

Research Article

Language Assessment Courses: An Investigation of the Course Instructors' Pedagogical and Assessment Practices

Masoomeh Taghizadeh  and Golnar Mazdayasna 

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

Correspondence should be addressed to Golnar Mazdayasna; golnarmazdayasna@yahoo.com

Received 26 February 2022; Accepted 31 March 2022; Published 15 April 2022

Academic Editor: Ayoub Bahnasse

Copyright © 2022 Masoomeh Taghizadeh and Golnar Mazdayasna. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Upon recognizing the significance of assessment literacy as a necessity for modern professional teachers, the quality of assessment courses has been examined from different angles. The present research aimed to examine the nature and functionality of assessment training at Iranian state universities in fostering the development of language assessment literacy (LAL), with a specific focus on the course instructors' pedagogical practices (in terms of instructional materials/content and pedagogical practices) and assessment practices (including both formative and summative assessment). One hundred three course instructors were surveyed through a set of open-ended questions, and two course instructors' pedagogical and assessment practices were observed throughout an educational semester. Based on the content and descriptive statistical analysis of the data, a lack of divergence in the instructors' teaching and assessment practices was observable, with the majority of the instructors adhering to traditional methods, showing no inclination to adopt innovative practices. Lack of a balanced focus on all the core elements of LAL, being theory-laden and disconnected from practice in assessment, and use of dated teaching materials were recognized as factors or inappropriate practices on the part of the course instructors that might hinder the development of LAL in the researched context. Furthermore, a set of key action which might be taken by the course instructors to reduce the obstacles in developing and maintaining LAL has been suggested. Finally, limitations of the study and suggestions for future studies on LAL have been mentioned.

1. Introduction

It is widely claimed that the quality of assessment affects the quality of learning and instruction, and significant gains in students' achievement and their motivation have been reported when assessment is integrated with instruction [1–4].

Upon recognizing such a critical role of assessment in education, assessment specialists and educational researchers have seriously called for teachers to enhance literacy in assessment [4–7]. As such, assessment specialists have argued that it is essential that teachers develop valid assessment expertise to be able to design and conduct quality assessments to achieve high results. Such assessment-linked competencies are labeled as *assessment literacy* (AL) in general education [7] and *language assessment literacy* (LAL) in language studies [8]. Such a knowledge base enables teachers to have a deeper understanding of the intricacies that existed in assessing learning

and helps them make logical choices about the domain or content to be tested and the design of appropriate tasks. Possessing the required knowledge base will also help them make accurate data interpretations and sound decisions. Familiarity with the innovative approaches to assessment (e.g., formative assessment and performance-based assessment) and the various assessment techniques (e.g., alternative forms of assessment, such as self-/peer assessment and portfolio) is being considered as an integral element of the current prevailing practices in many educational settings. Learning how to deal with bias and avoid the negative washback effect of the assessment they conducted has also been recognized as key components of teachers' assessment knowledge base [9].

Despite such concentration on the value of developing assessment literacy, research in the educational context has mainly shown an inadequate level of teachers' assessment literacy, and this has often led teachers to overlook quality

assurance in their activities and associate their assessment practices with traditional approaches and use of poorly designed tests [10, 11]. Hasselgreen et al. [11] investigated the effect of previous training with three types of stakeholders, including teachers, teacher trainers, and testers. The results suggested stakeholders' lack of formal training in assessment and the need for more extensive education. Fulcher [12] explored the training needs of some language teachers using an online survey. Results indicated teachers' need for more training. The findings also revealed serious problems associated with training materials and the requirement for more appropriate materials to expand teachers' level of LAL. Scarino [13] used collaborative dialogues with teachers to find teachers' weaknesses in using assessment. Findings showed that teachers had difficulty on theoretical, practical, and institutional levels. Tsagari and Vogt [14] investigated the language teachers' perceived levels of LAL and their training needs. The results indicated that the teachers' perceived LAL levels were inadequate to address the professional requirements of the field.

Upon the observance of such deficiencies in teachers' assessment practices, the focus on the development of sound assessment literacy by teachers through training programs has been highlighted in the literature [7, 15–17]. Stiggins [7] recommended that teachers be trained by assessment specialists in assessment issues and a monitoring system be applied for the assessment practices of in-service teachers. Such attention is echoed in Esfandiari and Nouri [17], indicating that raising teachers' understanding of assessment issues can facilitate the process of evaluating learners. DeLuca et al. [18] advocated for specifically designed and differentiated programs to respond to the professional training needs of teachers. However, in this regard, according to Looney et al. [19], teachers' identities and conception determine the way they approach assessment. Giraldo [20] points out that an amalgamation of knowledge, skills, and practices of language testing is required for teachers to be able to conduct a quality assessment.

Within this training-supportive perspective, the quality of assessment courses has been examined from different angles. A set of studies have investigated teacher candidates' perceived assessment literacy and their readiness for entering the teaching contexts after finishing a course on assessment. Most of the studies investigating teacher candidates' conceptualizations of assessment indicated them not to be sufficiently prepared for assessing student learning [21, 22]. Some studies investigated the relationship among training on assessment, perceptions toward assessment, and teacher assessment literacy [15, 17, 19, 23]. Some studies, investigating the effectiveness of preservice assessment training, have analyzed the assessment course syllabi [24, 25].

Bailey and Brown's [26], also replicated in [27], was an initiating effort in the analysis of language assessment courses in language education. The study aimed to investigate the topics instructors taught and their students' perceptions of the courses. In their two accounts, Bailey and Brown described and analyzed the significant developments in fostering an acceptable knowledge base in the assessment of language learning. The results of the recent iteration of the

survey [27] revealed that some areas were new, but generally, no dynamic change was observed.

Taken together, preservice quality teacher training programs, especially assessment courses, can ensure the development of AL/LAL for effective classroom-based assessment. Upon understanding the necessity of such training courses, most of today's teacher education programs have added a separate course on assessment to their preservice programs. However, despite such consideration, there are still reports of teachers' insufficient competence in or difficulty with utilizing assessment for learning purposes. As an example of a more recent study, reference can be made to Wu et al.'s [28] investigation of Chinese EFL teachers' perspectives toward formative assessment and the implementation of assessment for learning strategies. Based on the findings, the surveyed teachers did not place particular value on self- and peer assessment strategies and scarcely implemented them in their assessment practices. In a similar vein, Fazel and Ali [29] compared two groups of Canadian and Malaysian teachers' familiarity with learning-oriented assessment at both theoretical and practical levels. The results indicated that the teachers in both groups were, to some extent, acquainted with the concept of assessment for learning; however, on the practical level, they reported a lack of familiarity on how to implement the relevant principles. The authors call for the revision and remodeling of preservice teacher education programs.

