A Critical Review of the Role of Language Assessment Training in Developing Teachers' Language Assessment Literacy

Samia AMEZIANI[1], Lamia AMRANE[2]
[1] University of Algiers 2, Algeria
[2] University of Algiers 2, Algeria

Abstract

This study examines two research papers published in academic journals and available on the web. Although both articles hold a common objective, namely the evaluation of language assessment programs, a different methodology is followed. The aim of this study is to critically analyze and compare these two papers in terms of research background, research design and procedure, analysis and presentation of results, discussion of findings, conclusion and implications. Evaluating Teachers' Assessment training is of significance because it allows the stakeholders to figure out whether the future teachers are well prepared to conduct assessment efficiently in the classroom. The present study will help to identify valid and ethical papers that can serve the research field. . The results of this analysis show that Hatipoglo's research is valid and reliable, but it is poorly designed and structured whereas Lam's research is reliable, valid, and well designed.

Keywords: LAL, language assessment training, teacher training.

Introduction

The ultimate goal of teaching is to arrive at a good quality of learning, and it is only through assessment that teachers can know whether it is achieved or not. Assessment is about evaluating student's development and progression, as it includes many types and categories, such as: summative and formative assessment, criterion-based assessment and norm-based assessment. It can be said that assessment is the heart of effective teaching because it bridges the gap between teaching and learning. For this reason language teachers are required to know the assessment basics, to be skilled in implementing assessment in their classroom, and to ensure the design of good quality tests according to the five principles, namely practicality, validity, reliability, authenticity, and washback. Any   knowledge, skills, and principles related to assessment builds teachers' Language Assessment Literacy (LAL) and enables teachers to carry out assessment properly and effectively in the classroom. LAL is linked to the training that future teachers receive and which develops their knowledge, skills, and principles about assessment. However, poor assessment training can result in low teachers' assessment literacy, and as a consequence, this leads to failure of assessment practices in the classroom and teaching in general.

As mentioned above, Language Assessment Literacy (LAL) is regarded as one of the most important aspects that every language teacher should possess. Despite the need for teachers to be assessment literate, there is a lack in training teachers with the adequate LAL (Lam, 2015). Therefore, we have chosen two articles that deal with the evaluation of language assessment training to promote LAL in two different countries where English is regarded as a foreign language.

The main purpose of this study is to critically analyze and compare two different articles in terms of the layout, research design and background, and in terms of data collection and analysis. The purpose behind this analysis is to see how language assessment training impacts the development of teachers' language assessment literacy both in Turkey and in Hong-Kong and to evaluate the validity and reliability of both articles.

This study will serve to identify valuable, valid and reliable research papers that can be used by future master students. Through this research, it is expected that pre-service teachers and in-service teachers become aware of the importance of developing their assessment literacy.

1- Research Background
1.1- Assessment:

Assessment is very important for the development of quality education (Black & William, 1998; Crooks, 1988). It diagnoses the students learning, and examines their level of knowledge and understanding, . Also, it evaluates the effectiveness of teaching (Ibid).In other words, assessment plays a key role in the teaching and learning process as it measures the extent to which the students have progressed and achieved the desired learning outcomes.

In educational practice, the term assessment is popular and misunderstood .It is usually confused with the term test, these two terms are used interchangeably because they are thought to be synonymous but they are not. They are two different concepts (Brown,2003). Brown (2003) explains that tests are a subset of assessment, and a genre of assessment techniques. In scientific terms, a test is a method of measuring a person's ability, knowledge or performance in a specific field. It is also a process of quantifying a test-taker performance according to explicit procedures or rules (Bachman,1990).

Gardner (2006) explains that assessment is the collection of information about students' learning development. In other words, assessment seeks to determine how well students are learning and provides feedback to students, educators, parents, policy makers, and the public about the effectiveness of learning service.

According to Brown (2003) assessment is an ongoing process that involves a set of methodological techniques. Every time a student makes a comment or answers a question, the teacher responds with an appraisal to the student 's performance .Every single performance in the classroom from producing a phrase to a formal essay is subsequently « judged » by self, teacher, or mates .Therefore ,assessment implies all types of interactions between the students and their teacher in the classroom. Assessment can be incidental or intended. Incidental assessment occurs when the teacher gives feedback whenever he/she notices that it is necessary while teaching. On the other hand, intended

assessment takes place when the teacher sets a planned assessment instrument beforehand. A good teacher knows how to assess his /her students permanently.(Ibid)

Tusuonglo (2018) assumes that assessment enables teachers to grade and classify students according to their performance, and it provides feedback to enhance and adjust teachi1ng accordingly. Similarly, Taras (2005) affirms that assessment allows teachers and instructors identify the level of skills and knowledge learned by students. Also, it provides feedback about students' strength and weaknesses. Furthermore, Taras (2005) posits that assessment measures the development of students over a period of time, then, it classifies the students for selection or exclusion . Although many assume that assessment is the gathering of information about students performance, Buyyukarci (2014) argues that assessment is not a mere collection of grades and feedback.

