

## TEACHERS' LANGUAGE ASSESSMENT LITERACY: A CASE STUDY AT A HIGH SCHOOL IN HALONG CITY, VIETNAM

Thu Ha Bui

University of Western Australia, Australia; [buihuha2981998@gmail.com](mailto:buihuha2981998@gmail.com)

### ABSTRACT

This study sets out to explore high school teachers' perceptions of language assessment literacy (LAL) and their training needs on language assessment. In Vietnam, several issues concerning teachers' assessment practices have been identified, namely exam-oriented teaching, lack of expertise in test design, insufficient academic knowledge and professional training in language assessment (Hoang, 2017; Vu, 2017). The present study built on Davies's (2008) definition of LAL which includes knowledge and skills and Giraldo's (2018) dimensions of LAL. The data was collected via a qualitative approach as five English teachers from a high school in Ha Long were invited to participate in semi-structured interviews. The results revealed that teachers' knowledge of approaches, theories and concepts relevant to language assessment were insufficient and inaccurate. In terms of skills, several major findings indicated an emphasis on traditional testing over alternative methods; tendencies to design tests based on former experiences, personal instincts and colleagues' advice; and increased use of technology-based assessments. Further studies should investigate teachers' actual assessment practices and language assessment training needs in teacher education programs.

**Keywords:** *language assessment literacy; language testing and assessment; training needs*

### 1. INTRODUCTION

In the field of second language education, assessment is one of the most critical domains whose theory and practice continue to mature over time. Given the interconnectedness between teaching and assessment, it makes sense that an equivalent amount of attention is devoted to developing teachers' knowledge, understanding and practices of assessment - that is language assessment literacy (LAL).

Despite increased importance accorded to LAL on a global scale, it is widely observed that many teachers feel ill-prepared to execute and select a wide array of classroom assessments (Mertler & Campbell, 2005). In Vietnam, there is a strong line of evidence that corroborates a mismatch between expected assessment practices from teachers and their actual conducts. Overall, according to Duong (2016), teachers' performance of assessment is highly traditional and uninnovative, with assessment dominantly and rigidly focusing on discrete points of language knowledge and merely developing learners' cognitive abilities at low levels.

Specifically, an overwhelming majority of Vietnamese public high school teachers are heavily reliant on paper-based tests to evaluate their students' English proficiency. The tests themselves mostly touch upon lexis and grammar, while leaving out other essential competencies that allow successful communication to occur (Hoang, 2010). This undue emphasis on lexical and grammatical testing might negatively transform English teaching and learning into tips and tricks practice as well as prompting negative attitudes from students towards the subject (Hoang, 2010). Other than the test-based approach to assessment, several insensible assessment practices used by Vietnamese high school teachers have also been noted. According to Nguyen (2013), grades were often administered to students without feedback. A limited number of teachers made the effort to do so, which unfortunately could not add up to facilitate students' progress due to its insufficient or unconstructive nature.

These considerations lay the foundation for extensive research into English language teachers' LAL in Vietnam. After having thoroughly examined relevant materials in LAL, the researcher has detected the following gaps in the literature review. First and foremost, there is a dearth of prior research or publication on LAL in the context of Vietnam, since the majority of the existing studies have dealt with language assessment in general instead of teachers' language assessment literacy in particular. Specifically, Vietnamese teachers' perceptions of language assessment have not been discussed at length, whereas teachers' identified training needs for language assessment and testing have been absent from a larger part of available literature (Nguyen, 2013).

The present study seeks to investigate how teachers perceive their LAL in their teaching practices. It is also designed to find out their perceived training needs in terms of language assessment. Hence, this research purports to address the following question:

*What are the perceptions of English teachers at a high school in Halong of their LAL?*

## **2. LITERATURE REVIEW**

### **2.1. Defining LAL**

One of the earliest attempts to define LAL came from Paterno (2001) when he stated that LAL refers to a working knowledge about fundamental principles of well thought out assessment practices, which encompasses terminology, development and use of methods and techniques, and familiarity with quality standards in assessment. However, this definition, derived primarily from the concept of assessment literacy, was rather general as it made no reference to the language aspect.

