways of factor extraction. The Principle Components Analysis (PCA) analyses the entire correlation matrix and intends to reduce data while retaining information from the original data set. Williams et al. (2010) stated that the purpose of doing extraction of data is to decrease a large number of items into factors. This study uses PCA as the extraction method with a factor loading cut-off point of 0.60. According to Tabachnick and Fidell (2001), cut-offs going from 0.32 is poor, 0.45 is fair, 0.55 is good, 0.63 is very good and 0.71 is excellent. Therefore, only the item with a cut-off point of 0.60 is retained in this study. Out of 92 items, only 78 questions are retained and grouped into sub-constructs based on the factor loading. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was 0.922, which shows that the sample was adequate, and Bartlett's Test of Sphericity gave a p-value of <0.001.

RESULTS

EFA Result for Teachers' Beliefs in Assessment Practices

The KMO index for the first construct, teachers' belief in assessment practices, is .910, more than 0.5, indicating that the items are factorable and adequate for the next stage of factor extraction. Meanwhile, Bartlett's Test of Sphericity is highly significant ($p = .000 < .05$), indicating that the items are sufficiently correlated.

The result from the components and total variance explained shows that five components are based on the EFA procedure with an eigenvalue of more than 1.0. The eigenvalues ranged between 1.080 and 12.100. The total variance explained for Component 1 is 44.814, component 2 is 10.659, component 3 is 6.940, component 4 is 5.842, and component 5 is 4.000. The total variance explained cumulatively is 72.255, which is more than 60%. Table 1 shows the total variance explained.

TABLE 1
THE COMPONENTS AND TOTAL VARIANCE EXPLAINED FOR TEACHERS' BELIEF IN ASSESSMENT PRACTICES

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	12.100	44.814	44.814	12.100	44.814	44.814
2	2.878	10.659	55.473	2.878	10.659	55.473
3	1.874	6.940	62.412	1.874	6.940	62.412
4	1.577	5.842	68.255	1.577	5.842	68.255
5	1.080	4.000	72.255	1.080	4.000	72.255

Extraction Method: Principal Component Analysis

The factor loading for each item shows that out of 27 items, only 24 items have been retained. 24 items have a factor loading of more than 0.6. Items number 2, 4 and 19 are deleted. Table 2 shows the factor loading for each item.

TABLE 2
FACTOR LOADING FOR TEACHERS' BELIEF IN ASSESSMENT PRACTICES

Rotated Component Matrix[a]					
	Component				
	1	2	3	4	5
Q1					.662
Q2 Deleted					
Q3					.686
Q4 Deleted					
Q5			.840		
Q6			.725		
Q7			.738		
Q8			.774		
Q9			.644		
Q10		.606			
Q11		.803			
Q12		.818			
Q13		.714			
Q14		.756			
Q15				.752	
Q16				.747	
Q17				.787	
Q18				.649	
Q19 Deleted					
Q20	.607				
Q21	.797				
Q22	.758				
Q23	.888				
Q24	.802				
Q25	.819				
Q26	.679				
Q27	.623				

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Based on the factor loading, the items are clustered into five sub-constructs and renamed. So, the first construct which is teachers' belief in assessment practices is grouped into five sub-constructs namely Formative assessment, Formative assessment benefits, Assessment planning, Summative assessment and Alternative assessment. Table 3 shows the newly named sub-constructs and their internal reliability.

TABLE 3
NEW SUB-CONSTRUCTS AND THEIR INTERNAL RELIABILITY

Sub-Construct	Items	Total	Cronbach Alpha
Formative Assessment	1,3	2	.801
Formative assessment benefits	5,6,7,8,9	5	.880
Assessment planning	10,11,12,13,14	5	.901
Summative Assessment	15,16,17,18	4	.868
Alternative Assessment	20,21,22,23,24,25,26,27	8	.930

Internal reliability measures whether the results are consistent across items measuring the same construct. To ensure the items attain high internal reliability, Cronbach's Alpha value must be more than .7 (Rovai et al., 2014). The Cronbach Alpha for the sub-constructs above ranges from .801 to .930. Therefore, all the items in the sub-constructs have an internal reliability of more than .70. It means all the items grouped in the newly named constructs have acceptable internal reliability.