However, it is a systematic process in which teachers reflect on the appropriate way information can be delivered, in order to enhance and improve teaching and learning. Moreover, Pierce (2002) affirms that assessment does not merely inform about weaknesses and strengths of learners. However, it provides learners with feedback to assist their learning, and it enables teachers to adapt their teaching according to the learning styles of their students. Therefore, assessment is an important aspect in education, it is a technique which is used by teachers to gather information about students' achievement. It helps students improve as it provides constant feedback, and it tries to scaffold the weaknesses and difficulties in learning. Teachers can benefit from assessment by highlighting what really works in their teaching ,and changing what needs to be changed in order to deliver the information and knowledge properly.

Assessment is not a mere collection of grades as stated by some experts. However, it is an ongoing process that both students and teachers need in order to identify the strengths and weaknesses of learning in the classroom. It is recommended to allocate more time for assessment practices in the classroom because the ultimate goal of language teaching and learning is to improve and develop students' language ability, and this goal can be attainable by providing constant feedbacks.

1.2- Language assessment literacy:

In language education, the term assessment literacy has been used to describe the knowledge teachers should have about assessment. It has been considered as one of the most essential competencies of a language teacher.(Cheng, Rogers, & Hu, 2004, López & Bernal, 2009, Tsagari & Vogt, 2017).LAL has an important role to help teachers making assessment decisions, and any erroneous decision can lead to a negative impact on students' learning process (Purpura, 2016; Purpura, Brown, & Schoonen, 2015).Therefore, it is indispensable that teachers possess language assessment literacy owing to its major role in teaching practices and decision making. If teachers fail to measure what they taught effectively ,they would not pave the way for students to improve.

Fulcher (2012) states that the earliest attempt to define assessment literacy for teachers was made by the American Federation of Teachers in their Standards for Teacher Competence in Educational Assessment of Students(1990) .They proposed the term "Assessment Literacy" as guidelines to help teachers become aware of assessment. It was divided into two standards.

First, the instructional one where teachers should be able to know how to design and evaluate tests in relation to learning and teaching. Second, teachers should be able to communicate test results properly with other stakeholders. Later Stiggins (1991) identified an assessment literate as a person who can distinguish between good and poor-quality assessment and comprehend the basic concepts and procedures of education testing and then apply this knowledge to obtain information about student 's learning achievement.

Language assessment literacy is defined as the knowledge and skills that stakeholders such as language teachers, are supposed to have about assessment. Inbar-Lourie (2008) and Malone (2011),for instance consider that the term assessment literacy refers to the knowledge that instructors may have and how they choose to use it in their assessment practices. Taylor (2013) as well, agrees by adding that assessment literacy refers to teachers´ acquaintance with the basic ideas of testing, and how this knowledge is adapted in the classroom when assessing language.

It is also described as "the understanding and appropriate use of assessment practices along with the knowledge of the theoretical and philosophical underpinnings in the measurement of students' learning".(ibid)

Although many scholars agree that Language assessment Literacy (LAL) refers to the knowledge, skills, and principles that stakeholders involved in assessment activities are required to master, the components of LAL is a subject of debate (Fulcher 2012). Particularly, specifying the components of language assessment literacy for classroom assessment has proven to be a serious challenge (Inbar-Lourie 2008; Rea-Dickins 2008). For Popham (2009) , language assessment literacy includes teacher's awareness of the good test principles', along with reliability knowledge and its issues, besides the validity of the test and fairness of the results as well as, his/her ability to prepare student's tests and to design constructed-response and selected-response tasks and the use of alternatives in assessment properly such as portfolios. Similarly, Fulcher (2012)argues that language assessment literacy is the teacher's familiarity with test processes, and awareness of principles and concepts that guide him/her to design, develop, maintain or evaluate, large-scale standardized and/or classroom based tests ,with the inclusion of ethics and codes of practice . Brookhart (2011), on the other hand, argues that knowledge and skills are not exhaustive enough for language teachers. LAL includes knowledge of how students learn in a specific subject; link between assessment, curriculum, and instruction; design of scoring; administration of tests; and the use of feedback to promote learning. (Giraldo,2017,p.182).

According to the definitions above, it can be concluded that all the definitions of LAL share basic elements such as skills, abilities, principles, knowledge, and understanding that are related to assessment and needed by teachers for their own long-term professional journey, to benefit their students and programs or institutions in which they work (Popham, 2009).That is, LAL is the ability of teachers and stakeholders to design, apply ,and evaluate the exams based on theoretical knowledge ,skills, and principles (Sisan, 2019) that help them develop. As a result, it IS indispensable for teachers to be assessment literate.

Davies (2008) describes LAL in three fundamental components which are: knowledge, skills, and principles.

- **Knowledge**:

It represents what teachers should know about assessment, namely the theoretical and practical knowledge about general basics in assessment and applied linguistics, key comprehensions in all aspects related to the students' assessment and language learning .For instance, teachers should know the different types of testing and make the difference between testing and measuring .Teachers need also to learn about the assessment techniques ,and alternatives that are best to use. They should as well be acquainted with the models of language learning, teaching and testing in addition to the types of language learners, etc..

Stiggins (1991) defines assessment literacy as "having a basic understanding of the meaning of high- and low-quality assessment and being able to apply that knowledge to various measures of student achievement" (Stiggins, 1991,p. 535).