Meanwhile, according to Davies (2008), LAL is defined as a combination of knowledge of languages and skills of language testing. Specifically, *knowledge* is related to that of language and language methods such as communicative language testing and task-based assessment. Meanwhile, *skills* refer to those necessary for test design, use and interpretation of assessment data and test evaluation.

The dynamic relationship among the two mentioned components was well-

established in Davies's (2008) study. Knowledge offers a context setting for skills (Davies, 2008). For instance, in order to perform a skill such as item writing, test designers are expected to possess adequate knowledge of their assessment contexts and purposes of the written items.

Therefore, considering its comprehensiveness in covering all the relevant factors, its credibility given that Davies's global view of LAL is generally accepted by authors (Fulcher, 2012) and its suitability for serving as a basis for a thorough investigation into language assessment, this definition was used throughout this research.

## 2.2. Framework

Building on the definition of Davies's (2008) study, Giraldo (2018) proposed a list of five dimensions of LAL classified under the two central components of knowledge and skills. Each of the five dimensions entails a comprehensive list of descriptors which detail what language teachers are expected to know and do.

| <b>LAL components</b> | <b>Dimensions</b>   |
|-----------------------|---|
| Knowledge             | 1. Awareness of theory and concepts<br>2. Awareness of own language assessment context          |
| Skills                | 3. Instructional skills<br>4. Design skills for language assessments<br>5. Technological skills |

First and foremost, two dimensions namely awareness of theory and concepts and awareness of one's own language assessment context constitute *knowledge*. Within *knowledge*, theory and concepts are considered principal concerns given that they are directly related to the language-related aspect - a distinctive feature of LAL (Inbar-Lourie, 2013a). An awareness of the educational context (e.g.: local assessment policies) for language assessment is based on Scarino's (2013) study. *Skills* include three dimensions namely instructional skills, design skills and technological skills. Instructional skills, as described in the studies by McNamara and Hill (2011), refer to those such as using assessment methods or providing feedback. Design skills are concerned with test and item construction (Fulcher, 2012; Taylor, 2009). Lastly, technological skills, as suggested by Davies (2008) and Inbar-Lourie (2013a), are those that form teachers' use of technology in their assessment practices.

These descriptors are of great value in enabling teachers to evaluate their contexts of language assessment, and thus Giraldo's (2018) dimensions of LAL will be used in this research. Nevertheless, as some descriptors for one dimension may overlap those for other dimensions and a number of descriptors themselves require the acquisition of specific descriptors in advance, the provided descriptors would be adapted in accordance with the researcher's own context.

### 3. METHODOLOGY

#### 3.1. Research design

The study employed a qualitative approach via interviewing. Given that the primary focus of this research was to investigate teachers' perceptions of their LAL along with their training needs in language assessment, the qualitative approach was the most appropriate method at hand to allow the researcher to gain a sophisticated understanding of her research context.

#### 3.2. Research context

The research took place at a high school in Halong. Along with Chinese, English is one of the two foreign languages taught in this school. English lessons are designed and carried out in keeping with Dispatch No. 5333/ BGDĐT-GDTrH.

#### 3.3. Participants

Five English teachers from a high school in Halong were selected for the study based on convenience sampling. A brief introduction about their academic qualifications and teaching experience is as follows:

| Teacher | Academic degree (highest)                | Teaching experience |
|---------|--|---------------------|
| T1      | MA in English Teaching Methodology       | 14 years            |
| T2      | MA in TESOL                              | 11 years            |
| T3      | BA in English Language Teacher Education | 7 years             |
| T4      | MA in Linguistics                        | 13 years            |
| T5      | MA in TESOL                              | 14 years            |

