

## A review of teachers' assessment literacy proficiency measures

Geng Juanjuan<sup>1,2</sup>, Nurahimah Mohd Yusoff<sup>1</sup>

<sup>1</sup>School of Education, Universiti Utara Malaysia, Sintok, Malaysia

<sup>2</sup>School of Foreign Languages, Xianyang Normal University, Xianyang City, China

---

### Article Info

#### Article history:

Received Jun 11, 2021

Revised May 29, 2022

Accepted Jun 30, 2022

#### Keywords:

Assessment literacy

Conceptualization

Literature review

Measures

---

### ABSTRACT

At present, there are several literature reviews on teachers' assessment literacy (AL). However, little attention has been paid to a systematic analysis of AL measures. The review on AL measures helps to evaluate the available tools and provide recommendations for developing AL measures in further exploration. Hence, the purpose of the current literature review is to focus on analyzing the measures of teachers' AL proficiency from 1991 to 2020. Across the contexts, 28 quantitative tools made up of questionnaires or tests were collected. Among them, 10 were designed to investigate overall AL proficiency and 18 were constructed to examine the sub-categories of AL proficiency among teachers. Each instrument was analyzed from the format, guiding framework, item characteristics, targeting population, and psychometric quality. Results showed that the current AL measures were derived more from instrumental than social-cultural conceptualization, examined more on sub-categories than the overall proficiency level, targeted more at in-service teachers than pre-service teachers, and focused more on elementary and middle education than pre- or higher education. Further studies are welcomed from four aspects: construct social-culturally, test generalizability, target at specific groups, and cover various types of AL.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

### Corresponding Author:

Geng Juanjuan

Department of Foreign Languages, Xianyang Normal University

Wenlin Road, Weicheng District, Xianyang City, Shaanxi Province, China

Email: mylovegjj123@gmail.com

---

## 1. INTRODUCTION

Assessment literacy (AL) is a crucial element of teachers' professional development [1], [2] and it is necessary for teachers to have an adequate level in AL to realize the full potential of assessment in the teaching and learning process [3]–[6]. Regrettably, the existing studies suggest that teachers' AL is insufficient to meet the assessment requirements advocated in assessment policies or standards, even the excellent teachers feel challenged to prepare assessment [7]–[9]. The prerequisite to further improvement in AL is the accurate measure of teachers' current AL proficiency level to understand their performance in each dimension in AL.

The existing literature reviews mainly focus on reviewing all the studies conducted around the theme of AL [10], [11]. However, only a very few reviews confine the scope to teachers' AL measures. Previous researchers [12] analyzed eight prominent AL measures constructed after 1990 and 15 assessment standards. After the thematic analysis and close examination, they found that AL measures were less responsive to the updated assessment standards and were still framed in the early conceptualization of AL. Gotch and French [13] were merely interested in the psychometric evidence by collecting 36 measures in various formats of questionnaires, objective tests of AL, and rubrics. After their systematic analysis of the psychometric evidence, the results indicated that psychometric work in supporting AL measures was weak.

Overall, the reviews on AL measures are relatively small in number and need to be incorporated with updated research. The available reviews analyze AL measures merely from perspectives of assessment standards and its own psychometric quality. Further review is still warranted to analyze AL measures from various aspects, including but is not limited to, guiding framework, item characteristics, statistical evidence of quality. On top of that, the two reviews were conducted nearly five years ago, more latest measures need to be included and analyzed to provide a holistic picture of AL measures. Thus, to address the paucity, this study aims to review measures on teachers' AL proficiency by conducting a rather comprehensive search of the literature in order to shed light on the future construction of the instruments.

The initial conceptualizations of AL are described from an instrumental perspective, which is rooted in the acquisition of assessment knowledge and skills [14]. AL is outlined in Standards for Teacher Competence in Educational Assessment of Students issued by the American Federation of Teachers, National Council on Measurement in Education, and National Education Association [15] (hereafter 1990 Standards). The 1990 Standards have made a documentable contribution to the AL field [16] and have served as a primary guiding framework for the construction of several subsequent AL measures. The standards cover the following seven dimensions, teachers should be skilled in: i) Choosing assessment methods appropriate for instructional decisions; ii) Developing assessment methods appropriate for instructional decisions; iii) Administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods; iv) Using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement; v) Developing valid pupil grading procedures that use pupil assessments; vi) Communicating assessment results to students, parents, other lay audiences, and other educators; and vii) Recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

The subsequent broadened conceptualizations attempt to understand AL as a multi-dimensional framework. Over time, there has been a shift encompassing a set of knowledge, skills, and conceptions related to assessment. According to Inbar-Lourie [17], AL is conceived as a three-component model including teachers' theoretical knowledge about assessment (what), the performance of assessment practices (how), and the purpose of assessment (why). Similarly, Lam [18] stated that AL framework consists of teacher conception (belief system), knowledge, and practices about assessment.

From the social-cultural perspective, AL is reconceptualized as a social practice instead of the mere accumulation of assessment knowledge and skills. AL "involves teachers articulating and negotiating classroom and cultural knowledge with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students" [19]. According to Looney *et al.* [20], AL is not only about what teachers know, do, and feel about assessment, but also who they are. Likewise, teachers' identity construction as assessors is also emphasized by Xu and Brown [21], who provide a detailed profile of highly assessment literate teachers: "who constantly reflect on their assessment practice, participate in professional activities concerning assessment in communities, engage in professional conversations about assessment, self-interrogate their conceptions of assessment, and seek for resources to gain a renewed understanding of assessment and their own roles as assessors."

## 2. RESEARCH METHOD

The literature searching and sorting were carried out in two phases. In the literature searching phase, a systematic search of published articles from 1991 to 2020 was conducted by exploring across the online databases, such as ERIC, JSTOR, Psycho Info, ProQuest, EBSCOhost, and Google Scholar by using the terms as assessment literacy, assessment conception/belief, assessment knowledge, assessment practice, assessment confidence, and the combination of measure, instrument, tool, inventory, and validation. The search was limited to the published English language papers that examined AL among pre-service teachers or in-service teachers across the world.

The literature was sorted and organized for keeping focused on the quantitative measurement tools in AL. Thus, some research was excluded according to the following sorting criteria: within a large number of returning articles, non-empirical studies were excluded. The replication research adopting or primarily adapting prior questionnaires was also excluded for its limited originality. Besides, the qualitative studies using interviews, open-ended surveys, or classroom observation were also excluded for beyond the scope of the current literature review. Through this process, a total of 28 instruments were identified.