- **Skills:**

These consist of the competences of the teachers and the way they use the knowledge they learned. For instance, teachers should be able to design reliable ,valid and fair tests that are related to the curriculum objectives, they should be able to correct students' tests and communicate results properly ,record and report students' development use alternative means and provide feedback. Schafer (1993) similarly makes reference to a series of standards for teacher competence in the educational assessment of students. Thus, for him, teachers should develop appropriate assessment methods, interpret the results appropriately, and communicate them with the students, their parents and with other teachers, in order to make decisions for planning and improving teaching.

- **Principles:**

These concern any awareness and actions towards critical issues related to language assessment, where the teacher uses test processes and scores ethically . Fulcher (2012) defines language assessment literacy as "awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice" (Fulcher, 2012, p 125). That is, teachers need to be aware of the five assessment principles including; validity, authenticity, reliability, practicality, and wash back. They are supposed to assess the learners fairly and measures all users of language equally, inform them of what, how, and why they are going to be assessed on, and eliminate assessment bias that offends the test-takers because of their personal traits i.e. gender, race, cultural and socio-economical status.

On the other hand, Scarino (2003) suggests that in ,teacher's way of thinking should coexist with these components. Teachers thinking, ideologies and values can have an impact on assessment practices as well.

In a nutshell, knowledge, skills ,and principles are complementary ,i.e. ,each component completes the other: skills represent the ''how'' ,knowledge represents the ''what'' (Davies, 2008)

To sum up, a good language teacher should be an assessment literate person. He/she needs to be aware of the fundamental components of language assessment literacy

(knowledge ,skills ,and principles) because teacher language assessment literacy has a significant impact on education since it determines teachers' decisions making in the classroom. Thus, owing to the importance of LAL , significant attention should be given to language assessment courses and teacher training programs in order to train future teachers appropriately and to boost their language assessment literacy.

1.3- Language Assessment Training

The goal of teaching is achieving good learning which is guaranteed by the quality of teacher's assessment of students' achievement .Hence, the task of classroom assessment places very heavy demands on the teacher, and requires a complex set of measurement and assessment skills .These skills include the administration and interpretation of standardized tests and classroom tests,   the ability to make rapid in-classroom assessment of student understanding and progress,   the measurement of students' achievement, assignment of grades, and the ability to explain the results to parents.

Given these great demands, it is expected that a major portion of the teacher's professional preparation would be devoted to training them to meet these assessment requirements   under assessment training programs.

Assessment training programs are designed to improve teachers' assessment skills and strategies and make them more competent in their job. If they do not have a firm understanding of the basic principles of assessment, teachers are more likely to engage in unsatisfactory assessment practices. Hence, a necessary (though by no means sufficient) requirement for effective classroom assessment is that teachers be skilled in assessment.

Teachers spend as much as third or half of their professional time in assessment (Crooks,1988). Thus, it is crucial for teachers to be well prepared to handle assessment in their classrooms, therefore, an adequate assessment training should be provided for future teachers in teaching programs.

A language assessment training (course) is defined as a component that teaches future teachers how to use assessment in language classroom productively. (Richard J,& Stiggins,1993)

Language assessment course contributes to shape teachers' language assessment literacy. Therefore, it is expected to have a major part in teacher's professional training. Yet, it seems to be neglected in teaching courses(ibid) .According to Schafer & Lissitz (1987) over half teacher's institutions in the United States deliver no training in assessment. Nonetheless, there are language teaching programs that provide future teachers a course in assessment. Yet, the course may not be consistent and balanced .For instance, lecturers in Turkey who design language assessment courses are often faced with difficulties and dilemmas about what to cover in the course (theory, practice ,how to evaluate the students' performance.).

The language assessment course in all English departments in Turkey (Hatipoglu,2010) is compulsory ,every trainee must pass this course in order to be able to graduate. This course is placed in the seventh semester .It is either taken during Summer or Fall .For the Fall course, it consists of a three-hour class a week throughout fourteen weeks. The summer

course consists of a seven hour class per week throughout six weeks. The main texts used in the course are Genese &Upshur (1696). Besides, many articles, discussions and presentations are used as a support. The language assessment course in Turkey tackles : the notion of teaching and testing ,types of tests, validity, reliability, and techniques to assess language skills. In Algeria, in the University of Algiers 2, students take a language assessment course at Master's level .This course is compulsory with a high coefficient, which makes it a very important module. The trainees take this course three hours a week, throughout five months. This language assessment course consists of many chapters as follows: Notions about assessment and testing, types of assessment, Principles of assessment, and Alternative assessment. The main texts used in this course in the book of Brown and Abeywickrama (2010) entitled "Language Assessment Principles & Classroom Principles". Besides, discussions and presentations take place in the classroom to support trainees' assimilation. Students are also required to design test and evaluate them.

This course is worthwhile, it helps tremendously in shaping assessment literacy. As master 2 students ,we benefited from this module and we acquired a new vision towards assessment. However ,this module should have been taken in the third year of Bachelor's Degree because those graduates who did not peruse Master's class will miss a great amount of knowledge about assessment, especially for those who will teach in middle schools will have no notion about assessment principles and practices.

2. Overview of the Articles:

This section provides a general description of the two articles and provides information about what the two studies stand for.