#### 3.4. Instrument

In response to the two research questions, the baseline interview guide was developed using Giraldo's (2018) dimensions of LAL. The list of descriptors were adapted in accordance with the research context as the interview questions were developed. In total there were 7 questions, whose aim was to collect data about teachers' perceptions of their LAL which would correspond with the two components of language assessment literacy – knowledge and skills (Davies, 2008). A brief summary of the interview questions for part I is as follows:

| Category   | Questions                                    |
|--|--|
| <b>Knowledge</b><br>Awareness of theory and concepts<br>Awareness of language assessment context       | Question 1<br>Question 2                     |
| <b>Skills</b><br>Instructional skills<br>Design skills for language assessment<br>Technological skills | Questions 3-4<br>Questions 5-6<br>Question 7 |

### **3.5. Procedures**

#### **3.5.1. Data collection**

Invitation letters were sent to five teachers to seek their consent in participating in the research. Once the teachers agreed, agreements were made to interview them. The interview time and place were decided at the participants' convenience. Each interview generally lasted for nearly 30 minutes and was carried out in Vietnamese. At the beginning of each interview, the researcher delivered a brief summary of her study and outline of the interview scheme, as well as assuring that the interviewee's participation would be totally voluntary and that all of the answers would be kept anonymous. The interviews were subsequently voice-recorded provided that the participants' consent had been obtained.

#### **3.5.2. Data analysis**

The data analysis method for interviews was inductive thematic analysis. Transcribing and coding were required to analyze the available data using thematic analysis. For transcribing, as the language used in the interviews was Vietnamese, translation from Vietnamese into English was an additional step expected of the researcher. Also, in order to ensure the participants' anonymity, their names in this research were coded as numbers (T1, T2, T3, T4 and T5).

As the interviews were analyzed based on the inductive thematic analysis, firstly, the researcher reviewed the transcript of each interview to generate an initial list of codings. Similar codes were sorted into recurring themes. The next step was to review and refine identified themes. A final list of themes was then produced and analyzed in response to the research question. Expected findings were teachers' knowledge, skills and principles in language assessment.

## **4. FINDINGS AND DISCUSSION**

### **4.1. Knowledge**

#### **4.1.1. Awareness of theory and concepts**

#### **Insufficiency and inaccuracy in interpretations of major qualities for assessment**

All five participants were not able to recall the major qualities for assessment (reliability, validity, interactiveness, practicality, washback and authenticity) sufficiently and accurately. Participants considered **practicality** and **reliability** as two major qualities which need to be borne in mind when they conduct assessments. In further discussion on their interpretations of the chosen qualities, T1 defined practicality as the connection between assessment and the language knowledge that students have learned, whereas T2 suggested that practicality means that tasks should be based on local and school contexts. T2 also referred to reliability as teachers' ability to closely follow the matrix when designing a test and take their job seriously as examination supervisors.

It was suggested that practicality in the definition offered by T1 and T2 was associated with authenticity, which is defined as a task's ability to simulate real-life communication (Brown, 2005). In addition, T2's recollection of the term reliability was similar to the notion of fairness in the sense that it pointed to the role of test developers and administrators in achieving language fairness assessment (Spaan, 2000).

Overall, it is clear that all five teachers failed to grasp the basic concepts behind major qualities for assessment, which is in line with the studies by López Mendoza and Bernal Arandia (2009) which indicated teachers' lack of knowledge of assessment fundamentals such as reliability and validity. That they could not recall the exact terms related to assessment qualities might be put down to the fact that high school language teachers often pay more attention to what they are doing in the classroom than what they need to know about the theories behind assessment (Hoang, 2017). Similarly, as Alderson (2004) suggested, theoretical underpinnings of assessment are often presented in a way that only takes account of researchers as their main audience while dismissing those like teachers who may not take much of an interest in concepts and theories. Thus, it is justifiable why teachers often feel overwhelmed by abstract discussions on vague terms in language assessment (Alderson, 2004, p. 1).

#### ***4.1.2. Awareness of own language assessment context***

All five studied teachers agreed that guidelines and policies issued by the Ministry of Education and Training (MOET) have considerably influenced their assessment conducts.