## 3. RESULTS AND DISCUSSION

The 28 instruments are designed to examine either overall AL proficiency levels (n=10) or sub-categories in AL proficiency levels among teachers (n=18). A detailed analysis of these instruments is based

on: i) Item characteristics (e.g., number of the items, Likert-type items, or scenario-based items); ii) The guiding framework (e.g., the authoritative policies, standards, or literature used for the blueprint); and iii) The reported psychometric properties (e.g., reliability).

### 3.1. Measures of the overall AL proficiency

Among instruments measuring the teachers' overall AL proficiency (n=10), four are in the format of tests, five are in the format of Likert-scale questionnaires, and one is in the combination of test and questionnaire. In each sub-category, the instruments are listed in chronological order in Table 1. In the first sub-category, four tests are identified: Teacher Assessment Literacy Questionnaire (TALQ) [22], Assessment Literacy Inventory (ALI) [23], Assessment Literacy Test (ALT) [24], and Assessment Tasks (AT) [25]. The second type is the Likert-scale questionnaire. Five questionnaires are listed: Assessment Literacy Survey (ALS) [26], Questionnaire for Assessment Literacy (QAL) [27], Approaches to Classroom Assessment Instrument (ACAI) [28], English Language Teachers' Assessment Literacy (ELTAL) [29], and Performance Assessment Literacy (PAL) [1]. In the last category, the survey with the inclusion of both test and questionnaire is labeled as the Questionnaire for Measurement Literacy (QML) [30].

Table 1. The list of overall AL proficiency measures

Instrument (source)		Item characteristics	Framework	Respondents	Psychometric properties
Test	TALQ [22]	35 items	1990 Standards	555 elementary and middle school teachers in the USA	KR <sub>20</sub> =.54
	ALI [23]	35 classroom-based scenario items	1990 Standards	152 (phase 1) + 249 (phase 2) pre-service teachers in USA	KR <sub>20</sub> =.74
	ALT [24]	45 multiple-choice items	1990 Standards, Syllabus, and Competence Model	465 secondary school teachers in Malaysia	KR <sub>20</sub> =.85
	AT [25]	10 scenario-based tasks	Adopted from Folse <i>et al.</i> [31]	39 middle school English teachers in China	Cronbach $\alpha$ =.828
Questionnaire	ALS [26]	7 open- and closed-ended items	Earl [32]; Earl and Katz [33]	69 pre-service teachers in Canada	Not mentioned
	QAL [27]	50-item	Literature	310 university teachers in Iran	Cronbach $\alpha$ =.97
	ACAI [28]	53-item	Classroom Assessment Standards	404 pre-service and in-service teachers in Canada	Cronbach $\alpha$ ranged from .74 to .92 among the sub-sections
	ELTAL [29]	25-item	Literature and interview	150 teachers in institutes or universities in Iran	Cronbach $\alpha$ =.78
	PAL [1]	27-item	Quality Performance Assessment Framework [34]	1,080 teachers in England	Cronbach $\alpha$ ranged from .793 to .953 among the sub-scales; CFI=0.91, AIC=18204, BIC=18517, RMSEA= 0.09, and SRMR= 0.05
Test + Questionnaire	QML [30]	60-item (30-item test and 30-item questionnaire)	Literature	96 elementary and secondary teachers in USA	Cronbach $\alpha$ = .60 for measurement literacy and $\alpha$ =.93 for use of the technique

#### 3.1.1. Test

Plake *et al.*'s [22] test might be the earliest attempt to examine AL proficiency among teachers after the issue of the 1990 Standards. Their 35-item test was developed to measure the seven competence areas mentioned in the Standards with five items in each dimension to identify the specific knowledge, skills, and abilities that were meaningful to teachers' realistic assessment practices. Their test was validated through a rigorous process by a pilot study of 900 educational professionals and a peer review by assessment experts to ensure content validity. Concerning the reliability, the internal consistency reliability coefficient for the entire test was reported to be at a low level (KR<sub>20</sub>=.54) may be due to the mere five items in each standard.

The poor psychometric quality of previous study [22] was noted by Mertler and Campbell [23], who also criticized the test for being lengthy, difficult to read, and decontextualized. Hence, they redeveloped an instrument based on Plake *et al.* [22] and provided the psychometric qualities among the pre-service teachers. The revised version was embedded within five classroom-based scenarios, where teachers were faced with

various assessment-related decisions. Each scenario showed a brief classroom situation followed by seven multiple-choice items which were designed in alignment with the 1990 Standards. The satisfactory reliability coefficient ( $\alpha=.74$ ), fairly good item difficulty, and desirable item discrimination were reported. Thus, the instrument was tested to be an effective measure of teachers' AL with acceptable psychometric properties.

Talib, Ghafar, and Naim [24] incorporated AL tests with the local contexts by referring to the Malaysian Teacher Training Syllabus, 1990 Standards, and Stiggins Competency Assessment Model. The test included 45 multiple-choice items to cover the following five dimensions: assessment concepts, measurement methods, testing, scoring & grading, and statistics & reporting. The reliability of the test showed a high level of internal consistency ( $KR20=.85$ ) among 465 secondary school teachers in Malaysia. However, evidence of validity was not mentioned in the research.

The assessment test [25] adapted from previous study [31] was specifically designed for 39 middle school English teachers in China before the language assessment course. The instrument consisted of 10 tasks covering both the language-specific testing knowledge (the four skills: listening, speaking, reading, and writing) and the overall test-related knowledge (test preparation process, the statistic interpretation). The reliability coefficient Cronbach  $\alpha$  was estimated to be .828, showing a desirable level of internal consistency.

### 3.1.2. Questionnaire

The second sub-category is the Likert-type questionnaire, the most commonly used instrument for its convenience to examine the self-report AL proficiency among teachers. Volante and Fazio [26] used previous assessment framework [32], [33] which provided three conceptions of assessment: assessment *of*, *for*, and *as* learning to measure AL among Canadian pre-service teachers. The survey was composed of seven items in the form of open- and closed-ended questions in four areas: the self-described level of AL, perception of assessment purposes, frequency of using diverse assessment methods, and improvement suggestions for further training and the pre-service education program. Regrettably, the psychometric information of the questionnaire was not mentioned in their research.