EFA Result for Teachers' Knowledge of Assessing/Testing Students

The KMO index for the second construct which is teachers' knowledge in assessing and testing students is .927, indicating that the items are factorable and adequate for the next stage of factor extraction. Meanwhile, Bartlett's Test of Sphericity is highly significant ($p = .000 < .05$), indicating that the items are sufficiently correlated.

The result from the components and total variance explained shows three components based on the EFA procedure with an eigenvalue of more than 1.0. The eigenvalues ranged between 1.169 and 12.920. The total variance explained for Component 1 is 51.678, component 2 is 8.677, and Component 3 is 4.675. The total variance explained cumulatively is 65.030, which is more than 60%. Table 4 shows the total variance explained.

TABLE 4
THE COMPONENTS AND TOTAL VARIANCE EXPLAINED FOR TEACHERS' KNOWLEDGE IN ASSESSING AND TESTING STUDENTS

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	12.920	51.678	51.678	12.920	51.678	51.678
2	2.169	8.677	60.355	2.169	8.677	60.355
3	1.169	4.675	65.030	1.169	4.675	65.030

Extraction Method: Principal Component Analysis

The factor loading for each item shows that out of 25 items, only 18 items have been retained. Only 18 items have a factor loading of more than 0.6. Item 28, 29, 35, 37, 43, 44 and 45 is deleted. Table 5 shows the factor loading for each item.

TABLE 5
FACTOR LOADING FOR TEACHERS' KNOWLEDGE OF ASSESSING AND TESTING STUDENTS

Rotated Component Matrix[a]			
	Component		
	1	2	3
Q28 Deleted			
Q29 Deleted			
Q30			.829
Q31			.776
Q32			.636
Q33			.676
Q34			.712
Q35 Deleted			
Q36		.623	
Q37 Deleted			
Q38		.704	
Q39		.684	
Q40		.754	
Q41		.718	
Q42		.636	
Q43 Deleted			
Q44 Deleted			
Q45 Deleted			
Q46	.615		
Q47	.695		
Q48	.799		
Q49	.724		
Q50	.785		
Q51	.750		
Q52	.712		

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization [a]
a. Rotation converged in 6 iterations.

Based on the factor loading, the items are clustered into three sub-constructs and renamed. So, the first construct which is teachers' knowledge in assessing and testing students are grouped into three sub-constructs namely Teacher's assessment practices, Teacher's assessment knowledge and Skills in developing assessment tools. Table 6 shows the newly named sub-constructs and their internal reliability.

TABLE 6
NEW SUB-CONSTRUCTS AND THEIR INTERNAL RELIABILITY

Sub-Construct	Items	Total	Cronbach Alpha
Teacher's assessment practices	30, 31, 32, 33, 34	5	.861
Teacher's assessment knowledge	36, 38, 39, 40, 41, 42	6	.892
Skills in developing assessment tools	46, 47, 48, 49, 50, 51, 52	7	.909

The Cronbach Alpha for the sub-constructs above ranges from .861 to .909. Therefore, all the items in the sub-constructs have an internal reliability of more than .70. It means all the items grouped in the newly named constructs have acceptable internal reliability.

EFA Result for Knowledge of Validity and Reliability in Assessment

The KMO index for the third construct which is Knowledge of validity and reliability in assessment is .928, indicating that the items are factorable and adequate for the next stage of factor extraction. Meanwhile, Bartlett's Test of Sphericity is highly significant ($p = .000 < .05$), indicating that the items are sufficiently correlated.