Article X: Hatipoglu. C (2010) 'Summative Evaluation of an English Testing and Evaluation Course for Future English Language Teachers in Turkey. 13, pp. 40-51.

Article Y: Lam, R. (2015) 'Language assessment training in Hong Kong: Implications for language assessment literacy': Language Testing. 32,(2) pp. 169-197.

Article X is about a summative evaluation of an English Language testing and evaluation course for future English language teachers in Turkey (METU University). This study aims to evaluate an undergraduate ELTE course by English language trainees. Halipoglu wants to uncover how students view the course and whether they believe that it prepares them for professional life. This study was carried out with 81 students .The majority of these students were in their third year and fourth year, a small group consisted of second year students.80 students out of 81 stated that they would work as English Language Teachers after graduation. The data of this study were collected using questionnaires and interviews. The participants of the study were given a questionnaire after their final exam, the questionnaire consisted of two parts. The interview was conducted with 16 volunteer students. It took place after the analysis of questionnaires in order to have in-depth information. The qualitative data were analyzed through codifying the explanations into categories and later these categories were divided into sub-categories, and then, the data were quantified and analyzed using Predictive Analytics Software (PASW). Previous studies by Brown and Bailey are utilized as a support to show that the current study is worth investigation. The results show that the students find that three out of seven sections of assessment course very

useful .Nonetheless, they think that the remaining sections are irrelevant to their future professional career. Students recommend that a major addition should be added to the course. The addition is to allocate more time and opportunities for practice.

On the other hand, article Y aims at exploring the overall language assessment training landscape in Hong Kong and its impact on the pre-service teachers' development of their language assessment literacy in a context of assessment reforms. It also focuses on the perceptions of pre-service teachers and instructors of the improvement of language assessment literacy through course-based language assessment training. Thus, to reach that aim and to answer the two research questions, the researcher follows the positivist paradigm. To address Q1 he has conducted a survey of multiple data sources through the SCOLAR official website. Among a range of 84 program ,he has selected 5 teacher programs relying on 4 selection criteria and contacted program administrative staff by phone/mail to further validate the research .At the end in order to analyze the data collected, he uses categorization and systematic coding of the data. On the other hand, in order to investigate the second research question, he utilized semi-structured interview transcriptions to elicit pre-service teachers' and instructors' opinion on how course-based language assessment promotes LAL.40 teachers were selected to be interviewed in groups of four in addition to 9 instructors who were interviewed individually. To analyze the data collected, the author used a systematic coding and a categorization for the interviews' transcriptions and course materials. The results of the study indicate that language assessment training in Hong Kong are still inadequate ,and the courses selected do not fulfill the aim of bridging the gap between theory and practice within the assessment reform context.

3. Critical Evaluation of the Articles

3.1. Evaluation of the Research Design and Procedure:

Research design is very important in any research because it ensures the reliability of the findings in research and eliminates bias. In this section research design of the two articles is to be evaluated.

Sampling means defining population on which the research will focus (Cohen, 2007). The sampling size of the first study (article X) is 81(16 males, 65 females). The majority of the students were in their third year (39 students) and fourth year (36students). There was a small group of informants from second year (6 students). Their age range between18 to 25. 80 out 81 of the students plan to work as English language teachers when they graduate. According to Cohen (2007), in order to ensure reliability in quantitative research, a sampling size must be minimum of thirty participants. Therefore, the sampling size in this study is reliable since it is of 81 participants. The type of sampling used in this study is probability sampling method. All participants were chosen at random, and everyone had the chance to participate. The advantage of this method is that it minimizes the chance of sampling biases and inferences drawn from the sample can be generalizable to the population.

There is enough information regarding the participants, this includes the gender, age range, affiliation to any institutions, and level. However, the sampling presents at first glance a major weakness and incoherence. The researcher has stated that the ELTE course is placed in the seventh semester, i.e. in the fourth year .It is incoherent that students of second and third year

are part of this study since they have not taken the course yet. It appears illogical to choose students who have never studied ELTE course to evaluate it .Therefore we contacted the researcher to obtain some clarification about the sampling. The question was as follows:

"The ELTE course is placed in the seventh semester, it means that in the fourth year. How come that you chose students of third year and a small number of informants of second year? They have not taken the course yet. My question is how can they evaluate a course that have not studied yet "

Then, the researcher explained:

"... The ELTE course is placed in the 7th semester of our curriculum but second- and/or third-year students who : feel confident they could complete the requirements of the course ,want to graduate earlier ,plan to go for an ERASMUS exchange semester/year can/are allowed to take the course earlier than planned; hence the small number of 2nd and 3rd-year students in the course I talked about in my paper "

This explanation is not mentioned in the sampling of the article which makes it seem ambiguous and unclear. The researcher should have stated these details to avoid ambiguity. After we received the author's explanation, it became clear that the sampling was not done incoherently.

The participants were informed about the aim of the study and they were ensured that the data provided by them will not affect their grades (see page 45). Even though there was no explicit consent of the students, however, the consent was implicit since the participants were informed before-hand about the aim of the study, and they accepted to participate. Therefore, this study is ethical and abides by the ethical code of research.