Focusing on in-service teachers, Esfandiari and Nouri's [27] questionnaire of the self-rate AL proficiency consisted of 50 items on a five-point Likert-type scale. The questionnaire covered various aspects of AL ranging from knowledge of statistics and testing to interpretation of the test results. The participants needed to decide where they fit in the scale from 1= "not at all" to 5= "the highest degree". The development of the questionnaire was based on the literature of AL and the textbooks written by renowned scholars in the field of assessment. The questionnaire was checked for clarity, comprehensibility, relevance, and wording in the first step, and then went through the pilot study of 310 participating teachers. It showed a high level of reliability with Cronbach  $\alpha=.97$ .

To meet the call for developing an instrument reflecting the recent assessment standards and considering the rapid increase of accountability, DeLuca *et al.* [28] designed Approaches to Classroom Assessment Instrument (ACAI) framed by the latest revised Classroom Assessment Standards issued by Joint Committee on Standards for Educational Evaluation in 2015. This measure included three sections: Approaches to classroom assessment, perceived skill in classroom assessment, and assessment professional learning priorities and preferences. Part one of the ACAI adopting a scenario-based item format ( $n=20$ ) intended to generate teacher's overall approach to assessment in the four aspects: assessment purposes, assessment processes, assessment fairness, and measurement theory. Part two was comprised of 12 items to determine the self-assessed skill level on a five-point scale (1= novice, 5= expert). Part three was designed to examine teachers' professional learning priorities ( $n=10$ ) and preferences in assessment ( $n=11$ ). The validation of the instrument went through a multi-step process. The expert-panel review method (i.e., assessment expert panel + practitioner expert panel) was adopted to test the content validity for ACAI. A pilot testing among both in-service and pre-service teachers was implemented to collect evidence on internal structure. The internal consistency estimates for the subscales ranged from .74 to .92. Thus, ACAI was proven to be a reliable instrument to support professional development in AL.

More recently, based on the review of related literature and interview with professional, Nikmard and Zenouzagh [29] developed and validated a 30-item AL questionnaire targeting English language teachers only. Although it was labeled as an AL questionnaire, it focused more on examining the assessment knowledge of English teachers. The instrument covered four themes in assessment: validity, reliability, interpretability, and efficiency. Good reliability of the questionnaire ( $\alpha=.78$ ) was reported from a pilot study of 150 English language teachers. The instrument was also subjected to factor analysis, which revealed that the presence and classification of the four themes were reasonable in statistics.

Rather than measuring AL in general, a study [1] explored and confirmed a quantitative measure for teachers to self-evaluate their proficiency level on performance AL. The 27-item instrument was framed by five dimensions identified in the Quality Performance Assessment Framework by Brown and Mednick [34] as essential elements of performance assessment practices: valid assessment design, reliable scoring,

assessment data analysis, fairness in assessment, and student voice and choice. The items were answered by using a six-point Likert-type scale: 0=not confident, 1=somewhat confident, 2=moderately confident, 3=confident, 4=extremely confident, 5=completely confident. Validation of the questionnaire was conducted through a sample of 1,080 teachers by exploratory and confirmatory factor analysis (CFA), which indicated the five-factor hierarchical model was plausible to measure performance AL. It was recommended by the researcher to be added into the toolbox of practitioners to self-evaluate their confidence in performance AL.

### 3.1.3. Test and questionnaire

The third type of AL survey was characterized by the integration of assessment tests and Likert-type questionnaires, rather than relying on one single data collection method. With a reference to the AL literature, Daniel and King [30] developed a 60-item survey composed of two sections: a test (n=30) for measurement literacy and a Likert-type questionnaire (n=30) for the use of assessment techniques among elementary and middle school teachers. The test took the form of true-false choice and the questionnaire was in the form of a five-point Likert scale ranging from the least frequency of use to the greatest frequency of use. The reliability coefficients drawn from a sample of over 90 teachers were .60 and .93 for the two sub-sections respectively.

### 3.1.4. Discussion

Overall, concerning the total number of AL proficiency measures (n=10) available in the literature, the number of AL instruments is relatively small. Among the limited measures, three instruments are reported to be explicitly framed in 1990 Standards [22]–[24], which was a little outdated for not taking formative assessment into account [16]. Thus, the recently updated assessment standards or documents are suggested to be adopted as a guiding framework for the upcoming AL proficiency instruments in the future. Besides, regarding the contexts of research, the instruments are developed in various contexts, including the USA (n=3), Iran (n=2), Canada (n=2), China (n=1), England (n=1), and Malaysia (n=1). AL measures that fit into other local contexts also need more attention due to AL's context-sensitive nature [21].

The target population of the measures is primarily concentrated on the in-service teachers, and only three of them include pre-service teachers as participants [23], [26], [28]. Furthermore, teachers in higher education receive little attention with only two instruments designed to examine their AL proficiency level [27], [29]. More regrettably, teachers in pre-school education are still under-explored and more attention is deserved to develop the corresponding AL measures in accordance with the different assessment requirements.

Lastly, except for the measure by Volante and Fazio [26] who did not mention the reliability, the reliability coefficient of internal consistency of the remaining instruments varies from .54 to .97 with a mean of .80. It is revealing a satisfactorily level of internal consistency of the instrument. However, the validation process of some instruments is not presented in a detailed and transparent way, especially evidence on validity is not clearly mentioned in the research.

## 3.2. Measures of the sub-categories of AL proficiency level

Apart from the measures developed to examine the overall AL proficiency level, a large majority intend to focus on the sub-categories of teachers' AL proficiency (n=18). It is including assessment knowledge (n=1), assessment self-efficacy (n=2), assessment conception (n=5), assessment practices (n=7), and both assessment conception and practices (n=3). The detail is presented in Table 2.

### 3.2.1. Assessment knowledge

Only one instrument is available to explore the assessment knowledge and is designed especially for English as foreign language (EFL) teachers. Language Assessment Knowledge Scale (LAKS) developed by Ölmezer-Öztürk and Aydın [35] was composed of four constructs (assessing listening, speaking, reading, and writing) with 15 items for each and 60 items in total. The answers to each item should be chosen from three options (true, false, or not know).