The result from the components and total variance explained shows that there are three components based on the EFA procedure with an eigenvalue of more than 1.0. The eigenvalues ranged between 1.040 and 11.072. The total variance explained for Component 1 is 61.509, component 2 is 7.464, and Component 3 is 5.776. The total variance explained cumulatively is 74.749 which is more than 60%. Table 7 shows the total variance explained.

TABLE 7
THE COMPONENTS AND TOTAL VARIANCE EXPLAINED FOR KNOWLEDGE IN VALIDITY AND RELIABILITY IN ASSESSMENT

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.072	61.509	61.509	11.072	61.509	61.509
2	1.343	7.464	68.973	1.343	7.464	68.973
3	1.040	5.776	74.749	1.040	5.776	74.749

Extraction Method: Principal Component Analysis

The factor loading for each item shows that out of 18 items, only 15 items have been retained. Only 15 items have a factor loading of more than 0.6. Item numbers 60, 61 and 63 are deleted. Table 8 shows the factor loading for each item.

TABLE 8
FACTOR LOADING FOR TEACHERS' KNOWLEDGE OF VALIDITY AND RELIABILITY IN ASSESSMENT

Rotated Component Matrix[a]

	Component		
	1	2	3
Q53		.796	
Q54		.797	
Q55		.847	
Q56		.762	
Q57			.751
Q58			.771
Q59			.670
Q60 Deleted			
Q61 Deleted			
Q62		.645	
Q63 Deleted			
Q64	.660		

Q65	.675
Q66	.712
Q67	.657
Q68	.811
Q69	.841
Q70	.869

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]
a. Rotation converged in 6 iterations.

Based on the factor loading, the items are clustered into three sub-constructs and renamed. So, the third construct which is Knowledge of validity and reliability in assessment is grouped into three sub-constructs namely Validity, Reliability and Item Analysis. Table 9 shows the newly named sub-constructs and their internal reliability.

TABLE 9
NEW SUB-CONSTRUCTS AND THEIR INTERNAL RELIABILITY

Sub-Construct	Items	Total	Cronbach Alpha
Validity	53, 54, 55, 56, 62	5	.909
Reliability	57, 58, 59	3	.782
Item Analysis	64, 65, 66, 67, 68, 69, 70	7	.948

The Cronbach Alpha for the sub-constructs above ranges from .782 to .948. Therefore, all the items in the sub-constructs have an internal reliability of more than .70. It means all the items grouped in the newly named constructs have acceptable internal reliability.

EFA Result for Interpretability of Result and Washback

The KMO index for the fourth construct which is Interpretability of result and washback is .890, indicating that the items are factorable and adequate for the next stage of factor extraction. Meanwhile, Bartlett's Test of Sphericity is highly significant ($p = .000 < .05$), indicating that the items are sufficiently correlated.

The result from the components and total variance explained shows that there are two components based on the EFA procedure with an eigenvalue of more than 1.0. The eigenvalues ranged between 1.393 and 6.215. The total variance explained for component 1 is 56.497 and for component 2 is 12.667. The total variance explained cumulatively is 69.164 which is more than 60%. Table 10 shows the total variance explained.

TABLE 10
THE COMPONENTS AND TOTAL VARIANCE EXPLAINED FOR INTERPRETABILITY OF RESULT AND WASHBACK

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.215	56.497	56.497	6.215	56.497	56.497
2	1.393	12.667	69.164	1.393	12.667	69.164

Extraction Method: Principal Component Analysis

The factor loading for each item shows that all 11 items have been retained from this construct. All 11 items have a factor loading of more than 0.6. No items are deleted. Table 11 shows the factor loading for each item.

TABLE 11
FACTOR LOADING FOR TEACHERS' INTERPRETABILITY OF RESULTS AND WASHBACK

Rotated Component Matrix[a]		
	Component	
	1	2
Q71		.887
Q72		.835
Q73		.747
Q74	.700	
Q75	.697	
Q76	.629	
Q77	.866	
Q78	.835	
Q79	.848	
Q80	.816	
Q81	.777	

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]
a. Rotation converged in 3 iterations.

