



Research article

Language assessment course at Iranian state universities: An evaluation of the incorporation of assessment principles into the course content

Masoomeh Taghizadeh, Ph.D, Golnar Mazdayasna, Ph.D *

Department of English Language and Literature, Yazd University, Yazd, Iran

ARTICLE INFO

Keywords:

Assessment
Language assessment literacy
Assessment principles
Assessment training
Pre-service program

ABSTRACT

Upon perceiving the development of assessment literacy as a critical requirement for teachers' professional growth, the effectiveness of assessment training programs has been investigated from various perspectives. To add more to the research on this area, this study evaluated the efficacy of language assessment course at Iranian state universities in raising prospective teachers' awareness of the principles underlying language assessment. Accordingly, the course syllabus for English majors at the BA level, issued by the Ministry of Science, Research, and Technology (MSRT [1]), was content-analyzed. In addition, 92 language assessment course instructors' areas of content focus were explored through some Likert-scale items, and two instructors' teaching practices were observed throughout the course. Although the study did not claim that the course students lacked language assessment literacy (LAL), results cast doubts on the quality of their LAL concerning the principles of assessment. Overall, it seems that the assessment principles are given low priority by the syllabus designers and course instructors, which is contrary to the general debates on LAL. Concerning the underlying sub-components, reliability and validity qualities seemed to be incorporated into the course content, whereas other central themes (e.g., ethical considerations, fairness in assessment, test washback, etc.) were excluded. Finally, a set of key actions have been suggested to be taken by the course instructors and policymakers/syllabus designers to raise the course students' awareness of the principles of assessment.

1. Introduction

Assessment literacy (AL) is viewed as a necessity for modern professional teachers. For Stiggins [2,3], one of the noted advocates, assessment literacy is a crucial requirement for educators, the benefits of which can be their progressive professional achievements and the learning success of their students. While stressing the need for a profound consideration of what a test really entails and what the harmful effects of poor testing could be, Stiggins invited teachers to develop assessment competencies in assessment.

Since then, there have been increasing calls for a proper investigation of teachers' assessment literacy and designing suitable training programs to boost the functionality and efficiency of assessment in education [4–13]. In this case, pre-service teacher training programs can play a critical role in developing teachers' assessment literacy to implement effective assessment practices [14,15]. In

Abbreviations: LAL, Language Assessment Literacy; AL, Assessment Literacy; MSRT, Ministry of Science, Research, and Technology.

* Corresponding author.

E-mail address: gmazdayasna@yazd.ac.ir (G. Mazdayasna).

<https://doi.org/10.1016/j.heliyon.2023.e12857>

Received 5 July 2022; Received in revised form 31 December 2022; Accepted 4 January 2023

Available online 9 January 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

language teaching, these training programs have primarily been through a language assessment course, added as a necessity into the training programs. However, as Giraldo [15] mentioned, most of the studies on language assessment literacy (LAL) have investigated in-service teachers and their assessment-related professional needs, while studies on pre-service teachers are just starting to appear [16–18].

In line with these studies, the present research attempted to investigate whether the language assessment course, as a pre-service training course, at the BA level at Iranian state universities, builds up prospective teachers' LAL. The study focused on the efficacy of the course in raising pre-service teachers' awareness of the principles underlying language assessment, which seems to be an understudied and one of the poorest aspects of LAL [19]. For this purpose, the content of the official syllabus issued by MSRT [1] was analyzed. Furthermore, the course instructors' areas of focus were identified through a survey and classroom observations.

Accordingly, the present study stresses that assessment principles such as ethics, fairness, and washback effect require to be highlighted in pre-service and in-service programs. Hence, feedback from the research may be beneficial in LAL debates in various educational contexts and as an evaluation criterion to judge the programs based on their incorporation of the principles into the assessment courses. Besides, the findings can make policymakers aware of the nature of the course; hence, workable actions can be taken by them to address the problems associated with the course and design a more comprehensive one. In addition, the designed questionnaire and the observation scheme can be utilized as suitable scales for diagnostic purposes by teacher trainers and course instructors, for instance, for getting informed about the poor areas they need to focus on more in their prospective language assessment course and help them make a critical evaluation of their practices. Likewise, in-service teachers can identify their training needs and evaluate their practices. Findings can also benefit the design of other informed programs or professional initiatives such as workshops and conferences.

2. Literature review

Since three decades ago, scholars in education have called for the development of assessment literacy by different stakeholders, including teachers [2–4,8,11–13,19–22]. The significant contribution of being assessment literate is that teachers are far more likely to avoid making the assessment-related mistakes so common in educational systems, which can lead to the success of learners and schools [4,12,20]. On the other hand, a lack of assessment literacy results in the inappropriate use of tests/assessments in evaluating students' learning, which can even be detrimental to their lives [2,3,11–13,15,22,23]. Therefore, assessments produce better results in contexts enriched with high levels of assessment literacy, and malfunction occurs when assessment literacy levels are low. Such assessment-associated competencies are called *Language Assessment Literacy* (LAL) in language studies [24].

Generally, LAL has been defined as a set of competencies that help teachers in the test design, analysis of data, evaluation, and other related aspects [4,5,7,10,23,25–27].

Davies [23], perceiving assessment literacy as a field, categorized LAL into the three components of skills, knowledge, and principles. He drew attention to the general knowledge about language, practical aspects of assessment, and assessment principles, with a particular concentration on the ethics of assessment. Fulcher [5] defined LAL as:

... The ability to place knowledge, skills, processes, principles, and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (p. 125)

Scarino [26] pointed out that teachers' personal beliefs about assessment should be included in the concept of assessment literacy. He argued about the concept of assessment culture in which personal beliefs and social and cultural practices of teachers are reflected in their assessment practices. Taylor [27], reviewing studies on LAL, proposed the eight levels of knowledge for LAL as (a) knowledge of theory; (b) technical skills; (c) principles and concepts; (d) language pedagogy; (e) sociocultural values; (f) local practices; (g) personal beliefs/attitudes; and (h) scores and decision making. Lan and Fan [25], based on data from the CBLAL inventory, proposed four dimensions for LAL including (a) technical skill; (b) scoring and decision making; (c) language pedagogy; and (d) principles in language assessment. In a similar attempt, Mohammadkhah et al. [10] introduced the four factors of assessment in language pedagogy, assessment theories and principles, scoring and interpretations, and technical skills as the constructing cores.

Inherent in all these conceptualizations of LAL, in addition to the theoretical and methodological aspects underlying assessment, is the inclusion of the assessment principles. It can be inferred from the above views on LAL that in a standard-based system, language assessment literacy involves more than the mere knowledge of test instruments and testing theories.

The main issue concerning adherence to assessment principles relates to assessment quality management, e.g., ensuring reliability, validity, equity and fairness, ethicality, practicality, authenticity, and test washback, which test developers should take into account throughout all stages of the assessment process to certify favorable results [28,29]. The initial call for developing competence in these areas in language studies was echoed by Davies [23]. Davies mentioned the transition in focus from *skills* to *skills + knowledge* to *skills + knowledge + principles* in assessment. He further criticized the mere inclusion of knowledge and skills, overlooking the principles in assessment.