#### **Influence of guidelines and policies issued by the MOET**

With regard to guidelines set by the MOET, T1 emphasized two dispatches namely Dispatch No. 5333/ BGDĐT-GDTrH and No. 3333/ BGDĐT-GDTrH that were directly shaping teachers' assessment practices. Specifically, T5 recalled several guidelines including but not limited to the implementation of periodical tests in line with the curriculum plan; the inclusion of all four skills in the final term tests; the insertion of constructed response in tests for grade 10 and 11 and only MCQs in those for grade 12.

#### **Criticism of washback assessments on teaching context**

There were two teachers who raised concern about the mismatch between the expected outcome of high school students that is an overall command of all four language skills and the National High School Graduation Examination's exclusion of speaking, listening and writing. Specifically, it was worth noting from the assessment results that most students were reluctant to learn all four skills and instead paid full attention to what would be tested, as indicated by T2. Similarly, T5 affirmed that students excelled at grammar, vocabulary and reading, which were often included in national exams, while listening and speaking were English specialized students' strengths only.

Their critical comments on the negative washback effects imposed by the MOET's guidelines are aligned with the view of Damankesh and Babaii (2015) who contended that a testing culture may give rise to teachers' and students' optimal focus on areas

of knowledge which will be tested, thus shifting the focus of the curriculum from communicative competences to vocabulary and grammar.

## **4.2. Skills**

### **4.2.1. Instructional skills**

#### *4.2.1.1. Types of assessment*

##### **Prioritization of traditional testing**

All of the respondents reported traditional testing to be a predominant part of their assessment practices. At high schools, traditional testing consisted of oral tests, 15-minute tests, 45-minute tests and term tests. Specifically, according to T4, oral testing could be conducted via a short Q&A between teacher and student, or a conversation between two students who had been given a situation in advance and asked to work in pairs to provide a dialogue in response. 15-minute tests were normally used to evaluate what students have learned in the previous lesson, typically after a language section, and might vary from written tests to oral tests. Meanwhile, 45-minute and term tests were in strict conformity with the guidelines and policies provided by the MOET.

##### **Lack of practice of alternative assessment methods**

Limited use of alternative methods was found among almost all of the sampled teachers. Except for T1 who also adopted role-play and presentation, all five participants opted for project-based assessment as their primary alternative method.

The reason why project-based assessment was highly preferred among the respondents was mainly attributable to skill development offered by doing projects. T2 and T4 expressed their confidence in the possibility that projects can motivate students to not only acquire knowledge from textbooks but also soft skills like acting in a play, presentation, teamwork, etc.

Overall, the current findings present evidence in agreement with previous studies which highlighted how teachers continue to follow the traditional route of giving tests, whereas the demand for variations of types of assessment is left unsatisfactorily fulfilled (Pham et al., 2019; Nguyen, 2013; Hoang, 2017; Vu, 2017). One probable explanation for the overuse of paper-based testing might lie in the test-driven culture in Vietnam (Ho, 2013). Application of alternative assessment, as a consequence, comes at the expense of the prioritization of traditional testing, not to mention the fact that high school teachers are not well-trained to carry out this type of assessment (Nguyen, 2013).

While it is true that traditional testing is the main actor in the language assessment and testing arena, the presence of alternative assessments such as project assessment should nevertheless be dismissed. In this regard, the results draw a distinction from the findings by Le (2015) who argued that the use of project work or performance-based assessment was nowhere to be found. The discrepancy in the findings might be attributable to different sampling populations, given that the study by Le (2015) was conducted on a non-public university, hence resulting in different curriculum goals and structures.

#### 4.2.1.2. Feedback

Peer feedback and teacher feedback were used frequently by all five participants, particularly with respect to project-based assessment.