Based on the factor loading, the items are clustered into two sub-constructs and renamed. So, the fourth construct which is the Interpretability of result and washback is grouped into two sub-constructs namely Interpretability and Washback. Table 12 shows the newly named sub-constructs and their internal reliability.

TABLE 12
NEW SUB-CONSTRUCTS AND THEIR INTERNAL RELIABILITY

Sub-Construct	Items	Total	Cronbach Alpha
Interpretability	71, 72, 73	3	.824
Washback	74, 75, 76, 77, 78, 79, 80, 81	8	.922

The Cronbach Alpha for the sub-constructs above ranges from .824 to .922. Therefore, all the items in the sub-constructs have an internal reliability of more than .70. It means all the items grouped in the newly named constructs have acceptable internal reliability.

EFA Result for Knowledge of Assessment Methods

The KMO index for the fifth construct which is Knowledge of assessment methods is .887, indicating that the items are factorable and adequate for the next stage of factor extraction. Meanwhile, Bartlett's Test of Sphericity is highly significant ($p = .000 < .05$), indicating that the items are sufficiently correlated.

The result from the components and total variance explained shows that there are two components based on the EFA procedure with an eigenvalue of more than 1.0. The eigenvalues ranged between 1.258 and 5.845. The total variance explained for component 1 is 53.140 and for component 2 is 11.438. The total variance explained cumulatively is 64.577 which is more than 60%. Table 13 shows the total variance explained.

TABLE 13
THE COMPONENTS AND TOTAL VARIANCE EXPLAINED FOR KNOWLEDGE OF ASSESSMENT METHODS

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.845	53.140	53.140	5.845	53.140	53.140
2	1.258	11.438	64.577	1.258	11.438	64.577

Extraction Method: Principal Component Analysis

The factor loading for each item shows that out of 11 items, only 10 items have been retained. Only 10 items have a factor loading of more than 0.6. Item number 86 is deleted. Table 14 shows the factor loading for each item.

TABLE 14
FACTOR LOADING FOR KNOWLEDGE OF ASSESSMENT METHODS

Rotated Component Matrix[a]

	Component	
	1	2
Q82		.754
Q83		.810
Q84		.749
Q85		.620
Q86 Deleted		
Q87	.712	
Q88	.739	
Q89	.668	
Q90	.879	
Q91	.827	
Q92	.884	

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]
a. Rotation converged in 3 iterations.

Based on the factor loading, the items are clustered into two sub-constructs and renamed. So, the fifth construct, Knowledge of assessment methods, is grouped into two sub-constructs: Knowledge of objective-type assessment and Knowledge of subjective-type assessment. Table 15 shows the newly named sub-constructs and their internal reliability.

TABLE 15
NEW SUB-CONSTRUCTS AND THEIR INTERNAL RELIABILITY

Sub-Construct	Items	Total	Cronbach Alpha
Knowledge of objective-type assessment	82, 83, 84, 85	4	.763
Knowledge of subjective-type assessment	87, 88, 89, 90, 91, 92	6	.906

The Cronbach Alpha for the sub-constructs above ranges from .763 to .906. Therefore, all the items in the sub-constructs have an internal reliability of more than .70. It means all the items grouped in the newly named constructs have acceptable internal reliability.

DISCUSSION

This study intends to develop a valid and reliable instrument to measure ESL teachers' assessment literacy based on the five constructs identified based on document analysis of past literature and validated by a panel of experts. The five constructs are teachers' belief in assessment practices, teachers' knowledge of assessing and testing students, validity and reliability in assessment, interpretability of result and washback, and knowledge of assessment methods. Exploratory Factor Analysis (EFA) is used to validate the constructs of the instrument and Cronbach's Alpha is used to determine the reliability of the items. Based on the analysis, Bartlett's Test result is highly significant ($p = .000 < .001$) for all the items in all five constructs. This means that the items are sufficiently correlated. The KMO is used to test the sampling adequacy for the analysis. The KMO results for all five constructs are between 0.8 and 1, indicating that the data is adequate and factorable for the next stage.