A set of barriers and inappropriate practices may limit the development of assessment literacy within the preservice assessment courses. Different reasons are suggested for such deficiencies. Calderhead and Shorrock [30] attribute the divergence in methodology and especially the choice of course teaching materials to varying perspectives existed within the field of teaching and teacher education. Some studies have debated that textbooks used mostly emphasized traditional assessment issues, ignoring the more current critical aspects of assessment [5, 6, 9]. Kleinsasser [31] investigated language assessment courses from the instructors' points of view. Kleinsasser found the failure of bridging theory to practice as the major obstacle in teaching a language assessment course. The researcher pointed out that instead of saturating teachers' theoretical and statistical knowledge, they be empowered and promoted through dialogue and negotiation of assessment practices. Popham [2] and Yan et al. [32] also criticized the teaching content for being theoretical and irrelevant to classroom assessment practices.

In the Iranian EFL educational context, a set of similar studies have claimed the in-service teachers' poor knowledge base in assessment [33–35] or lack of readiness to implement innovative assessment methods [35]. Generally, these studies have indicated language teachers' lack of expertise in implementing appropriate assessment. Based on these findings, it may be concluded that in the Iranian EFL educational programs, preservice programs, which are mainly in the form of language assessment courses, have failed to prepare skilled teachers for the effective implementation of assessment. Hence, based on the literature reviewed, speculation might be that the course instructors' practices are not aligned with the requirements for fostering preservice teachers' LAL. As such, the present research aimed to examine the quality of

language assessment courses through an exploration of the course instructors' teaching methodologies (in terms of instructional materials/content and pedagogical practices) and assessment methodologies (in terms of both formative and summative assessment).

In the B.A. English language syllabus, defined by the Ministry of Science, Research and Technology (MSRT), a two-credit compulsory module on assessment and testing has been introduced for language students at state universities. The general syllabus of the course, including the number of sessions and time allocation, is introduced by MSRT; however, the course instructors are those who decide on the teaching materials, instructional procedures, assessment practices, and related issues. Language graduates usually enter the teaching context in private language institutes or public schools.

The study addressed the following two questions:

- (1) To what extent are the course instructors' teaching methodology (in terms of instructional materials/content and pedagogical practices)?
- (2) To what extent are the course instructors' assessment methodology aligned with the requirements for developing sound LAL?

2. Materials and Methods

As suggested in the literature, an amalgamation of methods and producing results from different angles would increase the validity of the results and contribute to findings which are more trustworthy and generalizable [36]. As such, two research methods, i.e., a questionnaire and classroom observation, were employed for the investigation of the research questions which set this particular research. The rationale for selecting each particular instrument, along with the critical aspects of design, data collection, and analysis stages, will be presented in each section.

2.1. Open-Ended Questionnaire. Questionnaire has been recognized as a suitable instrument for this particular research as it facilitates the collection of self-report practices of a group of instructors. Some concerns have been made for the possible weaknesses of using a questionnaire. For instance, Dörnyei (2003) discusses that questionnaires have often been misused, "partly due to the misconception that questionnaires can act as a rapid remedy to restrictions in time and other resources which researchers may be forced to contend with" (p. 9). Despite such arguments, questionnaires have been widely used in research conducted in different fields.

As such, for the present research, it was decided to use a questionnaire to collect the reports of a group of course instructors on their teaching and assessment methodologies. There were some considerations related to the type and format of the questions to be asked. Finally, it was decided to include open-ended questions to allow for flexible responses. The questions are listed below:

- (1) What textbook(s)/materials do you require your students to read for the course?

- (2) Explain your method of teaching the course and the learning activities you use briefly
- (3) Explain your method of assessing students learning (throughout and at the end of the course) briefly

2.1.1. Participants. One hundred three course instructors, including 64 holding Ph.D, 31 Ph.D candidates, and 8 MA graduates, returned the completed questionnaire. These participants were teaching at various state universities from different geographical locations within Iran.

2.1.2. Procedure. The questionnaire was delivered by e-mail, through personal visits, or via social networks. Respondents were asked to fill in the questionnaire if they had taught the course within the last three years. Instructors were not informed about the exact purpose of the study; however, they were assured about the confidentiality of their responses.

Participants were requested not to produce extended explanations; consequently, the responses provided could be more efficiently coded and grouped into categories. The results of this section were analyzed using content analysis, which enabled the identification of certain themes. Categorization was done by two coders (one of the researchers and her colleague) to check for the reliability of the coding. The initial analysis ensured a substantial kappa coefficient of agreement ($k = .86$). Double checking and more intercoder discussions were used to reach a consensus on the points of disagreement.

2.2. Observation. Questionnaires are suitable devices to get a general understanding of the efficacy of language assessment courses in developing LAL among learners; however, to get a deeper insight into what really happens in such courses, a field study was required. A series of potential qualitative research instruments were considered when deciding on a research method that could be used for collecting data for this particular research project. Finally, classroom observation was recognized as a suitable choice for the research. Observation gives the researcher the opportunity to see closely what teachers and learners are doing in the classroom rather than having to rely on what they say they do [36].

Observations were conducted to specify instructors' teaching methodology (in terms of the teaching materials/content and the pedagogical practices) and their assessment practices (considering both formative and summative assessment of the students learning).

2.2.1. Participants. Teaching practices of two language assessment course instructors, holding Ph.D. in TEFL, were observed throughout one academic semester. Both instructors had teaching experiences of more than ten years.

2.2.2. Procedure. To outline the instructors' pedagogical and assessment practices, an observational scheme was required to be designed. The designed observation scheme included three sections. The first section included information, such as class and session number and starting and ending time, filled out by the observer at the beginning of the session.