Therefore, as Davies discussed, teachers and assessment designers should consider issues beyond the mere design of assessment tasks. That is, teachers are required to raise awareness of the principles of assessment to promote their teaching and fulfil the needs of their learners and the context in which they work. As such, assessment literates are concerned about the consequences of using their assessments. They know how to select appropriate assessment methods that would not overburden learners and at the same time brings them the most benefits, use assessment to motivate learners, set standards and priorities for learning, and inform learners of their progress [7,15,17,23,27,30].

Although the topic is not in its infancy, there are still increasing calls for taking a more critical consideration of the development of and research on this dimension of LAL. Kremmel and Harding [7], underscoring the importance of developing competence in assessment principles argued that “a heightened awareness of the principles and practice of language assessment would ideally lead to more informed discussion of assessment matters, clarity around good practice in using language assessments, and ultimately more robust decision making on the basis of assessment data” (p.100). As one more recent endeavor to highlight the role of principles of LAL, reference can be made to the study by Tajeddin et al. [19], which aimed to design and validate an inventory to measure Iranian EFL teachers’ awareness of principles, referred to as ‘critical language assessment literacy’ (CLAL). According to their findings, CLAL comprised “(a) teachers’ knowledge of assessment objectives, scopes, and types; (b) assessment use consequences; (c) fairness; (d) assessment policies; and (e) national policy and ideology” (p. 649).

To sum up, it is necessary for teachers to develop an awareness of this dimension as an essential component of their LAL, and it is vital to investigate the efficacy of training in this area.

Nevertheless, research has generally pointed to the inadequate knowledge base of teachers with respect to the principles of LAL and underlying aspects or the lack of focus on this dimension in teacher training programs [5,15,16,25,26,31–38].

MacLellan’s [36] study showed that pre-service teachers were cognizant enough of assessment purposes, while they showed limited knowledge concerning theories and principles of assessment. Fulcher [5] indicated a low ranking for the principles of assessment on the part of the teachers. Scarino [26] indicated that teachers showed weaknesses in theoretical, practical, and institutional aspects. She suggested that the framework of pre-service training programs be designed based on a broader approach, providing much more than mere assessment knowledge.

In Kremmel et al. [35], teachers reported that they did not receive adequate training on principles such as ethicality. Lan and Fan’s [25] investigation of teachers’ assessment competence in the four dimensions in the CBLAL revealed low scores for the knowledge of theories and assessment principles. Anam and Putri’s [16] evidence from Indonesian EFL pre-service and in-service teachers indicated that both groups rated their ability in determining validity and reliability as low but high in ensuring fairness quality. Finally, based on a meta-analysis of studies on language assessment programs, Giraldo [15] indicated that only six out of 14 programs investigated in the studies included the principles for language assessment in their syllabi. Therefore, it seems that principles for LAL should be further incorporated into teacher training programs; particularly, because principles have become central considering the role of language assessment in society.

In the Iranian EFL context, despite such warnings addressing the hazards of poor language assessment literacy, there still appears to be reports of in-service teachers’ inappropriate assessment practices [39,40,41,42]. However, most of the studies on LAL have focused on the theoretical and practical components of LAL, while the contribution of teacher training programs to the development of competence in the principles of assessment has not been the focal topic of the studies. Hence, the current study was conducted to investigate whether language assessment course at state universities, as the prevalent pre-service teacher training program in the Iranian EFL context, is efficient in familiarizing prospective teachers with the principles of assessment.

In Iran, the educational system at state universities follows a centralized approach to policy making and practice controlled by MSRT. The organization exerts its influence by deciding on curriculum development and implementation, setting objectives, determining the type of teaching instruction and textbooks/materials, specifying the teaching time, and so on. Nevertheless, course instructors have a relative agency to decide on what to teach and how to teach and assess. The English Language BA program consists of 8 semesters, covering a number of general courses, basic courses, and specialized courses.

However, the BA level English language courses have not been thoughtfully and critically evaluated. Nevertheless, based on the few studies conducted [43–48], the program seems to be suffering from some critical drawbacks.

Fathi Vajergah [44] believed that the system permits little or no modification of course topics by universities. Similarly, Nasr-e-Esfehani [47] criticized the system for discouraging faculty members from manipulating the syllabi as such efforts do not have any positive effects on their salary or credit. Hayati [45] believed that the policies, decisions, and objectives set for the programs are not based on current research findings. Mehrmohammadi [46] referred to the ‘illusion of comprehensive curriculum’ and the simplistic views on different aspects of the curriculum, disregarding the notion of curriculum evaluation. In an investigation of the ELT curricula at BA and MA levels, Divsar and Jafari Gohar [43] stated that the curricula did not promote critical thinking skills.

Relatively, the research findings indicate that a top-down trend in curriculum design and policy-making is still dominant in Iran. However, it seems that further critical and evaluative efforts are demanded to provide more accurate evidence on the efficacy and quality of the BA English language curriculum and to identify the requirements for the needed reformations. Accordingly, this study aimed to evaluate the quality of the BA language assessment course, with respect to its efficacy in developing the course students’ LAL. In the BA English language syllabus, a two-credit assessment course has been presented for English language majors. The number of sessions and the teaching time has been assigned by MSRT. A general syllabus, including a suggestive content of instruction and a list of teaching textbooks, has also been introduced.

Based on the purpose introduced, the following two questions were addressed in the study:

1. To what extent are the principles of LAL incorporated into the language assessment course at the state university BA level?
2. What underlying aspects related to the principles of LAL are highlighted and/or overlooked in the language assessment course at the state university BA level?

3. Materials and methods

3.1. The conceptual framework and research design

The conceptual framework of this study was driven by Davies' [23] conceptualization of LAL, which defines LAL as comprising competence in knowledge, skills, and principles required to conduct language assessment in different educational settings. Davies' model of LAL has been frequently reported as a theoretical framework in research on LAL [5,17,24]. In this view on LAL, knowledge refers to the theoretical background in language, learning, and assessment. The skills dimension addresses the methodological aspects (issues of item construction, scoring, interpretation, etc.). Principles are the underlying concepts such as validity, reliability, and ethics.

Nevertheless, it needs to be noted that although LAL's three broad components have been stable in the literature, there have been gaps in grouping the underlying aspects of language assessment into the three components. Therefore, for this research, the initial step to the development of the required instruments was to determine which assessment areas belong to the principles component, as indicated in the literature and favored by the current prevailing standards (e.g., JCSEE [49]), and to map the categorization in detail. Based on an in-depth review of the relevant literature and experts' consultations, a set of sub-dimensions (see Table 1) were identified.

The identified areas provided the bases for the development of the instruments required to conduct the research. In the present study, it was decided that an amalgamation of methods would increase reliability and validity of the findings. Such a design could contribute to producing more trustworthy and generalizable results from various angles [50]. Based on the methodological approach taken, three research methods, i.e., the official syllabus introduced by MSRT [1], a survey, and classroom observation, were found suitable to explore the research questions addressed in the study. Before the research was conducted, the research proposal and the instruments were reviewed and endorsed by the Research Committee at the Faculty of Languages and Literature at Yazd University (Decision Number: 760), and ethical approval was obtained. The study was also approved (Decision Number: 2837341) by the Iranian documentation center (Iranodoc), an institution under the supervision of the MSRT. The course instructors were also assured by the researchers about anonymity, data protection, and ethical considerations.