The development of the instrument went through several stages. An item pool was created after a thorough review of related literature on AL and checked by experts for the comprehensibility and compatibility of each item. Later, feedbacks and opinions from practitioners in the field and experts in language testing and assessment were collected to ensure content validity. Also, a pilot study was implemented to test the scale. LAKS was administered to more than 500 university EFL teachers in Turkey for a tough validation process. The factor analysis confirmed the compatibility of items with the four constructs and the compatibility of these constructs with language assessment knowledge. The reliability analysis of the entire scale was estimated to be at a considerably high level with Cronbach  $\alpha$ =.91. Therefore, the statistical evidence, revealing LAKS had a good modal-data fit and a high level of internal consistency,

supported that LAKS could be utilized as a valid and reliable measure to examine university EFL teachers' language assessment knowledge.

### 3.2.2. Assessment self-efficacy

There exist two instruments intending to examine self-efficacy in engaging in assessment-related tasks, one for in-service teachers and the other for pre-service teachers. In order to differentiate in-service teachers' efficacy concerning the alignment of classroom-based assessment with state learning standards, Wolfe *et al.* [36] designed the 41-item Teacher Assessment Efficacy Scale (TAES), which covered six dimensions in assessment efficacy: confidence, impact, utilization, utility, experiences, and students. Each item was presented as a statement to which teachers needed to indicate their levels of agreement on a five-point Likert scale ranging from strongly disagree to strongly agree. The statistical evidence from a sample of over 600 primary and secondary school teachers was presented from four aspects: content, substantive, structural, generalizability, and validity. The analysis showed the six-dimensional classifications explained most per estimated parameter, the internal consistencies of the scale were reported to be at a high level (Cronbach  $\alpha$  ranging from .77 to .94 among the sub-scales), and the item fit indices were reasonable for most items. The instrument was concluded as an effective indicator of the teacher's assessment efficacy.

Instead of measuring the assessment efficacy of in-service teachers, another instrument targeting to measure the Confidence Level in Assessment Knowledge (CLAK) among pre-service teachers was developed by DeLuca and Klinger [37]. The guiding framework of the instrument was a series of assessment standards and policies at a national or provincial level. The questionnaire was framed in the two-fold model pairing assessment of learning with assessment for learning from three knowledge domains: assessment practice, theory, and philosophy. To be specific, the CLAK was composed of three sections: five checklist items in section A to collect demographic information of the participants, section B consisting of 45 scaled items to identify the confidence level by using a five-point Likert scale from 1="not very confident" to 5="very confident" and three items for ranking the priority, and one checklist item in section C to identify the primary context in which they learned the assessment items. Factor analysis was conducted to confirm the assessment construct. Also, reliability analysis of the internal consistency revealed a relatively high degree of response consistency; all Cronbach  $\alpha$  was above .82 with two exceptions.

### 3.2.3. Assessment conception

Teacher conception of assessment is used here as an umbrella term to refer to an organized belief system. It embraces all that a teacher thinks about the nature and purpose of assessment, encompassing beliefs, propositions, attitudes, and preferences [53]. To investigate assessment conception five tools are constructed, among which two for pre-service teachers and three for in-service teachers.

Pre-service teachers' attitudes towards the use of alternative assessment were examined by Allen and Flippo [38] through a five-point Likert-type Alternative Assessment Measure (AAM). The participants were asked to respond to 20 statements spread among three sub-scales (self-evaluation, peer evaluation, and instructor modeling). All these three sub-scales were tested to have a satisfactory internal consistency with a reliability coefficient of .83, .75, and .78 respectively.

With a similar focus on pre-service teachers, Hildén and Fröjdendahl's [42] Student Teachers' Assessment Literacy (STAL) questionnaire was aimed to explore the assessment conceptions of pre-service foreign language teachers in Finland. With reference to AL literature and national curriculum standards in Finland, this localized instrument entailed three sections: definition of core constructs (n=5), assessment practices on a five-point Likert scale from "always" to "never" (n=29), and conceptions of assessment (n=38). The factor analysis of the questionnaire confirmed the classification of the items and the reliability coefficient was estimated to be .92 for the pre-test among 77 participants and .93 for the post-test among 65 participants, indicating a high internal consistency of the questionnaire. However, as mentioned by the researchers, the validity of the instrument was subjected to criticism. The items needed to be cultivated by consultancy with the target population and other stakeholders (e.g., assessment experts, teacher educators).

Concerning the in-service teachers, Brown's [39] 50-item Teachers' Conceptions of Assessment (CoA-III) questionnaire was aimed to understand teachers' conceptions of assessment through their agreement or disagreement. It contained four purposes of assessment: improvement of teaching and learning, school accountability, student accountability, and treating assessment as irrelevant. Each statement needed to be responded to by using a six-point Likert scale from "strongly disagree" to "strongly agree". As to the overall fit of all the factors in the model, structural equation modeling (SEM) was run to determine the degree to which data fit the theoretically expected relationship. The Root Mean Square Error of Approximation (RMSEA) was estimated to be .58, which was below .80 indicating a well-fitting measurement model.

Brown [40] shortened 50-item CoA-III into Teachers' Conceptions of Assessment-III Abridged (TCoA-IIIa) inventory of 27 items because a larger number of items might be undesirable especially in a restricted responding time. The study reported the psychometric characteristics of the shortened version whether it measured the same theoretical framework with the full scale in a robust way. The statistical evidence showed that the factors had very similar inter-factor correlation values and direction as the full scale. Additionally, the fit characteristics was good for primary teachers in both New Zealand ( $\chi^2=841.02$ ;  $df=311$ ;  $RMSEA=.057$ ;  $TLI=.87$ ) and Queensland ( $\chi^2=1492.61$ ;  $df=311$ ;  $RMSEA=.074$ ;  $TLI=.80$ ) Thus, the abridged inventory was concluded to be an efficient and valid measure of teachers' assessment conceptions.

Another instrument was interested in exploring in-service teachers' beliefs about assessment. Thomas [41] formulated a seven-item questionnaire labeled as Teachers' Beliefs about Classroom Assessment (TBCA) with the help of a related AL literature review. Each belief was presented as a statement with a five-point Likert scale from "strongly disagree" to "strongly agree". A sample of 123 middle and secondary school teachers in Pakistan responded to the questionnaire, but the validation information about the instrument was not mentioned in the research. Therefore, to what extent the instrument was valid and reliable still needed further study and evidence.

### 3.2.4. Assessment practice

There are seven instruments designed primarily to examine the assessment practices among in-service teachers by self-report questionnaires. Among them, five are for teachers from primary and middle schools, one is for university teachers, and one is for K-12 teachers. However, little attention was paid to investigate the assessment practices of pre-service teachers may be due to the limited assessment practices initiated by them.