Section two and three, which explored instructors' pedagogical and assessment approaches and practices, consisted of a set of Likert scale items for which the observer needed to tick whether the variable was observed *extensively*, *moderately*, *a little*, or *not at all*. For sections two and three, some space was provided for taking field notes on the processes, situations, interactions, and tasks/activities. In addition, audio-recording of the classroom events were used as a supplementary technique for later checks. After-class informal talks with the instructors were also employed to probe into the instructors' explanations for using special activities/practices, in case of ambiguity. A post hoc rating scale coding procedure was used after the observed session, through which decisions were made on the frequency of each variable of consideration along the scale chosen. In addition to the pedagogical practices, quizzes, mid-term, and final exam sheets were also analyzed to determine the degree to which teachers' assessment practices were in congruent with new perspectives in teaching and assessment. The courses were surveyed for 15 sessions at two universities, which were accessible to the researcher assigned as the observer (i.e., convenience sampling). The instructors were kept uninformed of the purpose of the research. The researcher's consent with data protection, ethical considerations, and anonymity of the participants was assured.

The validity and reliability issues had to be considered in analyzing the observation data to confirm the accuracy and consistency of the findings. In qualitative research, reliability is associated with multiple coding to ensure the accuracy of the data. To address the reliability, notes were taken and, with the instructors' permission, sessions were audio-recorded, and inter-rater reliability (one of the researchers and her colleague as the coders) was carried out to determine if the two coders made the same coding decisions. When the initial coding procedure was completed, the Kappa coefficients of .85 coding agreement was obtained for sections two and three. Follow-up discussions were held between the two coders, and another colleague's consultation was sought to come to an agreement on the areas of discrepancy.

In qualitative studies, validity is defined in terms of rigor, which refers to the strength of the research design, the appropriateness of the methods, the accuracy of the data obtained, and the truthfulness of the analyses of the data and the decisions made [37]. In addition to ensuring the quality of the research design and the data collection methods, nonparticipatory role and self-awareness techniques were taken to avoid the researchers' personal perspectives affecting the results of the research and to minimize bias and enhance the validity. For the validity of the analyses and the decisions made, as mentioned above, colleagues and expert consultation was also sought.

Decisions were made, based on the Likert Scale, on the extent to which each variable related to teaching and assessment practices was observed. Independent *t*-test was used to compare the means for each dimension across the classes, and the Man-Whitney test was used to compare the distribution of each variable within each dimension.

3. Results

The findings in this section will be grouped according to each research instrument employed, initially presented for the questionnaire, followed by the classroom observations.

3.1. Questionnaire

3.1.1. Teaching Methodology

(1)*Instructional Materials/Content*. One purpose of the questionnaire was to outline the prevalent teaching sources employed by the course instructors. A total of six textbooks were listed. Some respondents utilized more than one textbook. Figure 1 displays the sources mentioned, along with the percentage of the participants using each particular textbook. Self-designed materials were also reported by 10% of the instructors. Two textbooks seemed to dominate the courses: (a) Farhady et al. [38], used with a percentage larger than that of all the other textbooks reported (82%), and (b) Heaton [39], with a percentage of 46. Four other textbooks [40–43] were utilized each with a percentage of less than 35.

(2)*Pedagogical Practices*. Through the second question, instructors briefly explained their method of instruction and the different learning activities they employ to help course students develop a valid level of LAL. Notably, 54% of the respondents reported using mainly lecturing for classroom instruction, and 28% claimed a combination of instructor lecture and student presentation on an assigned topic. Task-based learning, in the form of having students work by themselves or together to design an assessment instrument, was reported by 11%. Tasks involving assessment design, administration, and interpretation of results were used by only 4% (see Figure 2).

3.1.2. *Assessment Methodology*. The third open-ended question explored the instructors' assessment procedures and practices. Formal assessment in the form of final examination seemed to be the typical method of assessment as reported by 96% of the respondents, with mid-term exam almost enjoying similar position (reported by 88% of the instructors). Ninety-five percent of the instructors devoted a proportion of the total score to students' classroom attendance and participation in class activities. Some of the instructors (40%) had quizzes throughout the course of instruction or assigned a term project (20%). Assessing learning based on performance on a set of practical tasks was favored by 11% (see Figure 3).

3.2. *Observation*. Observations provided a more accurate outline of the course practices, including the content of teaching, method of instruction and delivery, and the assessment procedures used by the instructors.

A total of 932 minutes for class A and 1062 minutes for class B were recorded, excluding the greeting time and talks on unrelated topics. The distribution for the variable of time was approximately normal in each class. Results showed that the mean of time in class A (63.10 ± 6.09) and class B

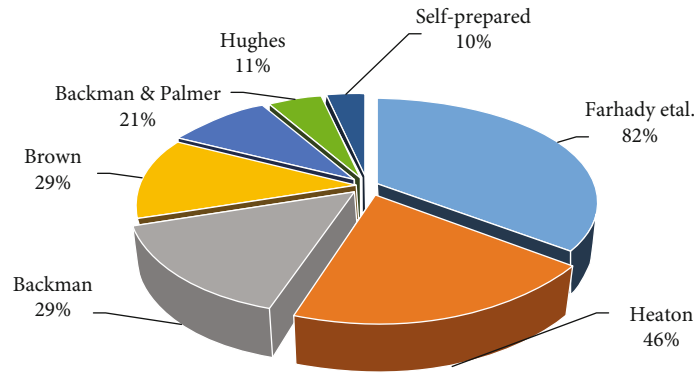


FIGURE 1: Teaching materials selected by the course instructors.

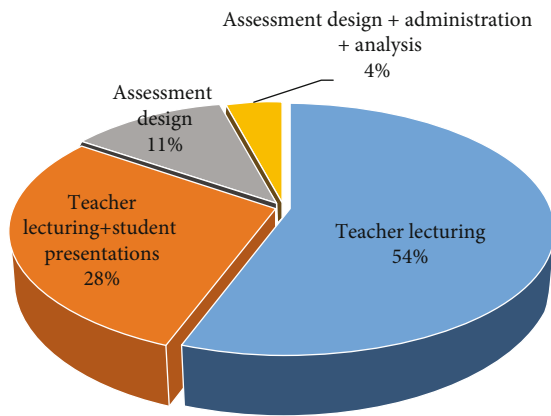


FIGURE 2: Teaching/learning activities reported by the course instructors.