3.2. Research instruments

3.2.1. The MSRT official syllabus

The university English course syllabus introduced by MSRT [1] was content-analyzed. To ensure reliability, coding was done by both researchers, and a relatively high Kappa coefficient of 0.94 was obtained. Further discussions and consultations were helpful to come into a compromise where opinions differed.

3.2.2. Questionnaire

For the present research, a questionnaire (see Appendix A) was developed to describe the degree to which language assessment course instructors at the BA level at state universities incorporate issues related to the principles of LAL. For this purpose, a set of Likert-scale items (based on the identified sub-dimensions) were designed through which the course instructors indicated the degree of focus they allocated to each sub-dimension, ranged from 0 (meaning none) to 4 (indicating extensively).

3.2.2.1. Participants. Ninety-two course instructors, including 33 female and 59 male instructors, 56 holding PhD, 28 PhD candidates, and 8 MA graduates, with the age range of 29–64, responded to the survey. Respondents were from universities in different provinces within Iran. Concerning their fields of expertise, 79 specialized in Teaching English as a Foreign Language (TEFL), five in Linguistics, and eight in Translation. Their teaching experience at university ranged from two to 33 years, and all of them had instructed the language assessment course during the past two years.

Table 1

Sub-components related to the principles of LAL.

Ethical considerations in testing/assessment
Making sound decisions based on test/assessment results
Considerations of test reliability
Considerations of test validity
Doing assessment based on multiple sources of evidence
Consideration of practicality in assessment
Importance of incorporating formative assessment
Importance of incorporating alternative assessment
Consideration of transparency in assessment
Consideration of authenticity in assessment
Importance of incorporating learner autonomy in assessment
Considerations of test washback/impact
Considerations of fairness/bias in assessment
Consideration of societal/cultural aspects in assessment/testing

3.2.2.2. Procedure. Initially, the designed questionnaire was reviewed by two experts to identify the possible pitfalls associated with the instrument from different perspectives. Following the development phase, the questionnaire was piloted among 30 instructors in order to predict and eliminate potential problems before the target administration. The finalized questionnaire was delivered to 121 instructors via e-mail, personal visits, or social networks. The recipients received a reminder some weeks later. Ninety-two course instructors returned the questionnaire, indicating a response rate of 76%. The exact purpose of the research was kept unknown; however, the confidentiality of the information was assured. Cronbach's analysis indicated a reliability index of 0.78 for the questionnaire. SPSS software (version 26.0.) was used to analyze the data.

3.2.3. Observation

The questionnaire was helpful in getting a general understanding of the course instructors' content focus in their classes; however, to see more closely what happens in the course, classroom observation seemed a rational choice. The content of instruction concerning the principles in assessment was the focus of the inquiry. For this purpose, an observation scheme (see Appendix B), including three sections, was designed. The first section collected general information about the course and each session. The second section investigated the instructors' amount of focus on the principles component based on the subtopics taught throughout the course. The third section targeted the amount of focus on each corresponding sub-component through a set of Likert-scale items checking if the underlying aspects were focused *extensively, moderately, a little, or not at all*.

3.2.3.1. Participants. Two course instructors' pedagogical practices, both holding Ph.D. in TEFL and with teaching experience of more than ten years, were pursued within one university semester. These two instructors were selected as the universities they were teaching at were accessible to the researchers (i.e., convenient sampling).

3.2.3.2. Procedure. The primary task to observe the classes was to design the required observation scheme. The courses were scrutinized for 15 sessions, held once a week. The aim of the research was not revealed to the course instructors. A non-participant observation technique was employed. Every session, the observation scheme was completed, field notes were taken, and each session was audio-recorded for later checks and reliability confirmation. In cases of ambiguity, after-session informal discussions were held with the instructors to explore more explanations.

As with the quantitative research method, to verify the accuracy and consistency of the results, validity and reliability qualities of the observation data needed to be addressed. In qualitative research, multiple coding is carried out to address the truthfulness of the data. For this purpose, field notes were taken, each session was audio-recorded, and inter-coder reliability was estimated to ascertain the decisions consistency. The Kappa coefficients of 0.72 and 0.81 coding agreement were achieved for sections two and three, respectively. Follow-up inter-coder discussions and colleague consultation were used to make decisions on the points of disagreement.

In qualitative studies, validity is related to the robustness of the research design, the soundness of the instruments, and the accuracy of the findings, analyses, and decisions made [51]. In the present research, efforts were made to ensure all these requirements. Throughout the data collection process, non-participatory stance and self-awareness strategies were employed to prevent the researcher's perceptions from affecting the findings, lessen bias, and improve the validity. Furthermore, consultations were sought from experts to boost the validity of the analyses and the decisions made.

In the analysis phase, for the second section (i.e., on the component focus), initially, the decisions were made on the component type. In the next step, the allocated time on each component was estimated through its transformation into 5-min units. Mean plot and Tukey HSD were carried out to compare means and significance of the differences. For the third section, post-hoc coding procedure was employed to decide the extent to which each variable was focused on, using a scale ranged from *extensively to not at all*. The Mann-Whitney test was applied to compare the mean differences for each variable across the two classes.

4. Results

Findings are presented in the following sub-sections with respect to each research instrument.

4.1. The MSRT official syllabus

The syllabus (written in Persian) expresses the goal of the course in a short sentence as "to familiarize the course students with different types of tests and testing procedures and to help them construct items for different language skills" [1, p. 42].

The official document has also made a set of content suggestions including (a) introduction to assessment/testing; (b) different types of tests; (c) approaches to testing; (d) various steps to test construction and constructing vocabulary tests; (e) constructing grammar and structure tests; (f) constructing reading tests; (g) constructing speaking tests; (h) constructing integrative tests; (i) reliability in assessment; (j) validity in assessment; (k) alternative assessment; (l) statistics in testing; and finally (m) a general review of the course.

As such, concerning the principles component of assessment, the suggestive areas included in the document for assessment training can be identified as reliability, validity, and focus on incorporating alternative assessment procedures. The suggested textbooks include Baker [52], Farhady et al. [53], Heaton [54], Weir [55], Hughes [56], McNamara [57], and Brown [58].

4.2. Questionnaire

4.2.1. Demographic data on the course

According to the findings, classes were held for about sixteen 90-min sessions. Class size ranged from 30 to 52 students. Farhady et al. [53] and Heaton [54] seemed to be the most frequently used textbooks.

4.2.2. Content focus

Table 2 shows the estimated degree of focus on the principles component, in general, and on the corresponding underlying aspects.

According to the findings, a general mean coverage of 1.81 (0.4) was estimated for the principles component. With respect to the underlying sub-components, the two qualities of test reliability (2.42 (0.79)) and test validity (2.34 (0.63)) were the prioritized variables, with considerations of practicality in assessment (2.28 (0.82)) coming next. Highlighting the importance of making sound decisions based on assessment (2.11 (0.79)) and ethical consideration in testing/assessment (2.03 (0.70)) were the next considered issue, yet receiving moderately low focus. Other underlying sub-components, such as considerations of transparency in assessment, consideration of test washback/impact, fairness issues in assessment, the effect of societal/cultural values in assessment, authenticity in assessment, doing assessment based on multiple sources of evidence, importance of incorporating formative/alternative assessment and learner autonomy in assessment, received mean coverage of less than 2.