Zhang and Burry-Stock [43] designed the Assessment Practices Inventory (API) with 67 items covering a broad range of classroom-based assessment practices in seven categories ranging from developing and administering tests to communicating tests results. Each item was required to be responded on a five-Likert scale from "not at all skilled" to "highly skilled". The verification of the construct reliability of the instrument was conducted through factor analysis and the Rasch model, which indicated the items defined the theoretical construct of classroom-based assessment practices with the distribution of items logits from -0.89 to 1.31. The reliability of the API was .97, indicating a high consistency in structure.

Based on primary and middle school teachers' perception of their assessment practices, the 21-item questionnaire developed by Bol *et al.* [44] intending to examine teachers using traditional and alternative assessment methods in three scales: frequency, preparation, and confidence. As to the quality of the instrument, the validity was tested by factor analysis, which confirmed the classification of traditional and alternative assessment. The reliability of some sub-scales was reported to be relatively low (e.g., Cronbach  $\alpha=.49$ , .51, .66) among a sample of over 800 teachers, possibly due to the small number of items in each scale. However, constrained by the response burden of the participants, the developers did not add more items in each scale in the questionnaire.

Suah and Ong's [48] 57-item Teachers Assessment Practice Inventory (TAPI) covered a wide range of assessment activities to ask about teachers' frequency of assessment practices on a five-point Likert scale. The instrument was proven to be satisfactory. It was supported in content validation from experts and statistical evidence for validity from CFA.

Confined to examine how the assessment reform was implemented by primary teachers only, Choi's [49] Implementation of Assessment Format Survey (IAFS) was released to collect data about teachers' assessment practices from three aspects: formats of assessment, cognitive demand, and assessment purposes. The survey was validated by factor analysis and the reliability of internal consistency was shown to be from .65 to .85, indicating the measure was reliable. However, more detailed information about the survey was not mentioned in the paper, for example, the number of items in total and in each sub-scale, the type of Likert scale, and the development of the survey.

Exclusively designed for EFL teachers, the instrument by Inbar-Lourie and Donitsa-Schmidt [47] focused on primary and middle school EFL teachers' classroom assessment practices, which were examined from four perspectives (i.e., technological, cultural, political, and postmodern) in teachers' alternative assessment practices. The main part of the questionnaire was composed of 65 items in a five-point Likert scale from "completely disagree" to "completely agree" to tap EFL teachers' perceptions. The detailed development and validation process of the instrument demonstrated high reliability among the sub-scales and the justification of the existence of the four perspectives. Also targeting at EFL and English as a second language (ESL) teachers but at the university level, Cheng *et al.* [46] developed Classroom Assessment Practices for EFL/ESL Instructors in higher education. They collected their purposes, methods, and procedures of classroom assessment practices through 51 multiple-choice items. However, any validation information of the instrument was not mentioned in their research.

Ohio Teacher Assessment Practices Survey (OTAPS) [45] extended the target population to include kindergarten teachers. With the focus on both traditional and alternative assessment practices, it consisted of 47 scaled and open-ended items in total to address the frequency and confidence level of using various assessment techniques by K-12 teachers. Regrettably, any information related to the validity and reliability of the instrument was not reported in the paper, so it was hard to judge the quality of the questionnaire.

### 3.2.5. Assessment conception and practice

Three measures were intended to investigate both conceptions and practices of assessment among teachers. One of the measures was targeted at teachers in general education. The Values and Practice of Classroom Assessment (VPCA) was constructed to investigate how teachers valued various classroom assessment practices and how congruent with these values they perceived their practices to be [50]. The 30 items in the questionnaire needed to be responded to twice: the first response concerning the assessment practices in four-point Likert scale from “never true” to “mostly true” and then a second response concerning the assessment values also in four-point Likert scale from “not at all important” to “important” but with an additional option offered as “bad practice”. The dual scale enabled the researchers to address the gap between values and current practices in assessment from the respondents. A pilot study and factor analysis were conducted to test the validity and reliability of the questionnaire, which was shown to be reliable and valid with the evidence that reliability coefficients among the subscales ranged from .54 to .76, and the factor analysis confirmed the underlying dimensions. The list of measures focusing on sub-categories of AL proficiency is presented in Table 2.

Table 2. The list of measures focusing on sub-categories of AL proficiency

Sub-categories	Instrument (source)	Item characteristics	Respondents	Psychometric properties
Assessment knowledge	LAKS [35]	60 true or false items	542 university EFL teachers in Turkey	Cronbach $\alpha$ =.91
Assessment self-efficacy	TAES [36]	41 items on a 5-point Likert scale	583 primary and secondary teachers in the USA	Cronbach $\alpha$ ranged from .77 to .94 among the sub-sections
	CLAK [37]	49 items of checklist, scale, and ranking	288 pre-service teachers in Canada	Cronbach $\alpha$ ranged from .55 to .93 among the sub-sections
Assessment conception	AAM [38]	20 items on a 5-point Likert scale	29 pre-service teachers in the USA	Cronbach $\alpha$ ranged from .75 to .83 among the sub-sections
	CoA-III [39]	50 items on a 6-point Likert scale	525 primary teachers in New Zealand	$X^2=3217.68$ ; $df=1162$ , $RMSEA=0.058$ ; $TLI=0.967$
	TCoA-IIIa [40]	27 items on a 6-point Likert scale	525 primary teachers in New Zealand + 692 primary teachers in Australia	New Zealand ( $\chi^2=841.02$ ; $df=311$ ; $RMSEA=0.57$ ; $TLI=0.87$ ) and Australia ( $\chi^2=1492.61$ ; $df=311$ ; $RMSEA=0.74$ ; $TLI=0.80$ )
	TBCA [41]	7 items on a 5-point Likert scale	123 middle and secondary school teachers in Pakistan	Not mentioned
	STAL [42]	72 items on a 5-point Likert scale	77 (pre-test) + 65 (post-test) pre-service teachers in Finland	Cronbach $\alpha$ =.92 for pre-test, Cronbach $\alpha$ =.93 for post-test
Assessment practice	API [43]	67 items on a 5-point Likert scale	311 primary and middle school teachers in the USA	Cronbach $\alpha$ =.97
	Survey [44]	21 items on a 5-point Likert scale	893 primary and middle school teachers in the USA	Cronbach $\alpha$ ranged from .49 to .76 among the sub-sections
	OTAS [45]	47 items on the scale and open-ended questions	625 K-12 teachers in the USA	Not mentioned
	CAP [46]	51 items in multiple choices	263 EFL/ESL teachers in Canada and China	Not mentioned
	APQ [47]	65 items in open-ended questions and a 5-point Likert scale	113 primary and middle school EFL teachers in the USA	Cronbach $\alpha$ ranged from .61 to .89 among the sub-sections
	TAPI [48]	57 items on a 5-point Likert scale	406 primary and secondary teachers in Malaysia	CFA: NFI, CFI, IFI, GFI, and AGFI exceeded .90.
	IAFS [49]	Not mentioned	700 elementary teachers in South Korea	Cronbach $\alpha$ ranged from .65 to .85 among the sub-sections
Assessment conception + practice	VPCA [50]	30 items on a 4-point Likert scale	558 teachers in England	Cronbach $\alpha$ ranged from .54 to .76 among the sub-sections
	BPMA [51]	48 items on a 4-point Likert scale, multiple choices, and open-ended questions	520 elementary EFL teachers in Taiwan in China	Cronbach $\alpha$ =.85
	CAPSQ-LT [52]	35 items on a 5-point Likert scale	115 language teachers in Japan and the Philippines	Cronbach $\alpha$ =.964