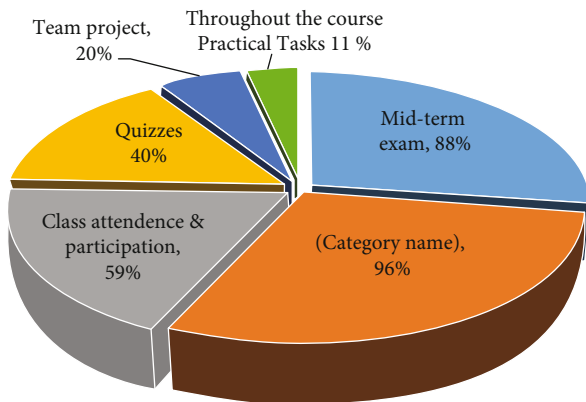


FIGURE 3: Assessment procedures reported by the instructors.

(68.13 ± 9.34) was not significantly different (p value = 0.353) (see Table 1).

A general outline for both classes is presented in Table 2, with respect to the teaching mode and end of the course evaluation of the students' achievement. Following, the findings are presented with respect to the three subsections.

3.2.1. Teaching Methodology

(1) *Instructional Materials/Content.* The common and main teaching textbook used by both instructors was Farhady et al. [38], which focuses dominantly on the knowledge and skills components of assessment (see Table 3 for the general outline of the topic covered). Although the instructor of class A, in addition to the mentioned textbook, used a more updated, renowned textbook (i.e., Brown [41], he still limited its use to four chapters, which dealt mainly with the knowledge and skills components). With respect to the knowledge component of assessment, the main focus in both classes was on traditional approaches and types of assessment (e.g., subjective/objective tests, high-stakes tests, achievement tests, norm-referenced tests (NRT), etc.). Criterion-referenced tests (CRT) was also covered to some extent in class A. In addition, in both classes, and to a greater degree in class B, students delivered lectures, which were mainly on aspects related to the knowledge component. The skills areas which were covered in both classes were related to the principles of item writing (e.g., multiple-choice, true/false, and matching items), analysis of item characteristics, graphic representation of test results, testing language skill/sub-skills, cloze/dictation tests, and strategies to estimate test reliability and validity. The content on the assessment principles was limited to the discussions of the reliability, validity, and practicality qualities in assessment.

(2) *Pedagogical Practices.* Section two in the observation scheme explored instructors' pedagogical activities and strategies with respect to a set of variables. As Table 4 indicates, generally, no significant difference was observed concerning the instructors' teaching activities and practices (mean differences = 0.44 ± 0.02 , p value = 0.214). Concerning the individual variables, significant differences were observed in the case of (a) allowing learners share their assessment-related experiences and concerns and (b) suggesting additional reading resources, with the instructor of class A paying more attention to these aspects. Assigning appropriate assessment-related tasks and tailoring instruction based on learners' needs were not considered noticeably by either instructor. Using various modes of instruction were, to some extent, taken into account by

TABLE 1: Comparing the mean difference of time between classes and components.

Dependent variable	Group	N	Mean \pm SD	Mean difference \pm Std.error	95 CI of difference	ANOVA, <i>p</i> value
Time	Class A	15	63.10 \pm 6.09	-6.03 \pm 2.75	0.13, 11.92	0.353
	Class B	15	69.13 \pm 9.34			

TABLE 2: General outline of the classes concerning teaching mode and end of the course evaluation.

Class	Prevalent classroom practices	End of the course evaluation
Class A	Teacher lecturing/student lecturing/class discussion/writing test items (mainly multiple-choice)	Classroom attendance&participation = 2 points 2 quizzes (paper&pencil multiple-choice format) = 3 points Mid-term exam (paper&pencil open-ended items) = 5 points Final exam (paper&pencil open-ended&multiple-choice items) = 10 points
Class B	Teacher lecturing/student lecturing/class discussion	Classroom attendance&participation = 3 points Mid-term exam (paper&pencil open-ended&multiple-choice items) = 6 points Final exam (paper&pencil open-ended&multiple-choice items) = 11 points

TABLE 3: A general outline of the topic covered in both classes.

Class	Teaching materials	Content covered
Class A	Farhady, et al. [38], chps 1-14 Brown [41], chps 1,2,4, and 5	Introduction to testing Different types of tests (mainly traditional assessment types, including objective/subjective, achievement tests, high-stakes/low-stakes, etc.) Functions of language tests (prognostic tests/selection Tests/placement tests/aptitude tests proficiency) NRT/CRT The structure of an item/ Classification of item forms Holistic/analytic approach to scoring NRT item analysis Tabulation of data Descriptive statistics (measures of central tendency & variability) Different steps to the process of testing Interpreting test scores Concept of reliability and validity and strategies to estimate them Testing language skills/subskills Concept of practicality in assessment Cloze and dictation type tests
Class B	Farhady, et al. [38] Chps1-15	Importance of testing/decision making/test, measurement and evaluation Functions of language tests (prognostic tests/selection Tests/placement tests/aptitude tests proficiency) Classification of item forms (selected/open-ended items) Structure of an items Principles for constructing selected-choice/matching items Tabulation of data Basic statistics in testing (measures of central tendency & variability) Different steps to the process of testing Concept of reliability and validity and strategies to estimate them Testing language skills/subskills Theories of language testing (discrete-point approach/integrative approach/functional approach) Problems with discrete-point/integrative tests Cloze and dictation type tests

both instructors; however, it was limited to teacher and student lecturing and class discussions (see Table 5).

3.2.2. *Assessment Methodology*. The purpose of section three was to investigate the assessment practices of the participat-

ing instructors throughout the course of instruction. As indicated in Table 2, formal assessment practices involved paper-and-pencil quizzes, mid-term, and final exams in open-ended and/or multiple-choice formats. Concerning the assessment practices employed by the instructors

TABLE 4: Comparing the mean of pedagogical approach between classes using independent *t*-test.

Dimension	Class	<i>N</i>	Mean \pm SD	Mean difference \pm Std.error	95% CI of mean difference	<i>p</i> -value
Pedagogical approaches	Class A	7	1.48 \pm 0.46	0.44 \pm 0.02	-2.04, 1.25	0.214
	Class B	7	1.04 \pm 0.48			

TABLE 5: Comparing the distribution of variables related to teaching methodology between class A and B.