##### **Use of peer feedback**

T4 explained that in the beginning, teachers would carefully instruct students on the evaluation criteria. Before a group delivered a presentation, evaluation criteria sheets were distributed to other groups. Students were asked to pay full attention to other groups' presentations and write down the comments. After each presentation, other groups gave feedback including the strengths of and limitations of the assessed group's performance.

Three participants pointed out the usefulness of peer feedback in their interviews. T1 believed that this introduced a proper way to include students in the assessing process, and their feedback could be used as a reference for teachers to decide on the final results. T4 commented on peer feedback as "very useful" in the sense that students would learn from the strengths and weaknesses of their peers' performances. Furthermore, peer feedback would help students develop critical thinking and self-confidence as well as increasing their engagement in their peers' presentations.

Teachers' reasons for employing peer feedback are similar to previous views regarding the merits of reviewing peers' projects, which include facilitating students' reflection and improvement of their own performance, development of critical thinking and increase in active engagement to articulate their knowledge of the matters being discussed (Liu & Carless, 2006).

##### **Use of teacher feedback**

Teacher feedback, meanwhile, was closely based on assessment criteria as for all five teachers. A common set of assessment criteria that could be drawn from their interviews might entail presentation skills, content delivery, level of interaction with the audience, use of technology (e.g.: PowerPoint, font size, visual aids, etc.) and time limit.

In addition, the nature of teacher feedback was claimed to be constructive. It was unanimously agreed upon that teacher feedback should aim at encouraging students rather than criticizing them. In order to do so, teachers would comment on the strengths of students' performances first, followed by limitations and suggestions. To illustrate this point, T4 gave a specific instance of how constructivity was incorporated into her feedback:

*T4: I divided my feedback into 2 things: what I like best is and what you need to improve. The things that my students are good at would be mentioned first. The things that they need to work on will be presented in a subtle manner so that they would not feel hurt and be motivated to work harder.*

Sampled teachers' attempts to provide constructive feedback give support to a range of theories and concepts in the matter of giving feedback. It is crucial that students upon receiving feedback should feel positive about it (Piccinin, 2003). By emphasizing

good performance, identifying poor performance and suggesting plans for improvements, language teachers are more likely to inspire and promote student learning.

It is interesting to note that the finding of teachers basing their feedback specifically on marking rubrics runs counter to that of the previous study by Pham et al. (2019) which pointed out teacher feedback being too general. Meanwhile, the evidence of participants articulating their feedback in a constructive fashion also presents a striking difference from the demotivating nature of teacher feedback as proposed by Nguyen (2013).

#### *4.2.2. Design skills*

##### **Search for ready-made sources**

Three out of five teachers started their test design process by looking for reliable sources with intended topics that they could readily apply. In the case of T5, she commended the quality of Oxford-published materials or Cambridge-published books. She preferred sources like these in which theories are accompanied by practice, including review practice, progress test and task revision. Thus, sources in which tests are logically organized would spare her the need to cut or edit the tests.

Inclination toward immediate adaptation of available sources in test design is confirmed by Nguyen (2013). According to Nguyen (2013), high school language teachers are inclined to collect and mimic sample tests without engaging in critical reflection upon their own assessment context or the objectives of the tests. Participants' use of ready-made sources might be pinned down to a lack of confidence in the validity of teacher-made tests, which prompted them to opt for internationally published sources with high credibility (Berry et al., 2019).

##### **Lack of association between test design and major qualities for assessment**

The relationship between major qualities for assessment and test design proved to be controversial and complex. Although major qualities were acknowledged in teachers' awareness of theory and concept on language assessment, actual practices of such qualities in test design were seriously limited.

All five participants found it incredibly challenging to apply all the assessment qualities in test design. Rather, they were heavily reliant on their own intuition, personal experiences and colleagues' professional advice. T3 confirmed her complete oversight of assessment qualities when designing 15-minute tests. She further noted that those qualities were only considered when she developed periodical or mock tests for the National High School Graduation Examination. Similarly, T4 asserted there was no tool available to gauge the validity of one's test. Instead, she built up her ideas of how the test should look primarily through her observation of students' learning. In addition, T1 proposed asking for professional advice from her colleagues to ensure qualities for assessment, specifically reliability or validity.