4.3. Observation

4.3.1. Demographic data on the course

Forty-five students in class A and 38 in class B enrolled in the course. In class A, Farhady et al. [53] and Brown [59] and in class B, Farhady et al. [53] were used as the teaching textbooks. Ignoring the greeting time and assessment-irrelevant talks, 932 min for class A and 1062 min for class B were recorded. Based on the Q-Q plot (Fig. 1), the time variable seemed to be distributed almost normally in both classes, and the mean difference of time in class A (63.10 ± 6.09) and class B (68.13 ± 9.34) did not seem significant (p -value = 0.353).

4.3.2. General content focus

Fig. 2 shows the amount of focus on the principles component of LAL for each class, in comparison to the two other core components of LAL (i.e., knowledge and skills components).

Based on the findings, for both classes, the least attention was allocated to the principles of assessment. Further statistical analysis indicated significant differences in time values for both classes concerning the knowledge vs. the principles and the skills vs. the principles components (see Table 3).

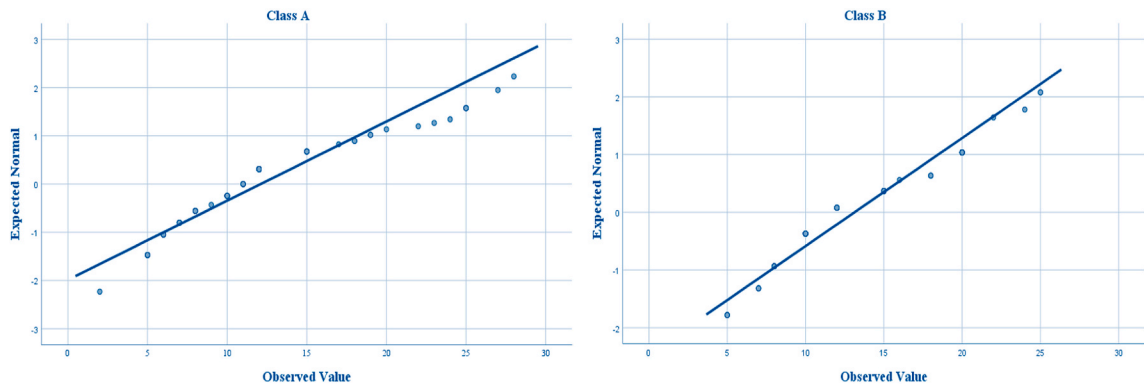
4.3.3. Sub-components content focus

This section aimed to explore the degree to which each aspect related to the principles of assessment was addressed in the content of the two classes observed. Results are presented in Table 4 below.

As the table indicates, concerning the underlying aspects related to the principles component, the two instructors seemed to act differently only in terms of their focus on the importance of incorporating alternative assessment and considerations of transparency in assessment. Both instructors focused primarily on reliability, validity, and practicality considerations in assessment. No significant differences were identified concerning the other relevant issues.

Table 2
Descriptive statistics for the underlying aspects related to the principles component.

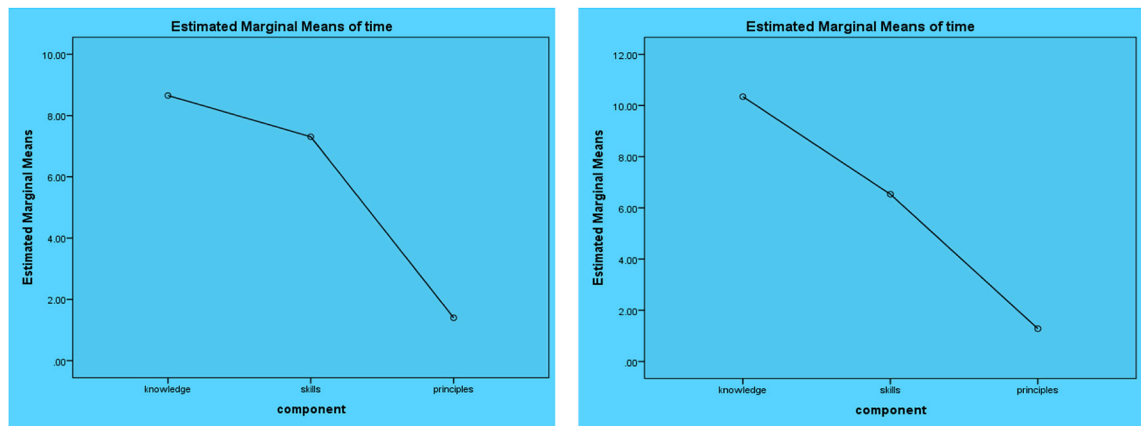
Variables	Variables		Main Dimension	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Ethical considerations in assessment	2.03 (0.70)	2.00 (2.00–3.00)	1.81 (0.4)	1.95 (1.32–2.09)
Consideration of transparency in assessment	1.86 (0.21)	2.00 (1.00–2.00)		
Making sound decisions based on sound samples of assessment	2.11 (0.79)	2.00 (1.25–3.00)		
Considerations of test reliability	2.42 (0.79)	2.00 (2.00–3.00)		
considerations of test validity	2.34 (0.63)	2.00 (2.00–2.00)		
Considerations of test washback/test impact	1.84 (0.77)	2.00 (1.00–2.00)		
Considerations of practicality in testing/assessment	2.28 (0.82)	2.00 (1.00–2.00)		
Considerations of test authenticity	0.99 (0.70)	1.00 (1.00–1.00)		
Considerations of fairness/bias in assessment/testing	1.79 (0.65)	2.00 (2.00–2.00)		
Doing assessment based on multiple sources of evidence	1.47 (0.80)	1.00 (1.00–2.00)		
Importance of incorporating formative assessment	1.63 (0.75)	2.00 (1.00–2.00)		
Consideration of societal/cultural values	1.51 (0.69)	2.00 (1.00–2.00)		
Importance of incorporating alternative assessment	1.66 (0.75)	2.00 (1.00–2.00)		
Importance of incorporating learner autonomy in assessment	1.48 (0.70)	1.00 (1.00–2.00)		



Distribution of time, Class A

Distribution of time, Class B

Fig. 1. The Quantile-Quantile plot representing the distribution of time across the classes.



Mean for components, Class A

Mean for components, Class B

Fig. 2. Mean plot for the LAL core components by class.

Table 3
Multiple comparisons of the component by class.