Two methods of data collection are used: a questionnaire and interviews. The questionnaire consists of two parts: part A and B. Part A comprises detailed information about students' background, this part is assumed to be important as it influences students' learning. Part B encompasses two questions related to the evaluation of ELTE course. As the questionnaire comprises only two questions, which is not enough. In order to carry out an in-depth and complete analysis the researcher should have added more questions. This has affected the validity of the questionnaire.

The questionnaire was designed by two experts who teach ELTE course in the department. It was piloted with 45 students who took the ELTE course in the fall semester of the 2009-2010 academic years. According to Cohen (2007) the questionnaire is a useful tool of data collection that provides numerical data and it can be administered without the presence of the researcher.

The second method which was used in this research is interview. Interviews were conducted with 16 volunteer students. They were set after the analysis of the questionnaires, in order to obtain more in-depth information to ensure reliable results. According to Cohen (2007) the interview is a flexible tool of data collection that serves to obtain more in depth information.

In this article little information was provided about the interviews and their content. Since the researcher has opted for two methods of data collection we can conclude that the results are reliable. The use of both quantitative (questionnaire) and qualitative (interview) methods is known as triangulation. Triangulation is a powerful way of demonstrating reliability .The more the methods contrast with each other, the greater is the researcher confidence.

Validity and reliability ensure the credibility and transparency of the research, and they decrease the researcher biases in qualitative and quantitative research. Validity ensures that the instrument measures what it purports to measure. On the other hand, reliability measures the extent to which the instrument is without bias , i.e. error free, thus, it insures consistent measurement across time and across the various items in the instruments .(Cohen , 2007). The use of different methods to gather data ensures the credibility of results. The researcher in this study (article X) has utilized triangulation (questionnaires, and interviews), which increases reliability.

For the questionnaire, the students had to answer the questionnaire anonymously and there was no force or power affecting their answers, in other words, they are free to answer the questions without any influence or guidance of the researcher. Cohen (2007) states that internal validity indicates whether the results of the study are legitimate because of the way the groups were selected, data were recorded or analysis was performed. It refers to whether a study can be replicated. Therefore, the questionnaire is internally valid as there is no external force from the researcher affecting the results.

External validity of the instrument indicates whether the findings can be generalized to other groups (ibid). Since the selection of the participants was random and not pre-determined, it can be said that the instrument is externally valid as every participant has the chance to participate .as a result these findings can be generalized in the context of university.

The methods are not described equally. The questionnaire is described in details. However, the content of the interviews is not mentioned. Positivism is associated with experiments and quantitative research. It is a form of philosophy that regards knowledge as objective and free from bias (Cohen, 2007). There is a clear rationale that justifies the choice of positivist approach in article X; the researcher adopts the positivist approach because she wants to rely totally on the results that she obtains objectively. Then, draw conclusions based on the evidence. The author of this article says:

" *The feedback that the students provided allowed me to see the course and my teaching from a different, more objective and more independent perspective that otherwise might have remained unnoticed. Thanks to the collaboration with my students, I now have empirical evidence based on which I can make some additions, deletions and revisions to the ELTE course in relation to the research questions scrutinized in this study* ". (Hatipoglu,2010,p.48)

This research is reliable as it is objective, and bias free. However, the validity is affected because it was conducted from students' perspective only. The researcher should have tackled the issue from different angles including teachers' viewpoint).

On the other hand in article Y, a noticeable methodological approach is presented next to the literature review. As the author presents the two research questions in a direct way and the variables are clear as well. It is very important to present the questions to ensure that the study is built upon their investigation. The author has also shown the way of collecting, analyzing, and discussing the data in an organized and simplified way as he used charts to illustrate.

For the data collection, he has used a survey and explicit selection criteria to identify relevant teacher education programs for the first research question, and a focus group interviews for the second question. For the analysis, he has utilized a systematic coding and a categorization for both questions.

The author has used two methods of collecting the data: surveys sampling and a case study (interviews and document analysis).To answer the first research question "What is the overall language assessment training landscape in Hong Kong and how does it influence the development of pre-service teachers' LAL in a context of assessment reforms?", Lam has surveyed all undergraduate English teacher education programs of primary and secondary schools in Hong Kong through an extensive online search in authentic websites of education/ testing agencies i.e., SCOLAR official website (www.language-education.com/eng/dc.esp) and a manual documentation search including government documents , course materials, student assessment tasks.

Among a range of 84 programs, he has selected 24 ELT-related programs with the use of four criteria namely; the programs should be part of the Joint University Programs Admission System, full-time and face-to-face undergraduate programs for Form Six leavers (grade 12), field of study restricted to Applied linguistics, English Language Teaching, and English Literature, and programs under the new-five year academic structure; that is to say, the reform. This has increased the reliability of the study.

He has skimmed university websites to get more details about their offers, and contacted University administrative staff by phone or email when queries arose. At the end he selected only five education programs; four government-funded TEIs and one self-financed TEI, (where 256 pre-service language teachers were enrolled) because they were pertinent to these criteria.

To answer the first research question in depth, he focused on one teacher education institution that provided pre-service teachers with language assessment training and coded the course materials and teaching evaluation documents. The analysis of these documents is triangulated with the interview so this increases the reliability of the research.