Two instruments were specially designed for language teachers only. Gonzales and Aliponga's [52] Classroom Assessment Preferences Survey Questionnaire for Language Teachers (CAPSQ-LT) was designed to explore their assessment beliefs and practices. The instrument consisted of 35 items on a five-point Likert scale from "very rarely" to "always" to inquire about the frequency of practicing assessment methods and assessment beliefs in five sections: assessment of learning, assessment for learning, assessment as learning, assessment for instruction, and assessment to inform. The reliability of the entire questionnaire was reported to be highly consistent ( $\alpha=.964$ ) and the five factors could explain nearly 65% of the variance measured in the instrument. Another survey was aimed to investigate EFL teachers' beliefs and practices of multiple assessments. Chan's [51] self-report questionnaire labeled as Beliefs and Practice of Multiple Assessment (BPMA). It was comprised of 48 items to understand multiple assessments from three dimensions: perceptions, beliefs, and practices. The survey was reviewed by the professors to ensure validity and undergone a pilot study to guarantee reliability. The Cronbach's  $\alpha$  was .85, indicating a satisfactory internal consistency of the questionnaire.

#### 4. CONCLUSION

Further efforts in developing AL measures are still warranted and more exploration is favored from four aspects. Firstly, with regard to the framework of the measures, the majority of the reviewed measures are developed within the instrumental conceptualization of AL from three dimensions of assessment knowledge, skills, and concepts. However, few studies construct AL measures from the social-cultural perspective, taking AL as a social practice instead of a repertoire of assessment-related knowledge and skills. This indicates a need to construct AL measures with the inclusion of more dimensions, for example, teachers' identity construction as assessors because by responding to the updated framework of AL, the measures' construct validity will be promoted. Secondly, concerning the generalizability, due to some AL measures designed especially for the purpose of the study, what is still unknown is to what extent are these AL measures valid and reliable to be generalized to other target groups in different contexts with varying assessment policies and priorities. Thus, applying or adapting these measures into different assessment contexts will further provide geographic validity evidence and test the generalizability of the tool.

Thirdly, the target group primarily focuses on in-service teachers in general education in primary and middle schools. What is still lacking is AL measures designed for teachers in a specific subject (physics, mathematics, arts), in different professional careers (novice, proficient, expert), in different education types (special education, adult education), and in indifferent educational level (college/university, kindergarten). The focal emphasis on the variance of the context and teachers' personal experiences will provide a more accurate and detailed picture of AL examined in the tool. Moreover, among the existing AL measures, only a few are developed for pre-service teachers. AL required to be possessed by pre-service teachers may be different from AL suggested being acquired by in-service teachers. In other words, constructs and degree of mastery level of AL may differ within the two teacher populations. Hence, developing AL measures with pre-service teachers as the target group is still deficient. Lastly, as to the content of the measures, the majority of AL measures aim to examine either the overall proficiency level or the proficiency level in some dimensions. However, the measures focusing on different types of AL are still under-explored, such as the measures for examining teachers' performance AL, authentic AL, and classroom AL.

#### REFERENCES

- [1] M. P. Kelly, R. Feistman, E. Dodge, A. St. Rose, and J. Littenberg-Tobias, "Exploring the dimensionality of self-perceived performance assessment literacy (PAL)," *Educational Assessment, Evaluation and Accountability*, vol. 32, no. 4, pp. 499–517, Nov. 2020, doi: 10.1007/s11092-020-09343-7.
- [2] M. Kyttälä, P. M. Björn, M. Rantamäki, V. Närhi, and M. Aro, "Assessment conception patterns of Finnish pre-service special needs teachers: the contribution of prior studies and teaching experience," *European Journal of Special Needs Education*, vol. 37, no. 1, pp. 131–145, Jan. 2022, doi: 10.1080/08856257.2020.1853972.
- [3] J. B. Napanoy and M. K. Peckley, "Assessment Literacy of Public Elementary School Teachers in the Indigenous Communities in Northern Philippines," *Universal Journal of Educational Research*, vol. 8, no. 11B, pp. 5693–5703, Nov. 2020, doi: 10.13189/ujer.2020.082203.
- [4] R. Watmani, H. Asadollahfam, and B. Behin, "Demystifying language assessment literacy among high school teachers of english as a foreign language in iran: Implications for teacher education reforms," *International Journal of Language Testing*, vol. 10, no. 2, pp. 129–144, 2020.
- [5] R. Yetkin and Z. Özer, "An Investigation into In-Service and Pre-Service English Teachers' Conception of Assessment," *Journal of Theoretical Educational Science*, vol. 13, no. 2, pp. 384–396, doi: 10.30831/akukey.582886.
- [6] K. Vogt, D. Tsagari, I. Csépes, A. Green, and N. Sifakis, "Linking Learners' Perspectives on Language Assessment Practices to Teachers' Assessment Literacy Enhancement (TALE): Insights from Four European Countries," *Language Assessment Quarterly*, vol. 17, no. 4, pp. 410–433, Aug. 2020, doi: 10.1080/15434303.2020.1776714.
- [7] M. N. A. Aziz, N. Mohd Yusoff, and M. F. Mohd Yaakob, "Challenges in using authentic assessment in 21st century ESL classrooms," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 9, no. 3, p. 759, Sep. 2020, doi: 10.11591/ijere.v9i3.20546.