Overall, it is conclusively demonstrated that the connection between test design and major qualities for assessment was frequently neglected due to teachers' reliance on their former experiences, professional instincts and staffroom knowledge sharing.

The result is comparable to that of the study by Nguyen (2013) who observed that teachers usually design tests based on their previous experiences without reflecting upon theoretical foundations of test design. In a similar vein, Sultana (2019) found that expertise in test design was loosely formed by teachers' intuition to write and grade items, on-the-job learning and suggestions from their colleagues.

#### *4.2.3. Technological skills*

Four out of five sampled teachers were reported to apply technology in testing and assessment. Their use of technology could be categorized into three major types: using smart classrooms, creating tests on online platforms, and assessing text difficulty.

##### **Use of smart classrooms**

T1 stated that in support of assessing students' capabilities, 13 smart classrooms at school were at her disposal. A software called ActivInspire was installed in advance, which would assist teachers in their assessment conducts, usually in the form of MCQs and short answers. T2 emphasized that such technological tools were brought to use for 15-minute tests rather than 45-minute tests. She explained that paper-based testing would be a safer choice as conducting 45-minute tests in smart classrooms was not time-efficient, given that teachers might have to deal with equipment, which could be distracting for students.

One important note to consider is that according to T1, young teachers applied technology in their assessments more often than their senior colleagues. This observation is in line with the finding by Vu (2017) who reported that young teachers are more predisposed towards innovative assessments.

However, using smart classrooms in testing and assessment is not without its downsides. T2 highlighted the fact that absence of Wifi would pose an obstacle to using interactive tablets, not to mention that utilizing smart classrooms was time-consuming, taking up on average an hour per lesson. These concerns might be associated with investment in facilities required for technology-enhanced assessment as well as staff time and training on how to use new technologies efficiently.

##### **Creation of tests on online platforms**

T3 and T4 recalled using SHub Classroom, while the latter added Google Forms as the platforms on which she could administer tests to students. Conducting online tests could guarantee immediate results without the need for teachers to grade the answers. Besides, T3 revealed that by doing tests online at home, students were able to make up for their previous scores. One obvious limitation of taking online tests, accordingly, was that students might discuss their answers with each other, which would compromise the ethics of testing.

Overall, the findings mark a visible improvement in teachers' technological skills in assessing their students, given that teachers' skills in using softwares to measure students' English competences were reported to be poor (Pham, 2013). Nevertheless, there were a number of barriers to a wider adoption of technology in assessments

including technical infrastructure, staff training and ethics. Furthermore, it should be noted that the teachers' use of technology in assessments was primarily limited to assessment of learning. Once again, the test-oriented culture among Vietnam's public high schools could be a possible explanation to justify teachers' disinterest in technology in assessment for learning (e.g.: improving classroom instructions (Chan, 2018).

## **5. CONCLUSION**

### **5.1. Summary of the main findings**

*Question: What are the perceptions of English teachers at a high school in Halong of their LAL?*

#### **1.1.1. Knowledge**

With regard to **awareness of theory and concepts**, inaccurate and insufficient interpretations of major qualities for assessment were found. As regards **awareness of language assessment context**, teachers confirmed that their assessment practices were decisively influenced by guidelines and policies issued by the MOET, which drew criticism for negative washback effects on teaching and learning.

#### **1.1.2. Skills**

##### **Instructional skills**

Traditional testing was found to be prioritized over alternative assessment methods, most notably project based assessment. Different deliveries of feedback involved peer feedback, which was reported to be highly useful, and teacher feedback which was primarily based on assessment criteria and constructively curated.

##### **Design skills**

With reference to **assessment methods**, selected response, particularly MCQs, was used the most frequently. A central focus on MCQs was attributable to the orientation towards large-scale and high-stakes tests, especially the National High School Graduation Examination.