Online surveys are effective for his study because knowing about the landscape of a program of the whole country is really time, money, and effort consuming. Using online surveys significantly reduces the set-up and administration costs. i.e., no money spent on paper, printing or postage, and it allows quickly creating, administering, collecting and analyzing surveys to save a considerable amount of time. For this, it can be said that the research method and the collection data are practical to the research question.

There is also enough information about the program educational institutions selected and the way the researcher selected them, this includes the setting, the number of programs, and the content. He has also described the methods and the instruments in a detailed way.

The methods used to collect the data are valid, because he wanted to explore the range of TEIs in Hong Kong, and set boundaries, as a result, the survey mirrors the situation. It is of significance to set boundaries for the survey to further enhance the ecological validity of the study (Creswell,2008). The methods are also reliable because he used specific selection criteria and fine-tuned them into 4 criteria in order to make it bias free.

To answer the second question "what are instructors' and pre-service teachers' perceptions of how course based language assessment training promotes LAL ?', Lam has used a case study in which he conducted semi-structured interviews in English, besides document analysis of two language assessment courses. The type of sampling used in this study is the voluntary response sampling method, as the participation was voluntary. The disadvantage of this method is that it maximizes the chance of sampling biases and inferences drawn from the sample cannot be generalizable to the population. In other words, the sampling does not represent the whole population because the participation was voluntary and participants were allowed to withdraw at any time. Lam has selected 40 pre-service teachers studying in their final year out of 83 from two different courses, in order to make a focus group interview and divided them into groups of four, each interview lasted 35 minutes. In focus group interviews; the answers that the participants gave were likely to be affected by the presence of the researcher and by the social pressure from the rest of the group.

He has also made individual interviews for nine instructors within the same institution which lasted for around 50minutes for each interviewee. Yet, interviewing each instructor for 50minutes is really time and effort consuming, the researcher could have used questionnaires in addition to the interview to arrive at a good level of reliability.

There was a clear rationale for choosing the semi-structured interviews as methods of collecting the data, because the researcher could elicit their opinions in a detailed way on the design, the content, usefulness, and the quality of the assessment courses related to the development of LAL. However there was no piloting in the study, which presents a weakness in the research.

In addition to interviews, the researcher has used course materials and assessment methods to triangulate the data and not to rely on one method only. The advantage of this study is that the researcher has carried out the evaluation from both perspectives: pre- service teachers and instructors which increases the validity of the study. There is not enough information about the participants. He has informed the reader only about their level.

The researcher has also sought consent from the participants before conducting the interviews. According to Bell, (2010) to secure the participants, volunteers should be informed consent before research gets embarked on, because it is considered the norm for the conduct of research. As a result, the research of article Y is also ethical and abides by the ethical code of research.

The researcher took measures to ensure inter-intra reliability as he used categories for coding the case study interviews' transcriptions and to code course materials about the process, content, and delivery of two language courses. This systematic coding was executed three times to arrive at a good level of intra-rater reliability and another research assistant performed the same coding procedure and compared it to the original to ensure inter-rater reliability.

Lam has separated the analysis of the survey and the case study to boost the objectivity and the validity of the study

In short, the two articles are evaluated in terms of validity and reliability of sampling and methods used. The next section is going to be devoted to the evaluation of data analysis and presentation of results.

3.2. Evaluation of Data Analysis and Presentation of Results:

Data analysis is very important in research as it provides an explanation of the findings, and shows the meaning that lies within data. Also, it helps in arriving at conclusions and proving the hypothesis. In this section an evaluation of data analysis of both articles is to be introduced.

In article X , the evaluated course consists of seven main topics : teaching and testing , kinds of tests , kinds of testing , validity , reliability , writing multiple choice item tests , testing skills/ knowledge .

The analysis of data generated by question one (list five topics that you think are useful for your professional career) was conducted by matching the topics mentioned by the students with the topics included in the course. The students were asked to list five topics , but there are some who listed less , thus , the number of topics stated is 339. (See table 1):

Table 1: topics that students think they will help them with their future career as English language teachers.

Section	N	%
Teaching and testing	25	31
Kinds of tests	37	46
Kinds of testing	21	29
Validity	**64**	**79**
Reliability	**76**	**94**
Writing multiple choice item tests	32	40
Testing skills / knowledge	**84**	**104**
Total	339	

As mentioned in table one, the most recurrent topics are: testing skills /knowledge, Validity, Reliability. . Some of the students even listed sub-topics, that is why the number of answers in this section exceeds the number of students.

The most popular topic for the students is the topic of section seven (testing skills/ knowledge). All the students assume that learning the techniques for assessing the skills (such as :reading , writing) is useful .Also , they consider that learning the process of designing exams of different skills , and the critical evaluation of questions of standardized English language tests will help them in their professional career as English language teachers.

The examination of the reasons behind choosing those topics revealed that the answers revolve around four reasons. First , parallelism between topics taught in class and the practices in FLE classes (one of the participants said that since the language skills in Turkish schools are taught separately , then they need to learn how to test them separately as well).Second, acquaintance with a wide variety of testing techniques(a participant said that she learned techniques that she had never heard of before).Third, appropriate tests for different groups of students. Students report that before taking the course, they did not know that young and adult learners are different from each other, and a number of factors should be taken into account when preparing tests. Finally, scoring procedures: participants have learned that different exams require different types of scoring and that students should not be penalized for irrelevant mistakes.