- [8] S. S. Jawhar and A. M. Subahi, "The Impact of Specialty, Sex, Qualification, and Experience on Teachers' Assessment Literacy at Saudi Higher Education," *International Journal of Learning, Teaching and Educational Research*, vol. 19, no. 5, pp. 200–216, May 2020, doi: 10.26803/ijlter.19.5.12.
- [9] M. Shahzamani and M. H. Tahririan, "Iranian Medical ESP Practitioners' Reading Comprehension Assessment Literacy: Perceptions and Practices," *Iranian Journal of English for Academic Purposes*, vol. 10, no. 1, pp. 1–15, 2021.
- [10] V. Nimehchisalem and N. Bhatti, "A review of literature on language assessment literacy in last two decades (1999-2018)," *International Journal of Innovation, Creativity and Change*, vol. 8, no. 11, pp. 44–59, 2019.
- [11] E. Pehlivan Şişman and K. Büyükkaracı, "A Review of Foreign Language Teachers' Assessment Literacy," *Sakarya University Journal of Education*, pp. 628–650, Dec. 2019, doi: 10.19126/suje.621319.
- [12] C. DeLuca, D. LaPointe-McEwan, and U. Luhanga, "Teacher assessment literacy: a review of international standards and measures," *Educational Assessment, Evaluation and Accountability*, vol. 28, no. 3, pp. 251–272, Aug. 2016, doi: 10.1007/s11092-015-9233-6.
- [13] C. M. Gotch and B. F. French, "A systematic review of assessment literacy measures," *Educational Measurement: Issues and Practice*, vol. 33, no. 2, pp. 14–18, 2014, doi: 10.1111/emip.12030.
- [14] C. DeLuca *et al.*, "Policies, Programs, and Practices: Exploring the Complex Dynamics of Assessment Education in Teacher Education Across Four Countries," *Frontiers in Education*, vol. 4, Nov. 2019, doi: 10.3389/feduc.2019.00132.
- [15] and N. American Federation of Teachers, National Council on Measurement in Education and E. Association, "Standards for Teacher Competence in Educational Assessment of Students," *Educational Measurement: Issues and Practice*, vol. 9, no. 4, pp. 30–30, Dec. 1990, doi: 10.1111/j.1745-3992.1990.tb00391.x.
- [16] S. M. Brookhart, "Educational Assessment Knowledge and Skills for Teachers," *Educational Measurement: Issues and Practice*, vol. 30, no. 1, pp. 3–12, Mar. 2011, doi: 10.1111/j.1745-3992.2010.00195.x.
- [17] O. Inbar-Lourie, "Constructing a language assessment knowledge base: A focus on language assessment courses," *Language Testing*, vol. 25, no. 3, pp. 385–402, Jul. 2008, doi: 10.1177/0265532208090158.
- [18] R. Lam, "Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong," *System*, vol. 81, pp. 78–89, Apr. 2019, doi: 10.1016/j.system.2019.01.006.
- [19] J. Willis, L. Adie, and V. Klenowski, "Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform," *Australian Educational Researcher*, vol. 40, no. 2, pp. 241–256, 2013, doi: 10.1007/s13384-013-0089-9.
- [20] A. Looney, J. Cumming, F. van Der Kleij, and K. Harris, "Reconceptualising the role of teachers as assessors: teacher assessment identity," *Assessment in Education: Principles, Policy and Practice*, vol. 25, no. 5, pp. 442–467, 2018, doi: 10.1080/0969594X.2016.1268090.
- [21] Y. Xu and G. T. L. Brown, "Teacher assessment literacy in practice: A reconceptualization," *Teaching and Teacher Education*, vol. 58, pp. 149–162, Aug. 2016, doi: 10.1016/j.tate.2016.05.010.
- [22] B. S. Plake, J. C. Impara, and J. J. Fager, "Assessment Competencies of Teachers: A National Survey," *Educational Measurement: Issues and Practice*, vol. 12, no. 4, pp. 10–12, 1993, doi: 10.1111/j.1745-3992.1993.tb00548.x.
- [23] C. A. Mertler and C. Campbell, "Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the Assessment Literacy Inventory," *American Educational Research Association*, p. 27, 2005.
- [24] R. Talib, M. N. A. Ghafar, and H. A. Naim, "Assessment Literacy: A Catalyst to the Success of School-Based Assessment in Malaysian Schools," in *Proceedings of the International Conference on Science, Technology and Social Sciences (ICSTSS) 2012*, 2014, pp. 197–202. doi: 10.1007/978-981-287-077-3\_23.
- [25] L. Dunlai and S. You, "An Investigation of Chinese Middle School In-Service English Teachers Assessment Literacy," *Indonesian EFL Journal*, vol. 1, no. 1, p. 1, Sep. 2017, doi: 10.25134/iefj.v1i1.607.
- [26] L. Volante and X. Fazio, "Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development," *Canadian Journal of Education*, vol. 30, no. 3, pp. 749–770, 2007, doi: 10.2307/20466661.
- [27] R. Esfandiari and R. Nouri, "A Mixed-methods, Cross-sectional Study of Assessment Literacy of Iranian University Instructors: Implications for Teachers' Professional Development," *Iranian Journal of Applied Linguistics*, vol. 19, no. 2, pp. 115–154, 2016, doi: 10.29252/ijal.19.2.115.
- [28] C. DeLuca, D. LaPointe-McEwan, and U. Luhanga, "Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy," *Educational Assessment*, vol. 21, no. 4, pp. 248–266, Oct. 2016, doi: 10.1080/10627197.2016.1236677.
- [29] F. Nikmard and Z. Mohamadi Zenouzagh, "Designing and validating a potential assessment inventory for assessing ELTs' assessment literacy," *Language Testing in Asia*, vol. 10, no. 1, 2020, doi: 10.1186/s40468-020-00106-1.
- [30] L. G. Daniel and D. A. King, "Knowledge and Use of Testing and Measurement Literacy of Elementary and Secondary Teachers," *The Journal of Educational Research*, vol. 91, no. 6, pp. 331–344, Jul. 1998, doi: 10.1080/00220679809597563.
- [31] C. Coombe, K. Folse, and N. Hubley, *A Practical Guide to Assessing English Language Learners*. Michigan Publishing, 2016, doi: 10.3998/mpub.170430.
- [32] L. M. Earl, *Assessment as Learning*. New York: Crowin, 2003.
- [33] L. Earl and S. Katz, "Rethinking Classroom Assessment with Purpose in Mind," *Learning*, pp. 1–98, 2006.
- [34] C. B. A. Mednick and Foreword, "Quality Performance Assessment: A Guide for Schools and Districts By Center for Collaborative Education," 2013.
- [35] E. Ölmezer-Öztürk and B. Aydın, "Investigating language assessment knowledge of EFL teachers," *Hacettepe Egitim Dergisi*, vol. 34, no. 3, pp. 602–620, 2019, doi: 10.16986/HUJE.2018043465.
- [36] E. W. Wolfe, S. G. Viger, D. W. Jarvinen, and J. Linksman, "Validation of Scores from a Measure of Teachers' Efficacy Toward Standards-Aligned Classroom Assessment," *Educational and Psychological Measurement*, vol. 67, no. 3, pp. 460–474, Jun. 2007, doi: 10.1177/0013164406292091.
- [37] C. Deluca and D. A. Klingerb, "Assessment literacy development: Identifying gaps in teacher Candidates' learning," *Assessment in Education: Principles, Policy and Practice*, vol. 17, no. 4, pp. 419–438, 2010, doi: 10.1080/0969594X.2010.516643.
- [38] D. D. Allen and R. F. Flippo, "Alternative Assessment in The Preparation of Literacy Educators: Responses from Students," *Reading Psychology*, vol. 23, no. 1, pp. 15–26, Jan. 2002, doi: 10.1080/027027102317345385.
- [39] G. T. L. Brown, "Teachers' conceptions of assessment: Implications for policy and professional development," *Assessment in Education: Principles, Policy and Practice*, vol. 11, no. 3, pp. 301–318, 2004, doi: 10.1080/0969594042000304609.
- [40] G. T. L. Brown, "Teachers' conceptions of assessment: Validation of an abridged version," *Psychological Reports*, vol. 99, no. 1, pp. 166–170, 2006, doi: 10.2466/PRO.99.1.166-170.
- [41] M. Thomas, "Teachers' Beliefs about Classroom Assessment and their selection of Classroom Assessment Strategies," *Journal of Research & Reflections in Education*, vol. 6, no. 2, pp. 104–115, 2012.