As for the components and total variance explained, the analyses indicated that there are five components with an eigenvalue of more than 1.0. The first construct (teachers' beliefs in assessment practices) accounted for 72.255% of the variance, the second construct (teachers' knowledge in assessing and testing students) accounted for 65.030%, the third construct (knowledge of validity and reliability in assessment) accounted for 74.749%, the fourth construct (interpretability of result and washback) accounted for 69.164%, and the fifth construct (knowledge of assessment method) accounted for 64.577%. Hence, it can be observed that all five constructs have the total variance explained cumulatively, which exceeded more than 60% (Bahkia et al., 2019).

The factor loading of more than 0.6 is retained while the items below the cut-off point of 0.6 are deleted (Awang, 2014; Yahaya et al., 2018). Therefore, only 24 out of 27 items are retained for the first construct while 3 items are deleted. As for the second construct, 18 out of 25 items have been retained while 7 items are deleted. For the third construct, 15 out of 18 items have been retained, whereas three items are deleted. All 11 items have been retained for the fourth construct, so no items were deleted. For the final construct, 10 out of 11 items were retained and only one was deleted.

After the factor loading is conducted, all the items for each construct were grouped into sub-constructs and renamed. Once the sub-constructs have been renamed, Cronbach's Alpha has been employed to compute the internal consistency of the items. Internal reliability tests the consistency of findings throughout the items assessing identical construct. The Cronbach's Alpha must be more than .70 for the items to have satisfactory internal reliability (Rovai et al., 2014). The internal consistency for the items in the sub-constructs for all five main constructs is between .763 and .948. Therefore, all the items have acceptable to excellent reliability. EFA and reliability analyses have shown that the survey instrument has good construct validity and reliability.

CONCLUSION

Five constructs on assessment literacy of ESL teachers have been identified based on analysis of past documents and literature. The constructs were then validated by a panel of experts for content validity. Exploratory Factor analysis is then used to validate the constructs. EFA results indicated that out of 92 items, only 78 items have a factor loading of above .60. The KMO and Bartletts' Test disclosed that the data is adequate to be used for factor analysis and are correlated. Besides, all the items have good internal consistency and are reliable based on Cronbach's Alpha value. Thus, this study has established that the instrument to measure ESL teachers' assessment literacy is valid and reliable. For future studies, confirmatory factor analysis (CFA) should be carried out to define the relationships among the latent variables.

ACKNOWLEDGMENTS

This research has been carried out under the Fundamental Research Grants Scheme 2019-0151-107-02 (FRGS/1/2019/SS109/UPSI/02/18) provided by the Ministry of Education of Malaysia. The authors would like to extend their gratitude to Universiti Pendidikan Sultan Idris (UPSI) who helped manage the grants. Equally, we would like to thank the participating in-service teachers who were truly cooperative.

REFERENCES

Alih, N.A.C., Yusoff, M.A., & Raof, A.H.A. (2020). Teachers' Knowledge and Belief on the CEFR Implementation in Malaysian ESL Classroom. *International Journal of Multidisciplinary and Current Educational Research (IJMCER)*, *2*(5), 126–134.

Arumugham, K.S. (2019). Teachers' Understanding Towards Portfolio Assessment: A Case Study among Malaysian Primary School Teachers. *Problems of Education in the 21st Century*, *77*(6), 695–704.

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

Asri, S. (2007). *Teaching and learning assessment practice in Bestari secondary school, Johor State*. [PhD dissertation, Universiti Teknologi Malaysia, Skudai, Malaysia].