The second most popular topic among students is Reliability, Seventy-six out of the 81 students (94%) think that the topics which are discussed in this section are the ones that would help them prepare better tests. One of the participants said that the preparation of a test is more than putting choices, the tests should have some features , notably reliability , to reach their goals .Other informants report that they have learned techniques to increase reliability in tests because they now realize that a test is not a test without reliability .

The last section which students find useful is Validity, Among the 81 students who participated in this study 64 think that the topics covered in this section will help them in their future career. Informants assert that validity can help them in preparing fairer and better tests by paying attention to what to include and how to include it.

The other four sections were not so popular among students. Section 2 (kinds of tests) is mentioned by 46% of the students and Section 6(Writing multiple choice item tests) by 40% of the participants. In this study, the least frequently mentioned topics by students are "Teaching and testing"(31%) and "Kinds of testing"(26%) . The next part will explain why these topics are not popular.

In the second question of the questionnaire, the participants were asked to list three things they believe should be changed in the course to improve it. Eight strands of suggestions emerged from students' answers. (see table 2)

Table 2 : Topics that need change

Topics	Number	%
1.Course books	4	2.4
2.Topics covered	15	8.9
3.Student presentations	37	21.9
4.Not enough practice	64	37.9
5.Slides before lectures	4	2.4
6.Assessment system used in ELTE course	11	6.4
7.Crowded classes	14	8.3
8.More testing Courses	20	11.8
Total	169	100

Students (8.9%) claim that the number of topics covered in ELTE course is too high and some of them are very abstract (for instance : kinds of testing , teaching and testing) .They recommend that these topics are removed and replaced by more practical practices such as Testing Skills / knowledge , which will help them to assess their students effectively.

37.9 % of the students feel that in the ELTE course theory and practice are not well balanced. They suggest to add more practical exercises that show the link between the theoretical materials covered in the lectures and the real daily challenges that face teachers in the classroom .Also , they say that more time should be allocated to the evaluation and discussion of exams (exams questions , the process of writing exams). 21.9 % of the students argue that the group presentations should be removed and replaced by individual presentation by the teacher. 11.8 % of informants think that more assessment courses should be added to the curriculum, because one single ELTE course is not enough to cover all the topics related to assessment.

The steps which are involved in the data analysis are well explained. First, the data are collected from the students using a questionnaire and interviews. Second, the data are processed and organized into numbers and statistics in tables. Then, the author describes the raw data into words to make it easy to identify and understand the patterns (the answers of participants) .Finally , the author interprets the findings and explains them.

The data analysis is rigorous enough to justify the claims because of the systematic analysis that occurs during each step of the process. For instance, the qualitative data were analysed in two stages: first, the listed reasons were coded Lmto categories and the information in each category was further divided into sub-categories. Then, the data in each sub-category was quantified and analysed using Predictive Analytics Software (PASW).

Not all the data are taken into account .For example the data of interviews are not mentioned at all although the researcher explained how they were analyzed. Only the questionnaire data are presented and analyzed. The presented results are relevant to the question of the researcher (the students' perceptions about ELTE course) ,ie , each   answer was related to students' opinion about the course. The tables make the data analysis clearer and easier to understand to the readership, comparing them with long paragraphs which are complicated to read especially when they contain statistics.

As for article Y, the author has used triangulation with multiple data sources. First, he used a survey to determine the programs that should develop LAL, and utilized a case study to determine the perception of LAL development from both perspectives: pre-service teachers and instructors. He analyzed the survey data collected, by means of content analysis i,e. he used a categorization and a coding system of the survey which included: program/government documents, course materials, student assessment tasks and websites of education/testing agencies (e.g., EDB and HKEAA).

From the analysis of the survey two themes emerged. First, there is an insufficient support to equip pre-service teachers with the adequate LAL, because each TEI offered one or two assessment courses and only TEI D which offered four courses and not all pre- service teachers receive LAL training. Second, there is a lack of training in managing classroom-based and large scale assessment. Although, all courses offered by the 5TEIs included theoretical rational of implementing AFL. Yet, only three courses which covered the practical part in utilizing APA.

From the results of the case study that was made on two language assessment courses (TEI D courses 4and 5) three themes emerged. First, the perception of LAL in an examination-oriented culture; thus, students have almost no access to knowledge base of the social dimensions of language assessment, as a result program administrators need to include LAL courses and acknowledge the learners with its influence on a larger test-driven culture. Second, the experience of course-based language assessment training, thus, it is important to contextualize the knowledge and skills of assessment teaching to advance pre-service teachers' LAL experience. Third, the restricted application of LAL in authentic school contexts, hence, it is of significance to apply knowledge and skills they learn because of there is a great detachment between theory and practice. As a result, it can be concluded that language assessment training in Hong Kong is still inadequate, and the courses selected do not achieve the aim of bridging the gap between theory and practice within the assessment reform context.

The steps involved in the data analysis are well explained: first, the data were collected from the survey and the case study. Second, the data were coded and categorized. Then, the author interpreted the raw data into tables to make it easy to identify and understand the coding of the courses and the answers of participants. Finally, the author interpreted the findings and explained them in details.