In terms of test design, it was revealed that teachers preferred to adapt ready-made tests. Major qualities for assessment were rarely taken into account in the process of test design as teachers mostly turned to their own intuition, personal experiences and colleagues' professional advice for guidance.

##### **Technological skills**

Increased use of technology was classified as three major types including using smart classrooms and creating tests on online platforms, while being predominantly focused on assessment of learning.

### **5.2. Implications**

The findings are hoped to offer strong theoretical and practical implications on the topic of LAL. Theoretically, the present study could provide a snapshot of English teachers' LAL using Davies's (2008) definition of LAL and Giraldo's (2018) dimensions of LAL. Specifically, a systematic and comprehensive illustration of Vietnamese language teachers' knowledge and skills is given, which is expected to enrich the existing body of

research on LAL in the context of Vietnam. Practically, the findings might present an ideal opportunity for English teachers to reflect upon their assessment practices.

### **5.3. Limitations**

Despite the researcher's best efforts, there are certain shortcomings of the present study that must be taken into consideration. First of all, the analysis of the results was performed with the assumption that participants were completely truthful in their answers. As Hammersley and Gomm (2008) suggested, interviewee's perceptions about a subject matter might be limited to what they were willing to reveal, therefore the accuracy of their responses might be questionable. Furthermore, due to time constraints and convenience, this study adopted interviewing as the only instrument. While interviews allow the research to explore the participants' perceptions at length, interviewing alone may not elicit sufficient data on the subject matter given that the available data are subject to both the interviewee's faulty memory and the interviewer's personal interpretations (Potter & Hepburn, 2005). In addition, considering that the research was conducted on five English teachers at a high school only, the findings might not be generalizable to all high schools in Vietnam and thus should be used with caution.

### **5.4. Suggestions**

In response to the limitations in the present study, several suggestions are made for future research into LAL in the context of Vietnam. While teachers' perceptions about their assessment practices remain a significant part of the existing scholarship into LAL, teachers' actual in-class assessment presents a potential area of future research. As there is obviously scope for further investigation into teachers' real-life practices of assessment in comparison to their perceptions, observation may serve as a useful check on and supplement to the data from the interviews. Also, as small sample sizes might interfere with the scalability of a study's findings, a bigger sample size consisting of multiple schools would allow for the results to be generalized. Last but not least, future studies could direct their attention to the population of teacher initial education with the aim of probing into specific needs on language assessment training (e.g.: mode of training, delivery, content, etc.). Given that pre-service education plays a key role in teachers' academic and professional development, language assessment in teacher education training is a pertinent issue worthy of further investigation.

## **REFERENCES**

- Alderson, J. C. (2004). The Shape of Things to Come: Will it be the Normal Distribution? (In) *European Language Testing in a Global Context Proceedings of the ALTE Barcelona Conference July 2001*. M. Milanovic, C. Weir, & S. Bolton (Eds.). Cambridge: CUP. (pp. 1-26).
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *English Language Teaching Journal*, 73(2), 113-123. <https://doi.org/10.1093/elt/ccy055>