Dependent variable: time							
Tukey HSD							
Class	(I) component	(J) component	Mean difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Class A	Knowledge	Skills	1.346	2.324	.832	-4.300	6.993
		Principles	7.253*	2.324	.009	1.606	12.900
	Skills	Knowledge	-1.346	2.324	.832	-6.993	4.300
		Principles	5.906*	2.324	.039	.259	11.553
	Principles	Knowledge	-7.253*	2.324	.009	-12.900	-1.606
		Skills	-5.906*	2.324	.039	-11.553	-.2596
Class B	Knowledge	Skills	3.813*	2.060	.041	-.343	7.970
		Principles	9.067*	2.060	.000	4.910	13.223
	Skills	Knowledge	-3.813*	2.060	.041	-7.970	.343
		Principles	5.253*	2.060	.014	1.097	9.410
	Principles	Knowledge	-9.067*	2.060	.000	-13.223	-4.910
		Skills	-5.253*	2.060	.014	-9.410	-1.097

*The mean difference is significant at the 0.05 level.

Table 4
Comparing the distribution of variables between class A and B.

Variables	Class	Mean \pm SD	Median (Q1, Q3)	Min, Max	P-value*
Ethical practices in students' preparation/scoring	Class A	1.08 \pm 0.52	1.00 (1.00, 2.00)	1, 2	<0.141
	Class B	0.90 \pm 0.51	1.00 (0.00, 1.00)	2, 3	
Consideration of transparency in assessment	Class A	1.80 \pm 0.51	2.00 (1.00, 2.00)	2, 3	<0.001
	Class B	1.10 \pm 0.51	1.00 (1.00, 2.00)	1, 2	
Making sound decisions based on sound samples of assessment	Class A	0.47 \pm 0.52	1.00 (0.00, 1.00)	2, 3	<0.421
	Class B	0.40 \pm 0.51	1.00 (0.00, 1.00)	1, 2	
Considerations of test washback/test impact	Class A	0.73 \pm 0.46	1.00 (0.00, 1.00)	1, 2	0.124
	Class B	0.50 \pm 0.41	0.00 (0.00, 1.00)	1, 2	
Considerations of test reliability	Class A	2.33 \pm 0.49	2.00 (2.00, 3.00)	2, 3	0.201
	Class B	2.20 \pm 0.68	2.00 (2.00, 3.00)	0, 2	
Considerations of practicality in testing/assessment	Class A	1.80 \pm 0.68	2.00 (1.00, 2.00)	0, 2	0.153
	Class B	1.47 \pm 0.52	1.00 (1.00, 2.00)	0, 1	
Considerations of test validity	Class A	2.40 \pm 0.51	2.00 (2.00, 3.00)	2, 3	0.145
	Class B	2.33 \pm 0.62	2.00 (2.00, 3.00)	1, 3	
Considerations of fairness/bias in assessment/testing	Class A	1.13 \pm 0.64	1.00 (1.00, 2.00)	1, 3	<0.311
	Class B	0.93 \pm 0.70	1.00 (1.00, 1.00)	0, 2	
Doing assessment based on multiple sources of evidence	Class A	0.23 \pm 0.51	0.00 (0.00, 1.00)	1, 2	0.105
	Class B	0.40 \pm 0.68	0.00 (0.00, 1.00)	0, 2	
Importance of incorporating formative assessment	Class A	0.93 \pm 0.70	1.00 (0.00, 1.00)	0, 2	0.114
	Class B	0.70 \pm 0.51	1.00 (0.00, 1.00)	0, 1	
Importance of incorporating alternative assessment	Class A	0.90 \pm 0.28	1.00 (0.00, 1.00)	0, 1	0.041
	Class B	0.38 \pm 0.32	1.00 (0.00, 1.00)	0, 1	
Considerations of test authenticity	Class A	0.60 \pm 0.51	1.00 (0.00, 1.00)	0, 1	0.269
	Class B	0.53 \pm 0.52	1.00 (0.00, 1.00)	0, 1	
Consideration of societal/cultural values in testing	Class A	0.70 \pm 0.21	1.00 (0.00, 1.00)	0, 1	0.106
	Class B	0.58 \pm 0.41	1.00 (0.00, 1.00)	0, 1	
Importance of incorporating learner autonomy in assessment	class A	0.60 \pm 0.51	1.00 (0.00, 1.00)	0, 1	0.103
	class B	0.33 \pm 0.49	0.00 (0.00, 1.00)	0, 1	

*The exact Man-Whitney test.

5. Discussion

The main issues regarding adherence to assessment principles are discussed below. However, it needs to be noted that in the case of the course instructors, generally, the study investigated whether the instructors focused on the related aspects in the teaching content. In other words, the instructors' adherence to these considerations in their assessment practices was not of concern as such aspects are generally not explicitly observable, and such exploration demands using a different research design (e.g., interviewing the instructors). Nevertheless, in some cases, reference might be made to some evidence in the assessment practices of the two instructors whose classes were observed.

There is a consensus that assessments should have adequate levels of validity and reliability in order to provide accuracy and appropriateness of the interpretations prior to making decisions. Based on the content analysis of the official syllabus, it seems that the policymakers/syllabus designers have explicitly stressed the importance of addressing these two qualities, which are dominant in language testing and assessment and are discussed in most of the textbooks written in the field. In the data from the survey and observations, correspondingly, these two qualities seemed to be the central themes. Practicality, which refers to the fact that the assessment procedures including construction, administration, and scoring should be simple and time-efficient, came next in the data from the survey and observation; however, it was not indicated in the official syllabus.

The transparency principle prescribes that it is merited and necessary that prior to taking the test, test takers be told about the purpose and requirements of the assessment, the content to be assessed, and the specific scoring criteria. Transparent feedback and interpretation of test results pertinent to the purpose of the assessment are also essential. This issue is not listed in the topics proposed in the official syllabus. In the case of the survey data, instructors did not report explicit focus on this issue. Similarly, the instructors whose classes were observed did not talk to their students about the importance of transparency; however, the principle was observed to some extent in their assessment practices. For instance, the instructors provided students with information on the date of the coming test and specified the content to be tested. However, detailed feedback on the test was almost lacking, and students were provided with a general score.

The significant function of ethics in assessment is to consider the critical matters of fairness in assessment and the appropriate use of assessment data. The issue of ethics is also closely connected to fairness in administering, scoring, interpreting, and reporting the results in a principled manner. Theoretically, results did not indicate the syllabus designers and instructors' concentration on teaching ethics and fairness. The selected course materials did not include content associated with ethics in assessment, either. Generally, it seemed that the approaches taken by the course instructors were not aligned with the recommendations in literature that teachers need to be aware of the significance of assessing their learners fairly and unbiasedly [9,19,28,60–62]. In practice, it cannot be claimed that the two course instructors were inconsiderate about fairness and ethical issues; however, the lack of such considerations was somehow noticeable in some aspects. For instance, both instructors released the quizzes/mid-term exam scores publicly. However, Bachman &

Palmer [28], warning against the unethical or unfair release of students' achievements, recommended that reports be personal to protect the rights of the learners and achieve fairness.

Another consideration relates to the washback effect of testing/assessment on teaching and learning. In the present study, the issue of test consequences also seemed to be an underestimated aspect by the syllabus designers and the course instructors. Similarly, analysis of the data from the three instruments indicated consideration of the societal/cultural values to be an overlooked topic.