Variables	Class	Mean \pm SD	Median (Q1, Q3)	Min, max	<i>p</i> -value*
Instructor uses multiple modes of instruction (i.e., small group discussion, practical works, and student lecturing)	Class A	1.73 \pm 0.46	1.00 (1.00, 2.00)	1, 2	0.104
	Class B	1.52 \pm 0.41	1.00 (1.00, 1.00)	1, 2	
Instructor allow students share their assessment-related concerns and experiences	Class A	2.60 \pm 0.51	2.00 (2.00, 3.00)	2, 3	<0.001
	Class B	1.93 \pm 0.46	2.00 (1.00, 1.00)	1, 2	
Instructor investigates students' learning needs	Class A	0.80 \pm 0.68	1.00 (0.00, 1.00)	0, 2	0.053
	Class B	0.58 \pm 0.52	0.00 (0.00, 1.00)	0, 1	
Tasks are assigned effectively to have practice in various aspects of assessment	Class A	0.93 \pm 0.70	1.00 (1.00, 2.00)	0, 2	0.084
	Class B	0.61 \pm 0.51	0.00 (0.00, 1.00)	0, 1	
Instructor suggests additional resources to facilitate student learning	Class A	1.23 \pm 0.64	1.00 (1.00, 2.00)	0, 2	<0.00
	Class B	0.60 \pm 0.63	1.00 (0.00, 1.00)	0, 2	

* The exact Man-Whitney test.

TABLE 6: Comparing the mean of assessment approaches between classes using independent *t*-test.

Dimension	Class	<i>N</i>	Square mean \pm SD	Mean difference \pm Std.error	95% CI of mean difference	<i>p</i> -value
Assessment approach	A	5	3.53 \pm 1.30	1.00 \pm 0.43	0.11, 1.89	0.029
	B	5	2.53 \pm 1.06			

TABLE 7: Comparing the distribution of variables related to assessment practices between classes.

Variables	Class	Mean \pm SD	Median (Q1, Q3)	Min, max	<i>p</i> -value*
The instructor implemented diagnostic assessment at the beginning or end to adjust subsequent instruction	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	
Students' understanding is checked throughout the instruction	Class A	1.27 \pm 0.56	1.00 (1.00, 2.00)	0, 2	0.026
	Class B	0.87 \pm 0.74	1.00 (0.00, 1.00)	0, 2	
The instructor provides feedback based on the assessment he/she does	Class A	1.40 \pm 0.74	1.00 (1.00, 2.00)	0, 2	0.038
	Class B	0.73 \pm 0.68	1.00 (0.00, 1.00)	0, 2	
The instructor involved alternative assessment (self/peer/whole class, portfolios, etc., in evaluation process)	Class A	0.60 \pm 0.51	1.00 (0.00, 1.00)	0, 1	0.153
	Class B	0.33 \pm 0.49	0.00 (0.00, 1.00)	0, 1	

* The exact Man-Whitney test.

throughout the instruction, a set of variables (i.e., use of diagnostic assessment to adjust instruction, checking students' understanding throughout the instruction, feedback provision, and use of various types of alternative assessment) were taken into consideration. Generally, the mean for class A was significantly higher than that in class B (mean difference = 1.00 \pm 0.43, *p*-value = 0.029) (see Table 6). However, significant differences were limited just to two of the variables under consideration (checking students' under-

standing throughout the instruction and providing feedback based on the assessment). No significant differences were observed for other assessment-related aspects (Table 7).

The use of alternative assessment types, such as observations and portfolios, was lacking in both classes. The extent to which the instructors incorporated diagnostic assessment to adjust instruction was also focused. As the table shows, not many instances of diagnostic assessment were observed in either class.

4. Discussion

4.1. On the Teaching Methodology. The first research question explored the pedagogical approaches and practices used by the course instructors. The first two open-ended questions in the questionnaire and data from section two in the observation scheme provided information on the teaching practices of the instructors. The teaching methodology of the instructors is discussed with respect to the instructional materials/content and the pedagogical practices employed.

4.1.1. Instructional Materials/Content. Generally, to foster the theoretical background of the prospective teachers in assessment and testing, content and teaching materials play a pivotal function. Utilizing appropriate assessment textbooks is essential as they have the potential to equip course students with the required in-depth competence in all necessary aspects of the assessment process and facilitate the process of developing assessment literacy [9].

Based on the questionnaire data, the two key textbooks reported to be used by the majority of the instructors were Farhady et al. [38] and Heaton [39], with both dating back to more than 25 years ago. While both sources arguably aim to provide an introductory function for trainee language teachers, they do not place considerable emphasis on some important aspects of assessment. More importantly, these textbooks are drastically limited in encompassing new perspectives and approaches in assessment. In addition, the courses are strongly ineffective in stressing the value of principles in assessment and testing. That is, content including ethical considerations, fairness, test washback, and similar issues is not presented. However, it seems essential for these programs to incorporate principles as they have become pivotal due to the function and impact of assessment in society [12, 20, 44, 45].

Moreover, assessment design aspects such as test specification, designing scoring rubrics, different scoring methods, and some other skill-related aspects have been likewise neglected. Besides, these textbooks do not provide useful exercises that trainee teachers need to work on. Nonetheless, as Davies [9] recommends, teaching materials should remark both theoretical and practical aspects. The observation data also confirmed the findings from the questionnaire. Concerning the two courses observed, whereas the content varied to some extent, the data suggested almost consistent tendencies on the part of the instructors at the theoretical level, with both focusing more on knowledge and skills in assessment. Findings documented by the classroom observations highlighted that instructors predominantly provided students with information on traditional assessment, such as objective tests or psychometric properties of the tests and showed less tendency toward innovative assessment types.

Hence, the findings confirmed the reports in the studies indicating the use of inappropriate teaching materials [5, 6, 9]. Generally, the lack of variation in the textbooks employed and the limitation of topics were noticeable in the data obtained through both the questionnaire and observations, despite the availability of many textbooks written by signifi-

cant scholars in the field. Instructors might select such textbooks as they might believe they cover what they think is of primary important in assessment.