Awang, Z. (2014). *A Handbook on SEM for Academicians and Practitioners: The Step-by-Step Practical Guides for the Beginners*. MPWS Rich Resource.

Bahkia, A.S., Awang, Z., Afthanorhan, A., Ghazali, P.L., & Foziah, H. (2019). Exploratory factor analysis on occupational stress in the context of Malaysian sewerage operations. *The 4th Innovation and Analytics Conference & Exhibition (IACE 2019)*.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In H. Wold, & K. Jöreskog (Eds.), *Systems under indirect observations* (pp. 147–173). New York: Elsevier North-Holland.

Creswell, J.W. (2012). *Educational Research; Planning, Conducting and Evaluating Quantitative and Qualitative Research* (4th Ed.). University of Nebraska-Lincoln: Pearson Education.

Frankel, J.R., Wallen N.E., & Hyun, H.H. (2016). *How to Design and Evaluate Research in Education* (9th Ed.). New York: McGraw Hill.

Gan, Z., He, J., & Liu, F. (2019). Understanding Classroom Assessment Practices and Learning Motivation in Secondary EFL Students. *The Journal of Asia TEFL*, *16*(3), 783–800.

Ghazali, N.H.C.M., Rabi, N.M., Hassan, N.M., & Wahab, N.A. (2018). A Confirmatory Factor Analysis of Classroom Assessment Practises Scale in a Malaysian Context. *International Journal of Academic Research in Progressive Education and Development*, *7*(3), 516–529.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2011). *Multivariate Analysis* (7th Ed.). New Jersey, Prentice Hall.

Kalajahi, S.A., & Abdullah, A.N. (2016). Assessing Assessment Literacy and Practices among Lecturers. *Pedagogika*, *124*(4), 232–248. https://doi.org/10.15823/p.2016.65

Kothari, C.R. (2004). *Research Methodology: Methods and Techniques* (2 Revised Edition). New Age International Publishers.

Lee, S.M., Loo, F.Y., & Chua, Y.P. (2018). School-Based Assessment: A Case Study of a Lower Secondary Schools Music Education in Malaysia. *Malaysian Journal of Performing and Visual Arts*, *4*(1), 18–35. https://doi.org/10.22452/mjpva.vol4no1.2

Lian, L.H., & Yew, W.T. (2016). A Framework for Examining Assessment Literacy of Preservice Teachers. *US-China Education Review A*, *6*(5), 294–300.

Lian, L.H., Yew, W.T., & Meng, C.C. (2014). Enhancing Malaysian Teachers' Assessment Literacy. *International Education Studies*, *7*(10), 74–81.

Mansor, A.N., Vikaraman, S.S., Medina, N.I., & Alias, B.S. (2019). Managing School-Based Assessment: Challenges and Solutions for Educational Practice. *International Journal of Innovation, Creativity and Change*, *7*(7), 63–84.