There are other critical aspects related to the principles of assessment that need attention on the part of the syllabus designers and the instructors. It is proposed in the literature that decisions made based on the assessment results be an accurate picture of student learning outcomes. Thus, it is essential to generate sound samples of students' performance. There is also a consensus that decisions should not be based on a single assessment task/test result. Instead, decisions need to be made using multiple sources of evidence collected on students' achievements throughout the course [28,62]. This can be justified with respect to the criterion of the reliability of the scores and the validity of the decisions made [28]. Analysis of the syllabus, the questionnaire, and the observation data indicated this aspect to be a skipped topic in the course content.

There have also been calls for educational systems to employ various assessment methods to permit all learners to demonstrate their knowledge and skills. This is also closely linked to the issue of equity, which recognizes that there must be a correspondence between the assessment method and the students learning styles, and to achieve this, employing various assessment methods has been advocated [28,63]. Such a focus on alternative assessment was greatly neglected by the course instructors, despite the syllabus designers' consideration. Adherence to such principle was not even observed in the assessment practices of the two instructors, as they mostly used paper and pencil, multiple-choice, or open-ended questions to assess the students' learning.

It has also been suggested that learners be involved in the assessment process [6,64,65]. For instance, Fulcher [6] called for training and adapting a learner-centered approach in assessment. That is, teachers and assessment developers have to work to foster learner autonomy in the assessment process, which will ensure greater learning. The view is in accord with the critique dimension of Chan and Luo's [66] learners' assessment literacy model. Chan and Luo highlighted that students should have critical engagement with assessment; that is, they should have "the right to critique the assessment used, the feedback provided, and take control of their holistic competency development" [66, p.12]. The topic is not included in the official syllabus. On the theoretical ground, little focus was reported on the issue through the survey data. Observation data also indicated a lack of attention to this aspect of assessment. On the practical level, the approaches seemed to be teacher-directed as the instructors determined what to assess, how to assess, and so on. Moreover, feedback on the teaching and assessment process was not sought from the students.

Finally, another principle associated with assessment literacy relates to domain sampling or collecting assessment examples representative of real-world context, referred to as authenticity in assessment. Although authenticity is not mentioned as a critical feature of assessment, especially in EFL contexts, controversies arise when the richness of assessment literacy is weakened by the overuse of assessment forms that are unidimensional or inauthentic (e.g., multiple-choice items). The lack of authenticity of such assessment methods can be frustrating as results do not usually represent students' knowledge and skills [28]. The present research findings did not seem to align with this recommendation, as the principle was overlooked by the syllabus designers and the course instructors. In the observed classes, the two instructors seemed to avoid considering authenticity in designing assessments in their practices as they mainly employed traditional assessment procedures. Notwithstanding, more real-life assessment tasks, such as observing students or engaging them in performing assessment tasks, could be much more informative and rewarding.

Generally, data from the analysis of the official syllabus and the course instructors' practices revealed that principles of assessment were not adequately addressed in the language assessment course. Familiarity with and consideration of principles is not included in the goals presented for the course and the content suggested in the official syllabus. More alarming is the fact that the syllabus has not been revised for more than a decade. Besides, the textbooks being utilized by most instructors and those suggested by the official syllabus date back to more than two decades ago. These textbooks are seriously restricted in marking the significance of principles in assessment and do not familiarize the users with content on issues such as ethical considerations, fairness, and so on. Hence, it seems that the policies and practices employed do not align with the recommendations in literature [7,15,19,23,27,30]. For instance, as mentioned earlier, Davies [23] believed that it might be a detrimental disservice not to incorporate language assessment principles into the course syllabus. Hence, the misuse or misapplication of tests by unprepared teachers might impact the lives of their learners.

Taken together, it seems that the course students do not receive the training they need to become familiar with the significant role of language assessment and the impact it can have on learning and learners' lives. Hence, they will not be able to competently implement sound assessment and make sound decisions. A logical conclusion to the situation is that teacher candidates do not leave their pre-service education program assessment literate with respect to the assessment principles. More disappointing is that there are no requirements for teacher qualification in assessment literacy in the Iranian educational system. As such, these teachers probably enter the teaching arena without sufficient competence in the assessment essentials. As such, there is a recommendation for more concentration on enhancing pre-service teachers' awareness of what constitutes teacher's LAL and sharpening their understanding of the assessment principles, significantly, to avoid the demerits of unethical practices associated with language assessment.

6. Conclusions

This paper investigated the efficacy of language assessment course in raising the course students' awareness of the principles underlying LAL. Overall, the research results were not in line with the general debates on LAL, contending that principles (e.g., ethics, fairness, washback, etc.) form a key piece of the puzzle.

To address the drawbacks identified, the researchers present a set of recommendations for the better design of language assessment course at state universities which can also be applied to other similar educational contexts. The findings can provide significant

implications for policymakers/syllabus designers to become aware of the nature of language assessment course. In addition, they are recommended to be mindful of the current assessment training needs, prioritize assessment training in principles of assessment, and furnish more opportunities for the course students to receive the training they need in assessment. Hence, modification of policies and practices, with respect to presenting more comprehensive goals and teaching content, is highly advocated. The research also suggests critical implications for policy development to reduce undesirable external factors, such as the course length and class size, which may impede the course students' development of LAL. In addition, the researchers suggest more systematic attention to teachers' LAL on the part of the educational system, for example, by making it as a requirement before their recruitment.

The study also recommends certain critical measures to be taken by the course instructors to enhance the course students' awareness of the essential principles in assessment. First, appropriately selected assessment textbooks can provide the opportunity for the course students to acquire a deep understanding of the various aspects of assessment. Hence, employing more updated textbooks, including the key components of LAL, is stressed. In addition, the course instructors need to inform students about the hazards of unethical practices related to assessment. This may include familiarizing them with published codes of ethics through class discussions of ethical concerns, permitting students to share their learning experiences, and disclosing their assessment-related preoccupations. There are also online forums available that can be introduced to raise awareness of the significant role of assessment principles. Besides, instructors are suggested not only to present explicit instruction on the assessment principles but also highlight the importance of considering principles in assessment by modeling acceptable assessment practices and showing their tendencies and concerns toward observing them in their assessment practices. This is critically important as it is argued in the literature that teachers' conceptions of assessment are typically shaped by their experiences as learners [11,67–69]. Of course, experts, policymakers, and course instructors' cooperation and dialogue are necessary to obtain the requirements of a more quality language assessment course.

Nevertheless, despite the significant points associated with this study, it suffers from a number of limitations. First, it is a small-scale study involving ninety-two questionnaire respondents, which could otherwise provide more assured results and a deeper understanding of the course instructors' practices. Likewise, the observation data of this study only came from two classrooms at two universities. Hence, generalizing the findings should be done with caution. Besides, despite the researchers' attempts to incorporate all underlying aspects of assessment principles into the instruments designed, there might be features that have been left out. Finally, due to the broad area of LAL and to avoid the project getting vast and unwieldy, it was impossible to investigate the efficacy of the language assessment course from different perspectives. Nevertheless, there are still many unexplored areas related to the development of competence in the principles underlying assessment via the course.