Concerning the use of educational assessment textbooks, it is suggested that care should be taken on the part of users not to adopt dated textbooks, which to a great extent only focus on psychometric aspects of educational assessment and summative assessment of learning [5, 6, 9, 46]. Davies [9] comments that what is missing in most textbooks is a focus on the principles of assessment. Masters [5] criticizes preservice teacher education programs for treating assessment literacy at a superficial level, attributing the problem to the choice of unsound assessment textbooks. Taylor [47] also comments that the majority of teachers find most of the available textbooks too much technical to be absorbed or applied in practice as they are written from the perspective of research academics and measurement specialists rather than teachers.

Nevertheless, as some scholars have argued, although textbooks serve as a cornerstone to support LAL, other resources such as online content can be employed to foster LAL [9, 45, 48, 49]. However, in the studied context, the textbook(s) assigned by the instructors seemed to be the only source of knowledge the course students were exposed to.

4.1.2. Pedagogical Practices. At the practical level, programs planned to develop teachers' LAL should certainly give special attention to having practice in assessment [8, 9, 12, 20, 50].

As mentioned earlier, without adequate LAL levels, teacher candidates cannot have the knowledge and confidence to select, design, or implement practical language assessment, make legitimate decisions, and differentiate between good and bad language testing practices. Accordingly, to enter their classrooms with the knowledge and confidence they need, preservice training programs should be provided. However, as Malone [16] asserts, mere training is inadequate for teacher candidates to respond to the language assessment needs. It can be surmised that a course with a specific focus on practical tasks in the assessment of students' learning can better contribute to teacher candidates' assessment literacy and should be pivotal in teacher training programs. This perspective is also reiterated by Jeong [15], who believes that teachers tend to see language assessment from a practice-based view; for example, "when it comes to statistics, teachers seem to want clarity and practical examples rather than abstract notions" (p.371). Particularly, it seems that, through a practice-focused methodology, teachers also get familiar with the conceptual aspects, which are part of the knowledge component in LAL (Malone, [49]).

Based on the present research findings, it seems that the courses do not credit the design of language assessments. In the rare examples of the practices observed, students were engaged with constructing item types linked to traditional assessment. According to the results, confirming the findings in Kleinsasser [31] and Yan et al. [32], the programs appeared to have considered LAL as conceptually driven rather than design-based. Use of task-based learning could bring dramatic outcomes; yet, theory-only courses do not have the potential to help the achievement of the required

levels of LAL or even be based on teachers' real training needs. In short, it may be inferred that these courses have not addressed the needs of current assessment requirements.

4.2. On the Assessment Methodology. Another purpose of the study (research question 2) was to investigate the types of assessment approaches/practices utilized by the course instructors to evaluate students learning. The third open-ended question in the questionnaire explored the instructors' assessment practices. Classroom observations also provided information on the assessment practices of the two instructors.

As mentioned in the introduction section, in current assessment perspectives, formative view of assessment for learning and the use of alternative types of assessment is seen as paramount.

In the present study, findings indicated the instructors' adherence to traditional assessment practices as the instructors mainly seemed to treat summative test and the end-product as the norm for assessment. Data showed that traditional approaches, including teacher-led assessment activities and focus on paper-and-pencil tests, were dominant in the courses, and these instructors did not take into account changes that favor formative assessment. The types of assessment mainly being utilized were selected-response assessments (mostly multiple-choice items) and constructed-response assessments (e.g., gap-filling and questions requiring extended written answers). The same results were obtained through both the observations and the survey. It can be concluded that the assessment practices in these courses can be characterized as being traditionally oriented and teacher-directed, not enjoying ongoing, formative, and learner-centered classroom assessment.

The problem with using tests/exams is that they usually yield minute information about students' progress. On the contrary, performance-based, portfolio, and self- and peer-assessments have been prioritized in current assessment practices as they have the potential to furnish multidimensional information on students' learning. Besides, traditional assessment suffers from low face validity as they do not involve meaningful and real-life knowledge and skills needed for everyday lives [51]. Contrary to the traditional assessment, innovative, or alternative assessment measures higher-order thinking skills and broader course objectives [51, 52]. However, considering the dual role of classroom assessment, it is recommended that instructors balance the formative and summative assessment approaches to assessment, avoid the mere use of monodimensional assessment types, and utilize assessment for collaborative construction of meaning [6]. As such, decisions based solely on the results of traditional assessment methods should be forewarned as they might incorrectly narrow the learning objectives of the program [6].

A critical point in this regard is that it is very vital that teacher educators model acceptable practices in assessment throughout the program. As echoed in the literature, when teachers have not undergone adequate training on how to assess students' learning, make decisions, and report the results, they begin to assess their learners as they were assessed throughout their education. That is, the assessment experience teachers have had as learners might influence

their perceptions and practices in assessment [14, 53–55]. For instance, Green and Stager [53] indicated that teachers' previous learning experience (i.e., as students) influenced their beliefs about classroom tests. Bandura [54] also showed that the assessment experience (i.e., as students) with traditional assessment (e.g., tests) led teachers to resort to such methods. According to Bandura, this type of learning is highly influential on teachers' future performances.

Consequently, EFL teacher candidates will focus on the course instructors' assessment practices and shape their literacy in language assessment by their own assessment experiences in teacher education programs. As such, students in the present research who were encountered by traditional assessment methods, as the prevalent type of assessment during their B.A. education, will probably resort to traditional assessment in their future teaching experiences, which can be detrimental to their learners.

5. Conclusion

The present research analyzed the pedagogical and assessment practices of the language assessment course instructors at state universities to determine if they are likely to foster the course students' development of sound LAL.

Although this research did not establish that students lack LAL, findings raised questions about the quality of the courses in helping them develop their LAL. A set of barriers and inappropriate practices on the part of the course instructors were identified, which may limit the development of students' LAL within the studied context. Based on the most frequently used course books reported, lack of balanced attention to all core elements of LAL was outstanding. Instructor' choice of the teaching materials and the content analysis of the two classrooms observed implied that the instructors primarily concentrated on teaching knowledge and skills of assessment, giving little importance to teaching assessment principles. Besides, instructors' classroom teaching and assessment practices were not aligned with new trends in the field, with the evident inclinations toward traditional approaches. Being theory laden and disconnected from practice in assessment, use of dated teaching materials, and employing inappropriate assessment procedures were also recognized as other factors which might hinder the development of LAL in the researched context.