Mertler, C.A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, *33*, 49–64.
Mertler, C.A., & Campbell, C. (2005). *Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the Assessment Literacy Inventory*. Montréal, Quebec, Canada: American Educational Research Association.
Ministry of Education, Malaysia. (2013). *Malaysia Education Blueprint 2013-2025*. Putrajaya, Malaysia.
Mohamed, S., Kamis, A., & Ali, N. (2016). Gauging the assessment literacy of Malaysia's home economics teachers: An empirical study. *Malaysian Journal of Society and Space*, *12*(3), 130–138.
Muhammad, N., Ali, N., Zamani, S., Yamin, N.A., & Ismail, N.N. (2020). Examining Assessment Literacy: A Study of Technical Teacher. *European Journal of Molecular & Clinical Medicine*, *7*(8), 705–717.
Nimehchisalem, V., Foo, E.S.K., & Nowrouzi, S. (2019). English as a Second Language Learners and Teachers' Conceptions of Language Assessment. *The Journal of Asia TEFL, 16*(4), 1348–1359. https://doi.org/10.18823/asiatefl.2019.16.4.18.1348
Rohaya, T., & Mohd Najib, N.G. (2008, November 25–27). Development and validation of an instrument to measure Malaysian secondary school teachers' assessment literacy level. In *Proceedings of Seminar Penyelidikan Pendidikan Pasca Ijazah*, Universiti Teknologi Malaysia.
Rovai, A.P., Baker, J.D., & Ponton, M.K. (2014). *Social science research design and statistics: A practitioner's guide to research methods and IBM SPSS.* Watertree Press LLC.
Saad, L., Jani, J., & Rahmat, A. (2017). The Performance Standard of Students Learning based on Assessment Instruments in Physical Education. *International Journal of Academic Research in Business and Social Sciences*, *7*(4), 927–935.
Shorten, A., & Smith, J. (2017). Mixed methods research: Expanding the evidence base. *Evidence-Based Nursing*, *20*(3), 74–75. https://doi.org/10.1136/eb-2017-102699
Sidhu, G.K., Kaur, S., & Chi, L.J. (2018). CEFR-aligned school-based assessment in the Malaysian primary ESL classroom. *Indonesian Journal of Applied Linguistics*, *8*(2), 452–463.
Singh, P., Supramaniam, K., & Teoh, S.H. (2017). Re-assess or Risk the Slow Death of School-Based Assessment. *Pertanika Journals of Social Sciences and Humanities*, *25*(S), 71–80.
Stiggins, R.J. (1991). Assessment Literacy. *Phi Delta Kappan*, *72*, 534–539.
Suah, S.L., & Ong, S.L. (2012). Investigating Assessment Practices of In-service Teachers. *International Online*.
Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, *9*(1), 1–14.
Tabachnick, B.G., & Fidell, L.S. (2001). *Using Multivariate Analysis*. Northridge: California State University, Harper Collins College Publishers.
Tajularipin, S., Fauzi, A., & Sulaiman, S. (2015). Curriculum Change in English Language Curriculum Advocates Higher Order Thinking Skills and Standards-Based Assessments in Malaysian Primary Schools. *Mediterranean Journal of Social Sciences*, *6*(2).
Talib, R., & Naim, H.A. (2012, November). *Assessment literacy: A catalyst to the success of school-based assessment in Malaysian schools* [Paper Presentation]. International Conference on Science Technology and Social Sciences, Kuantan, Pahang.
Talib, R., Kamsah, N.Z., Naim, H.A., & Latif, A.A. (2014). From principle to practice: Assessment for learning in Malaysian school-based assessment classroom. *International Journal of Social Sciences & Education*, *4*(4), 840–857.
Varatharaj, R.K., Abdullah, A.G.K., & Ismail, A. (2014). Assessment Practices among Malaysian Cluster School Teachers. *International Journal of Research in Management and Business Studies, 1*(3), 23–27.
Veloo, A., Ramli, R., & Khalid, R. (2016). Assessment Practices among English Teachers in Malaysian Secondary Schools. *International Journal for Infonomics*, *9*(4), 1220–1227. https://doi.org/10.20533/iji.1742.4712.2016.0149

Vogt, K., & Tsagari, D. (2014) Assessment Literacy of Foreign Language Teachers: Findings of a European Study. *Language Assessment Quarterly*, *11*(4), 374–402.
Watkins, M.W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219–246.
Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, *8*(3).
Xie, Q., & Tan, S. (2019). Preparing Primary English Teachers in Hong Kong: Focusing on Language Assessment Literacy. *The Journal of Asia TEFL*, *16*(2), 653–673.
Xu, Y., & Brown, G.T.L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, *6*(1), 133–157.
Yahaya, T., Idris, K., Suandi, T., & Ismail, I. (2018). Adapting instruments and modifying statements: The confirmation method for the inventory and model for information-sharing behaviour using social media. *Management Science Letters*, *8*(5), 271–282.