As such, findings from the present research call for investigations into the assessment literacy of the course instructors themselves and if they adhere to the principles in their assessment practices. An investigation of the policymakers and course instructors' rationale for their practices and decisions and what may deter them from including the principles of assessment in the course content is also necessary. It is also worth investigating whether instructors with various teaching experiences have different predispositions toward teaching the course. Investigation of the course students' perceptions of the assessment principles, their readiness to implement appropriate assessment, and their assessment training needs concerning assessment principles could also provide more insightful findings. Finally, further research is needed to identify other probable hindering factors to provide more comprehensive information for the design of a better course.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.heliyon.2023.e12857>.

References

- [1] Course Syllabus, Language and Literature, Ministry of Science, Research, and Technology (MSRT), Iran, 2009.
- [2] R.J. Stiggins, Assessment literacy, *Phi Delta Kappan* 72 (7) (1991) 534–539.
- [3] R.J. Stiggins, Assessment crisis: the absence of assessment for learning, *Phi Delta Kappan* 83 (10) (2002) 758–765.
- [4] C. Coombe, H. Vafadar, H. Mohebbi, Language assessment literacy: what do we need to learn, unlearn, and relearn? *Lang. Test. Asia* 10 (1) (2020) 1–16, <https://doi.org/10.1186/s40468-020-00101-6>.
- [5] G. Fulcher, Assessment literacy for the language classroom, *Lang. Assess. Q.* 9 (2) (2012) 113–132, <https://doi.org/10.1080/15434303.2011.642041>.
- [6] G. Fulcher, Operationalizing assessment literacy, in: D. Tsagari (Ed.), *Language Assessment Literacy: from Theory to Practice*, Cambridge Scholars Publishing, 2020, pp. 8–28.
- [7] B. Kremmel, L. Harding, Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: developing the language assessment literacy survey, *Lang. Assess. Q.* 17 (1) (2020) 100–120, <https://doi.org/10.1080/15434303.2020.191674855>.
- [8] M.W. Latif, Exploring tertiary EFL practitioners' knowledge base component of assessment literacy: implications for teacher professional development, *Lang. Test. Asia* 11 (1) (2021) 1–22, <https://doi.org/10.1186/s40468-021-00130-9>.
- [9] G.N. Masters, Reforming educational assessment: imperatives, principles and challenges, in: *Australian Education Review* 57, Australian Council for Educational Research, 2013.
- [10] E. Mohammadkhah, et al., EFL teachers' assessment literacy: a contextualized measure of assessment theories and skills, *Lang. Teach. Res. Quart.* 29 (2022) 102–119, <https://doi.org/10.32038/ltrq.2022.29.07>.
- [11] W.J. Popham, Assessment literacy for teachers: faddish or fundamental? *Theory Into Pract.* 48 (1) (2009) 4–11, <https://doi.org/10.1080/00405840802577536>.
- [12] W.J. Popham, Assessment literacy overlooked: a teacher educator's confession, *Teach. Educat.* 46 (4) (2011) 265–273, <https://doi.org/10.1080/08878730.2011.605048>.
- [13] M. Price, Assessment standards: the role of communities of practice and the scholarship of assessment, *Assess Eval. High Educ.* 30 (3) (2005) 215–230, <https://doi.org/10.1080/02602930500063793>.