All things considered, the language assessment courses studied here need a radical shift in various aspects. The study, identifying the deficiencies of the courses, highlights certain vital actions to be implemented by the course instructors for the more effective design of the courses to promote students' LAL.

A major suggestion in literature is that theory should be connected to practice in assessment. As such, the researcher emphasizes teachers' need for a deeper focus on practical, hands-on training, i.e., development of the skills side of LAL through engagement in assessment practice. Hence, design-based learning should be fundamental in these courses.

Another way of equipping the course students with valid assessment literacy is through the course instructors' employing efficient pedagogy. Research suggests that assessment

literacy can be expanded through teachers' self-reflection and/or their engagement in dialogue with colleagues on their assessment practices and assessment-related issues [56, 57]. In the case of assessment courses, this could also be achieved through engaging students in self-/peer assessment and collaborative learning within the context of practice. Specially, through team works, students have the opportunity to share their personal beliefs about different aspects of assessment, participate in dialogues with their classmates on developing quality assessment tasks, and discuss the problems they encounter in their assessment practices. Such collaborative learning in the context of practice can also activate their theoretical underpinnings of the assessment process.

Most significantly, a quality assessment course, in itself, entails the use of valid assessment procedures. As such, course instructors can have an influential role in helping students' development of LAL by modeling appropriate assessment practices. Hence, they should not just provide explicit instruction but also model acceptable assessment by demonstrating more predisposition toward using more innovative assessment methods, being more conservative about ethical issues in their assessment practices, etc.

Moreover, employing more updated textbooks, comprising the balance of the core components of LAL, is suggested. Besides, course books should provide more comprehensive coverage of formative and summative assessments and the specific aspects, such as assigning scores, selecting and constructing item specifications, developing rubrics, interpreting assessment results, and issuing of washback and fairness. Besides, introducing further study resources can enlighten students' perceptions and practices of assessment.

Finally, mere adherence to published textbooks is not sufficient as students need to be updated with respect to the latest knowledge base, research, and innovations in educational assessment. Students need to be exposed to the relevant research to assessment literacy development that is user-friendly, accessible, and easily absorbed. Hence, introducing practitioners to appropriate internet-based resources and published research is advocated. Online video tutorials and forums can raise students' awareness in different assessment-related aspects.

Therefore, the results can be beneficial to course instructors as they may disclose the strong and weak aspects of the courses and help them evaluate their preferences and practices. Of course, policymakers, university, and departmental cooperation are essential to improve the quality of the instructors' assessment practices. Hence, there should be interactions between these stakeholders to achieve the requirements of a comprehensive language assessment course. Finally, it should be pointed out that although this research has been conducted in a particular educational field, the results can make a special contribution to the overall understanding of assessment literacy development in other teacher education programs.

Addressing the limitations of the present study, more research is needed to give us a deeper insight into how to make the courses more efficient through surveying more course instructors and observing a greater number of classrooms. In line with this study, future research is needed to explore the course instructors' conceptualization of LAL

and the reasons behind their actions. An investigation of the course students' perceived/actual LAL levels, their training needs, and their readiness to perform assessment-related tasks is also suggested.

Data Availability

The data used to support the findings of this study are available from the first author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors would like to express their sincere gratitude to all the course instructors who participated in the study.

References

- [1] J. Gardner, *Assessment for learning: a compelling conceptualization*, Assessment and Learning, 2006.
- [2] W. J. Popham, "Assessment literacy for teachers: faddish or fundamental?," *Theory Into Practice*, vol. 48, no. 1, pp. 4–11, 2009.
- [3] W. J. Popham, "Assessment literacy overlooked: a teacher educator's confession," *The Teacher Educator*, vol. 46, no. 4, pp. 265–273, 2011.
- [4] M. Price, "Assessment standards: the role of communities of practice and the scholarship of assessment," *Assessment & Evaluation in Higher Education*, vol. 30, no. 3, pp. 215–230, 2005.
- [5] G. N. Masters, *Reforming educational assessment: imperatives, principles and challenges*, 2013.
- [6] W. J. Popham, *Classroom Assessment*, 2014.
- [7] R. J. Stiggins, "Assessment literacy," *The Phi Delta Kappan*, vol. 72, no. 7, pp. 534–539, 1991.
- [8] O. Inbar-Lourie, "Constructing a language assessment knowledge base: a focus on language assessment courses," *Language Testing*, vol. 25, no. 3, pp. 385–402, 2008.
- [9] A. Davies, "Textbook trends in teaching language testing," *Language Testing*, vol. 25, no. 3, pp. 327–347, 2008.
- [10] J. C. Alderson, *Diagnosing Foreign Language Proficiency. The interface between learning and assessment*, A&C Black, 2005.
- [11] A. Hasselgreen, C. Carlsen, and H. Helness, *European Survey of Language and Assessment Needs. Part One: General Findings*, vol. 22, 2004/<http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>.
- [12] G. Fulcher, "Assessment literacy for the language classroom," *Language Assessment Quarterly*, vol. 9, no. 2, pp. 113–132, 2012.
- [13] A. Scarino, "Language assessment literacy as self-awareness: understanding the role of interpretation in assessment and in teacher learning," *Language Testing*, vol. 30, no. 3, pp. 309–327, 2013.
- [14] D. Tsagari and K. Vogt, "Assessment literacy of foreign language teachers around Europe: research, challenges and future prospects," *Papers in Language Testing and Assessment*, vol. 6, pp. 41–63, 2017.