- [42] R. Hildén and B. Fröjdendahl, "The dawn of assessment literacy – exploring the conceptions of Finnish student teachers in foreign languages," *Apples – Journal of Applied Language Studies*, vol. 12, no. 1, pp. 1–24, 2018.
- [43] Z. Zhang and J. Burry-Stock, "Assessment Practices Inventory: A multivariate analysis of teachers' perceived assessment competency," *Annual Meeting of the National Council on Measurement in Education, Chicago, IL*, 1997.
- [44] L. Bol, P. L. Stephenson, A. A. O'connell, and J. A. Nunnery, "Influence of Experience, Grade Level, and Subject Area on Teachers' Assessment Practices," *The Journal of Educational Research*, vol. 91, no. 6, pp. 323–330, Jul. 1998, doi: 10.1080/00220679809597562.
- [45] C. A. Mertler, "Classroom assessment practices of Ohio teachers," A Paper Presented at the 1998 Annual Meeting of the Mid-Western Educational Research Association, 1998.
- [46] L. Cheng, T. Rogers, and H. hu, "ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures," *Language Testing*, vol. 21, no. 3, pp. 360–389, 2004, doi: 10.1191/0265532204lt288oa.
- [47] O. Inbar-Lourie and S. Donitsa-Schmidt, "Exploring classroom assessment practices: the case of teachers of English as a foreign language," *Assessment in Education: Principles, Policy & Practice*, vol. 16, no. 2, pp. 185–204, Jul. 2009, doi: 10.1080/09695940903075958.
- [48] S. L. Suah and S. L. Ong, "Investigating assessment practices of in-service teachers," *International Online Journal of Educational Sciences*, vol. 4, no. 1, pp. 91–106, 2012, [Online]. Available: [http://www.iojes.net/userfiles/Article/IOJES\\_724.pdf](http://www.iojes.net/userfiles/Article/IOJES_724.pdf)
- [49] C. Jinyoung, "Understanding elementary teachers' different responses to reform: The case of implementation of an assessment reform in South Korea," *International Electronic Journal of Elementary Education*, vol. 9, no. 3, pp. 581–598, 2017.
- [50] M. James and D. Pedder, "Beyond method: Assessment and learning practices and values," *Curriculum Journal*, vol. 17, no. 2, pp. 109–138, 2006, doi: 10.1080/09585170600792712.
- [51] Y. Chan, "Elementary school EFL teachers' beliefs and practices of multiple assessments," *Reflections on English Language Teaching*, vol. 7, no. 1, pp. 37–62, 2008, [Online]. Available: <http://nus.edu.sg/celc/publications/RELT71/37to62-chan.pdf>
- [52] R. D. Gonzales and J. Aliponga, "Classroom Assessment Preferences of Japanese Language Teachers in the Philippines and English Language Teachers in Japan," *SSRN Electronic Journal*, 2017, doi: 10.2139/ssrn.2746287.
- [53] A. G. Thompson, "Teachers' beliefs and conceptions: A synthesis of the research," in *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics*. Macmillan Publishing, 1992, pp. 127–146.

## BIOGRAPHIES OF AUTHORS



**Geng Juanjuan**    is a PhD candidate at Universiti Utara Malaysia, and currently a lecturer of Xianyang Normal University. Her main research directions are pre-service and in-service teacher education. Relating to her research area, she has written and published over 10 articles in journals and proceeding of international conference. She can be contacted at email: [mylovejijl23@gmail.com](mailto:mylovejijl23@gmail.com).



**Nurahimah Mohd Yusoff**    is a Professor at the School of Education, Universiti Utara Malaysia. Her research interests include curriculum design and development, learning and teaching in higher education, educational evaluation, and qualitative research. She had her B.A. (Linguistics) from Victoria University of Wellington, New Zealand in 1987; M.A. (Applied Linguistics) from University of Edinburgh, United Kingdom in 1991, and Ed.D (Curriculum and Instruction) from University of Virginia, USA in 1999. She can be contacted at email: [nurayusoff@gmail.com](mailto:nurayusoff@gmail.com).