- [14] K. Vogt, D. Tsagari, Assessment literacy of Foreign Language teachers: findings of a European study, *Lang. Assess. Q.* 11 (4) (2014) 374–402, <https://doi.org/10.1080/15434303.2014.960046>.
- [15] F. Giraldo, Language assessment literacy and teachers' professional development: a review of the literature, *Profile - Issues Teach. Prof. Dev.* 23 (2) (2021) 265–279, <https://doi.org/10.15446/profile.v23n2.90533>.
- [16] S. Anam, N.V.W. Putri, How literate am I about assessment: evidence from Indonesian EFL pre-service and in-service teachers, *Engl. Rev.: J. Engl. Educ.* 9 (2) (2021) 377–388, <https://doi.org/10.25134/erjee.v9i2.4374>.
- [17] F. Giraldo, D. Murcia, Language Assessment literacy and the professional development of pre-service language teachers, *Colomb. Appl. Linguist. J.* 21 (2) (2019) 243–259, <https://doi.org/10.14483/22487085.14514>.
- [18] S. Lutovac, M.A. Flores, Conceptions of assessment in pre-service teachers' narratives of students' failure, *Camb. J. Educ.* 52 (1) (2022) 55–71, <https://doi.org/10.1080/0305764X.2021.1935736>.
- [19] Z. Tajeddin, M. Khatib, M. Mahdavi, Critical language assessment literacy of EFL teachers: scale construction and validation, *Lang. Test. X* (2022) 1–30, <https://doi.org/10.1177/02655322211057040>.
- [20] O. Inbar-Lourie, Language Assessment literacy, in: E. Shohamy, I.G. Or, S. May (Eds.), *Language Testing and Assessment*, Springer International Publishing, Cham, 2017, pp. 257–270, https://doi.org/10.1007/978-3-319-02261-1_19.
- [21] J. Willis, Assessment for learning as a participative pedagogy, *Assess. Matters* 2 (2010) 65–84. <https://search.informit.org/doi/10.3316/informit.330525572573096>.
- [22] Y. Xu, G.T. Brown, Teacher assessment literacy in practice: a reconceptualization, *Teach. Teach. Educ.* 58 (2016) 149–162, <https://doi.org/10.1016/j.tate.2016.05.010>.
- [23] A. Davies, Textbook trends in teaching language testing, *Lang. Test.* 25 (3) (2008) 327–347, <https://doi.org/10.1177/0265532208090156>.
- [24] O. Inbar-Lourie, Constructing a language assessment knowledge base: a focus on language assessment courses, *Lang. Test.* 25 (3) (2008) 385–402, <https://doi.org/10.1177/0265532208090158>.
- [25] C. Lan, S. Fan, Developing classroom-based language assessment literacy for in-service EFL teachers: the gaps, *Stud. Educ. Eval.* 61 (2019) 112–122, <https://doi.org/10.1016/j.stueduc.2019.03.003>.
- [26] A. Scarino, Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning, *Lang. Test.* 30 (3) (2013) 309–327, <https://doi.org/10.1177/0265532213480128>.
- [27] L. Taylor, Communicating the theory, practice and principles of language testing to test stakeholders: some reflections, *Lang. Test.* 30 (3) (2013) 403–412, <https://doi.org/10.1177/0265532213480338>.
- [28] L. Bachman, A. Palmer, *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*, Oxford University Press, 2022.
- [29] C. Gipps, Developments in educational assessment: what makes a good test? *Assessment in Education: Principles, Pol. Pract.* 1 (3) (1994) 283–292, <https://doi.org/10.1080/0969594940010304>.
- [30] G. Brindley, Language assessment and professional development, in: C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, K. O'Loughlin (Eds.), *Experimenting with Uncertainty: Essays in Honour of Alan Davies*, Cambridge University Press, 2001, pp. 137–143.
- [31] S.J. Chung, Y. Nam, Language assessment literacy of Korean EFL teachers: an investigation of their training experiences and needs, *Mod. Engl. Educ.* 19 (1) (2018) 38–48, <https://doi.org/10.18095/meeso.2018.19.1.04>.
- [32] I. Fitriyah, F. Masitoh, U. Widiati, Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers, *Indones. J. Appl. Linguist.* 12 (1) (2022) 124–134, <https://doi.org/10.17509/ijal.v12i1.46539>.
- [33] M.A. Flores, Feeling like a student but thinking like a teacher: a study of the development of professional identity in initial teacher education, *J. Educ. Teach.* 46 (2) (2020) 145–158, <https://doi.org/10.1080/02607476.2020.1724659>.
- [34] L. Gan, R. Lam, Understanding university English instructors' assessment training needs in the Chinese context, *Lang. Test. Asia* 10 (1) (2020) 1–18, <https://doi.org/10.1186/s40468-020-00109-y>.
- [35] B. Kremmel, et al., Fostering Language assessment literacy through teacher involvement in high-stakes test development, in: D. Xerri, P. Vella Briffa (Eds.), *Teacher Involvement in High-Stakes Language Testing*, Springer International Publishing, Cham, 2018, pp. 173–194, https://doi.org/10.1007/978-3-319-77177-9_10.
- [36] E. Maclellan, Initial knowledge states about assessment: novice teachers' conceptualisations, *Teach. Teach. Educ.* 20 (5) (2004) 523–535, <https://doi.org/10.1016/j.tate.2004.04.008>.
- [37] A.A.L. Mendoza, R. Bernal Arandia, Language testing in Colombia: a call for more teacher education and teacher training in language assessment, *Profile - Issues Teach. Prof. Dev.* 11 (2) (2009) 55–70.
- [38] X. Yan, C. Zhang, J.J. Fan, Assessment knowledge is important, but...: how contextual and experiential factors mediate assessment practice and training needs of language teachers, *System* 74 (2018) 158–168, <https://doi.org/10.1016/j.system.2018.03.003>.
- [39] H. Farhady, K. Tavassoli, Developing a language assessment knowledge test for EFL teachers: a data-driven approach, *Iran. J. Lang. Teach. Res.* 6 (2018) 79–94, <https://doi.org/10.30466/ijltr.2018.120602> (3 Special Issue).
- [40] T. Firoozi, K. Razavipour, A. Ahmadi, The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies, *Lang. Test. Asia* 9 (1) (2019) 1–14, <https://doi.org/10.1186/s40468-019-0078-7>.
- [41] S. Jannati, ELT teachers' language assessment literacy: perceptions and practices, *Int. J. Res. Teache. Educ.* 6 (2) (2015) 26–37.
- [42] M. Nemati, et al., Teachers' writing proficiency and assessment ability: the missing link in teachers' written corrective feedback practice in an Iranian EFL context, *Lang. Test. Asia* 7 (1) (2017) 1–18, <https://doi.org/10.1186/s40468-017-0053-0>.
- [43] H. Divsar, G.M. Jafari Gohar, *Evaluation of the Curricula for TEFL in BA and MA Levels Based on Bloom's Revised Categorization*, Payam-e-Noor University Publications, 2014.
- [44] K. Fathi Vajergah, A Comprehensive model for curriculum needs analysis, *Q. J. Econ.* 57 (2) (2008) 204–221.
- [45] A.M. Hayati, Teaching English for special purposes in Iran: problems and suggestions, *Arts Humanit. High Educ.* 7 (2) (2008) 149–164.
- [46] M. Mehrmohammadi, *Curriculum: Viewpoints, Approaches and Visions*, Astane-Qods-Razavi Publications, 2012.
- [47] A.R. Nasr Esfahani, *Curriculum Planning and Course Design at Universities and Institutes of Higher Education*, Iranian Academic Center for Education, Culture and Research, Isfahan University, 2007.
- [48] H. Soodmand Afshar, H. Movassagh, EAP education in Iran: where does the problem lie? Where are we heading? *J. Engl. Acad. Purp.* 22 (2016) 132–151.
- [49] Joint Committee on Standards for Education Evaluation (JCSEE), *Classroom assessment standards for PreK-12 teachers*. [Kindle version]. <https://www.amazon.com/Classroom-Assessment-Standards-PreK-12-Teachers-ebook/dp/B00V6C9RVO>, 2015.
- [50] M. Chalhoub-Deville, C.A. Chapelle, P.A. Duff, *Inference and Generalizability in Applied Linguistics: Multiple Perspectives vol. 12*, John Benjamins Publishing, 2006.
- [51] B.S. Cypress, Rigor or reliability and validity in qualitative research: perspectives, strategies, reconceptualization, and recommendations, *Dimens. Crit. Care Nurs.* 36 (4) (2017) 253–263.
- [52] D. Baker, *Language Testing: A Critical Survey and a Practical Guide*, Edward Arnolds, 1989.
- [53] H. Farhady, A. Jafarpur, P. Birjandi, *Testing Language Skills: from Theory to Practice*, SAMT Publications, Tehran, 1994.
- [54] J.B. Heaton, *Writing English Language Tests*, second ed., Cambridge University Press, 1998.
- [55] C.J. Weir, C.J. Weir, *Understanding and Developing Language Tests*, Prentice Hall, New York, 1993.
- [56] A. Hughes, *Testing for Language Teachers*, Cambridge university press, 2003.
- [57] T. McNamara, *Language Testing*, Oxford University Press, 2000.
- [58] H.D. Brown, P. Abeywickrama, *Language assessment, in: Principles and Classroom Practices*, Pearson Education, NY, 2004.
- [59] J.D. Brown, *Testing in Language Programs: a Comprehensive Guide to English Language Assessment*, McGraw-Hill, 2014.

- [60] A. Davies, Demands of being professional in language testing, *Lang. Test.* 14 (3) (1997) 328–339, <https://doi.org/10.1177/026553229701400309>.
- [61] A.J. Kunnan, Recent developments in language testing, *Annu. Rev. Appl. Ling.* 19 (1999) 235–253.
- [62] M. Miller, R. Linn, N. Gronlund, *Measurement and Assessment in Teaching*, tenth ed., Pearson, NY, 2013.
- [63] T. McNamara, C. Roever, *Language Testing: the Social Dimension*, Blackwell, 2006.
- [64] Y.G. Butler, Exploring the idea of “assessment with learners” in discussing Language Assessment literacy, *Lang. Teach. Res. Quart.* 29 (2022) 92–101, <https://doi.org/10.32038/ltrq.2022.29.06>.
- [65] M. Taras, Student self-assessment: processes and consequences, *Teach. High. Educ.* 15 (2) (2010) 199–209, <https://doi.org/10.1080/13562511003620027>.
- [66] C.K.Y. Chan, J. Luo, A four-dimensional conceptual framework for student assessment literacy in holistic competency development, *Assess. Eval. High Educ.* 46 (3) (2021) 451–466, <https://doi.org/10.1080/02602938.2020.1777388>.
- [67] J.C. Alderson, *Diagnosing Foreign Language Proficiency: the Interface between Learning and Assessment*, A&C Black, 2005.
- [68] G. Brown, Assessment literacy training and teachers’ conceptions of assessment, in: C. Rubie-Davies, C. Rawlinson (Eds.), *Challenging Thinking about Teaching and Learning*, Nova Science, 2008, pp. 285–302.
- [69] S.M. Quilter, J.K. Gallini, Teachers’ assessment literacy and attitudes, *Teach. Educ.* 36 (2) (2000) 115–131.