- Brown, J. D. (2005). *Testing in language program*. New York: McGraw-Hill.
- Chan, K. K. (2018). The Effect of Teachers' Perceptions on the Role of Technology in Assessment: the Case of Macau. *International Journal of Learning, Teaching and Educational Research*, 17(2), 127–137. doi: 10.26803/ijlter.17.2.8
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, 45, 62-69.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3): 327-347.
- Duong, M. T. (2016). The Modern Assessment Paradigm and the Methods to Assess Students' Competences in Vietnam's Education. *VNU Journal Of Science: Education Research*, 32(1). Retrieved from <https://js.vnu.edu.vn/ER/article/view/666>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- Giraldo, F. (2018). Language assessment literacy: implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195.
- Hammersley, M. & Gomm, R. (2008). Assessing the radical critiques of interviews. In: M. Hammersley, (Ed.), *Questioning Qualitative Inquiry: Critical Essays* (pp. 89-100) London: Sage.
- Hoang, V. V. (2017). The 2016 national matriculation and general certificate of secondary education English test: A challenge to the goal of foreign language education in Vietnamese schools. *VNU Journal Of Foreign Studies*, 33(2). doi:10.25073/2525-2445/vnufs.4138 <http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>.
- Hoang, V. V. (2010). The Current Situation and Issues of the Teaching of English in Vietnam. *Ritsumeikan Language and Culture Research*, 22(1), 7–18. Retrieved from [http://www.ritsumei.ac.jp/acd/re/k-rsc/lcs/kiyou/pdf\\_22-1/RitsIILCS\\_22.1pp.7-18\\_HOANG.pdf](http://www.ritsumei.ac.jp/acd/re/k-rsc/lcs/kiyou/pdf_22-1/RitsIILCS_22.1pp.7-18_HOANG.pdf)
- Inbar-Lourie, O. (2013a). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923–2931). Oxford: Blackwell.
- Le, T. V. (2015). An Investigation into the Effectiveness of Learning Assessment for Non-English Major Students at the Tertiary Level. *IJSELL* 3(5), pp:14-31.
- Liu, N. -F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290.
- López-Mendoza, A. A., & Bernal-Arandia, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teacher' Professional Development*, 11(2), 55-70.
- McNamara, T., & Hill, K. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395-420.

- Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the "Assessment Literacy Inventory." *Online Submission*. Retrieved from <http://files.eric.ed.gov/fulltext/ED490355.pdf>
- MOET (2014). *Dispatch No. 5333/BGDĐT-GDTrH: Implementing language testing and assessment in order to develop learners' English proficiency at secondary level from academic years 2014-2015*. Hanoi.
- MOET (2016). *Dispatch No. 3333/BGDĐT-GDTrH: Using test formats to evaluate students' English proficiency at primary, secondary and high school levels from academic years 2015-2016*. Hanoi.
- Nguyen, K. C. (2013). Innovations in assessing high school students using a competency-based approach. <https://doi.org/shorturl.at/qJNSW>
- Paterno, J. (2001). Measuring success: A glossary of assessment terms. In *Building cathedrals: Compassion for the 21st century*. Retrieved July 24, 2003, from <http://www.angelfire.com/wa2/buildingcathedrals/measuringuccess.html>
- Piccinin, S. J. (2003) *Feedback: Key to learning*. Halifax, NS: Society for Teaching and Learning in Higher Education.
- Potter, J. & A. Hepburn. (2005). Qualitative interviews in psychology: Problems and possibilities. *Qualitative Research in Psychology*, 2: 281–307. <http://dx.doi.org/10.1191/1478088705qp045oa>
- Pham, A. T. N., Nguyen, N. V. Q., & Nguyen, T. H. H. (2019). Vietnamese EFL Teachers' Classroom Assessment Practice at the Implementation of the Pilot Primary Curriculum. *International Journal of Language and Linguistics*, 7(4), 180. doi: 10.11648/j.ijll.20190704.15
- Pham, H. Q. (2013). *Development of teacher training programs, theoretical and practical issues*, Thai Nguyen Publishing House.
- Scarino, A. (2013). Language assessment literacy as self-awareness: understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327 <https://doi.org/10.1177/0265532213480128>.
- Spaan, M. (2000). Enhancing fairness through a social contract. In A. J. Kunnan (ed.), *Fairness and Validation in Language Assessment: selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida: Studies in Language Testing 9*. Cambridge, UK: Cambridge University Press, 35–37.
- Sultana, N. Language assessment literacy: an uncharted area for the English language teachers in Bangladesh. *Lang Test Asia* 9, 1 (2019). <https://doi.org/10.1186/s40468-019-0077-8>
- Vu, A. T. P. (2017). Classroom-based assessment in Vietnam: An investigation into teachers' beliefs and practices. *Vietnam Language Assessment Symposium*.