- [15] H. Jeong, "Defining assessment literacy: is it different for language testers and non-language testers?," *Language Testing*, vol. 30, no. 3, pp. 345–362, 2013.
- [16] M. Malone, "Training in language assessment," *Encyclopedia of language and education*, vol. 7, 2008.
- [17] 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.
- [18] C. DeLuca, D. LaPointe, and U. Luhanga, "Teacher assessment literacy: a review of international standards and measures," *Education, Assessment, Evaluation, and Accountability*, vol. 28, no. 3, pp. 251–272, 2016.
- [19] 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 & Practice*, vol. 25, no. 5, pp. 442–467, 2018.
- [20] F. Giraldo, "Language assessment literacy: implications for language teachers," *Professional Development*, vol. 20, no. 1, pp. 179–195, 2018.
- [21] L. Volante and X. Fazio, "Exploring teacher candidates' assessment literacy: implications for teacher education reform and professional development," *Canadian Journal of Education*, vol. 30, pp. 749–770, 2007.
- [22] S. L. Wise, L. Lukin, and L. L. Roos, "Teacher beliefs about training in testing and measurement," *Journal of Teacher Education*, vol. 42, no. 1, pp. 37–42, 1991.
- [23] S. M. Quilter and J. K. Gallini, "Teachers' assessment literacy and attitudes," *The Teacher Educator*, vol. 36, no. 2, pp. 115–131, 2000.
- [24] C. Deluca and A. Bellara, "The current state of assessment education," *Journal of Teacher Education*, vol. 64, no. 4, pp. 356–372, 2013.
- [25] J. Greenberg and K. Walsh, *What Teacher Preparation Programs Teach about K-12 Assessment: A Review*, National Council on Teacher Quality, 2012.
- [26] K. M. Bailey and J. D. Brown, "Language testing courses: what are they?," *Language Testing*, vol. 25, no. 3, 1996.
- [27] J. D. Brown and K. M. Bailey, "Language testing courses: what are they in 2007?," *Language Testing*, vol. 25, no. 3, pp. 349–383, 2008.
- [28] X. Wu, L. Zhang, and H. R. Dixon, "Implementing assessment for learning (AFL) in Chinese university EFL classes: teachers' values and practices," *System*, vol. 101, article 102589, 2021.
- [29] I. Fazel and A. M. Ali, "EAP teachers' knowledge and use of learning-oriented assessment: a cross-contextual study," *System*, vol. 104, article 102685, 2022.
- [30] J. Calderhead and S. B. Shorrock, *Understanding Teacher Education*, 1997.
- [31] R. C. Kleinsasser, "Transforming a postgraduate level assessment course: a second language teacher educator's narrative," *Prospect*, vol. 20, no. 3, pp. 77–102, 2005.
- [32] X. Yan, C. Zhang, and J. J. Fan, "Assessment knowledge is important, but how contextual and experiential factors mediate assessment practice and training needs of language teachers," *System*, vol. 74, pp. 158–168, 2018.
- [33] E. Babaii and F. Asadnia, "A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice," *Reflective Practice: International and Multidisciplinary Perspectives*, vol. 20, no. 6, pp. 745–760, 2019.
- [34] H. Farhady and K. Tavassoli, "Developing a language assessment knowledge test for EFL teachers: a data-driven approach," *Iranian Journal of Language Teaching Research*, vol. 6, no. 3, pp. 79–94, 2018.
- [35] T. Firoozi, K. Razavipour, and A. Ahmadi, "The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies," *Language Testing in Asia*, vol. 9, no. 1, pp. 1–14, 2019.
- [36] Z. Dörnyei, *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*, Oxford University Press, 2007.
- [37] B. S. Cypress, "Rigor or reliability and validity in qualitative research," *Dimensions of Critical Care Nursing*, vol. 36, no. 4, pp. 253–263, 2017.
- [38] H. Farhady, A. Jafarpur, and P. Birjandi, "Testing language skills: from theory to practice," *Tehran: SAMT Publications*, 1994.
- [39] J. B. Heaton, *Writing English Language Tests*, Longman, 1988.
- [40] L. F. Bachman, *Fundamental Considerations in Language Teaching*, Oxford University Press, 1990.
- [41] J. D. Brown, *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*, McGraw-Hill, 2014.
- [42] L. F. Bachman and A. Palmer, *Language Testing in Practice*, Oxford University Press, 1996.
- [43] A. Hughes, *Testing for Language Teachers*, Cambridge University Press, 2nd ed. edition, 2003.
- [44] T. F. McNamara and C. Roever, "Language Testing: The Social Dimension," *International Journal of Applied Linguistics*, vol. 16, no. 2, 2006.
- [45] O. Inbar-Lourie, *Language assessment literacy*, Language testing and assessment: Encyclopedia of language and education, 2017.
- [46] C. Campbell and V. L. Collins, "Identifying essential topics in general and special education introductory assessment textbooks," *Educational Measurement: Issues and Practice*, vol. 26, no. 1, pp. 9–18, 2007.
- [47] L. Taylor, "Developing assessment literacy," *Annual Review of Applied Linguistics*, vol. 29, pp. 21–36, 2009.
- [48] F. Giraldo, "Language Assessment Literacy and Teachers' Professional Development: A Review of the Literature," *Profile: Issues in Teachers' Professional Development*, vol. 23, 2021.
- [49] M. E. Malone, "Training in language assessment," in *Encyclopedia of language and education*, Encyclopedia of Language and Education, E. Shohamy, I. G. Or, and S. May, Eds., Springer, Cham, 3rd edition, 2017.
- [50] O. Kvasova, *Will a bloom lead to a bloom? Or how to secure a launch of language assessment literacy in Ukraine*, Language assessment literacy: From theory to practice, 2020.
- [51] M. D. Miller, R. L. Linn, and N. E. Gronlund, *Measurement and Assessment in Teaching*, Pearson, 11th edition, 2013.
- [52] E. Maclellan, "Initial knowledge states about assessment: novice teachers' conceptualisations," *Teaching and Teacher Education*, vol. 20, no. 5, pp. 523–535, 2004.
- [53] K. E. Green and S. F. Stager, "Measuring attitudes of teachers toward testing," *Measurement and Evaluation in Counseling and development*, vol. 19, no. 3, pp. 141–150, 1986.
- [54] A. Bandura, *Self-Efficacy: The Exercise of Control*, W. H. Freeman and Company, New York, 1997.
- [55] M. A. Siegel and C. Wissehr, "Preparing for the plunge: pre-service teachers' assessment literacy," *Journal of Science Teacher Education*, vol. 22, no. 4, pp. 371–391, 2011.

- [56] L. Adie, "The development of teacher assessment identity through participation in online moderation," *Assessment in Education Principles, Policy & Practice*, vol. 20, no. 1, pp. 91–106, 2013.
- [57] M. D. Howley, A. Howley, J. E. Henning, M. B. Gilla, and G. Weade, "Intersecting domains of assessment knowledge: school typologies based on interviews with secondary teachers," *Educational Assessment*, vol. 18, no. 1, pp. 26–48, 2013.