The data analysis is rigorous enough to justify the claims, because the researcher has used a systematic analysis in each step of the process. All the data are taken into account, and the results presented are relevant to the research questions .He has used tables to make the data

analysis clearer and easier to understand for the readership. This work shows that the researcher is very qualified and professional in carrying out research, because the style of writing and the layout of the article are very effective.

The results of both articles were evaluated in a way that enables the readership to understand the meaning that lies within data. The next section is going to tackle the evaluation of the discussion.

Conclusion:

This article has provided a comparative and critical analysis of the two selected articles. The latter have been examined and evaluated in terms of title, abstract, keywords, introduction, research background, research design and procedures, data analysis and findings, discussion, conclusions and implications, and the references. The first article (article X) deals with an evaluation of English language testing course for future teachers in Turkey ,it uses a questionnaire and interviews as data collection procedures. For the layout, article X it is poorly structured. The second article (article Y) discusses the landscape of language assessment training in Hong Kong and focuses on implications of language assessment literacy and the extent to which it facilitates and/or inhibits the development of pre-service teachers' LAL. A survey, documentation, and semi-structured interviews were used as data collection procedures. The article is well-presented and structured which reflects a high academic background of the researcher.

References

1. Alderson, J. Charles. (2005). Diagnosing foreign language proficiency: the interface between learning and assessment. London: Continuum.

2. -Al-Mahrooqi.R .(2018). Revisiting EFL assessment: Critical perspectives (pp. 191–208). New York, NY: Springer.

3. -Anderson, J. (1989). Evaluation of student achievement :Teacher Practices and Educational Measurement .The Alberta Journal of Educational Research.36(1),1-3

4. -Aschbacher, P. A. (1991). Performance assessment: State activity, interest, and concerns. Applied Measurement in Education, 4, 275-288 .

5. -Bachman, L F. (1990). Fundamental Consideration in Language Testing. Oxford: Oxford University press.

6. -Bachman, L, F & Palmer, A.S. (2009). Language Testing in Practice: Designing and Developing useful Language tests. New York. Oxford University Press.

7. -Bachman, LF.(1995). Fundamental Considerations in Language Testing. Oxford: Oxford University Press.

8. -Bailey ,k,M .(1996). Working for Washback: a Review of the Washback Concept in Language Testing .https://doi.org/10.1177/026553229601300303 .

9. -Barnes,S. (1985).A study of classroom pupil evaluation:The missinglink in teacher education.Journal of Teacher education,1985, 36(4), 46-49.

10. -Black, P and Wiliam, D (1998) .Assessment and Classroom Learning, Assessment in Education: Principles, policy & practice. DOI:10.1080/0969595980050102 .

11. -Bond, L.(1996). Norm and Criterion-referenced Testing:Practical Assessment: Research & Evaluation (O).http://www.ericae.net/pare/getvn.asp?v=5&n=2(Accessed 10 November 2014).

12. -Brakefield,G .(2009) .Authentic Language Testing Design: A Practical Approach .

13. Brown J ,Hudson, T .(1998). The Alternatives in Language Assessment. TESOL QUARTERLY .Volume 32.issue 4 .

14. -Brown,H.(2003) . Language Assessment :principles and classroom practices. Pearson EsL (1700).

15. -Buyukkarci,K.(2014). Assessment Beliefs and Practices of Language Teachers in Primary Education . International Journal of Instruction. Vol.7,No.1.

16. -Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), Encyclopedia of language and education. Volume 7: Language testing and assessment (2nd ed., pp.349-364). New York: Springer Science and Business Media LLC.

17. -Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. Language Testing, 21(3), 360-389. http://dx.doi.org/10.1191/0265532204lt288oa.

18. -Cheng, L., Rogers, T., & Hu, H. (2004). esl/efl instructors' classroom assessment practices: Purposes, methods, and procedures. Language Testing,21(3) 360-389. https://doi.org/10.1191/0265532204lt288oa.

19. -Clark, J. (1975). Theoretical and Technical Considerations in Oral Proficiency Testing. In S. Jones& B. Spolsky (Eds.), Language testing proficiency. (pp. 10-24). Arlington, VA: Center for Applied Linguistics.

20. -Cohen,L. (2007).Research Methods in Education .Sixth Edition. Routledge. London & New York .Colombia vol 20 n°1, jan-jun.

21. -Coombe, C., Troudi, S, & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, an.d recommendations.

22. -Coughlan,M .(2009). I nterviewing in Qualitative Research. International Journal of Therapy and rehabilitation .June 2009.vol16. No 6 .

23. -Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58(4), 438-481 .

24. -Davies, A. (2008). Textbook trends in teaching language testing. Language Testing, 25, 327–348.

25. -Davies, A. (2008). Textbook trends in teaching language testing. Language Testing, 25(3), 327-347. https://doi.org/10.1177/0265532208090156.

26. -Fulcher,G. (2012) assessment literacy for the language classroom language assessment

27. -Gardner, J. (2006). Assessment and Learning . London: Sage.

28. -Giraldo, F.(2018) Language Assessment Literacy: Implications for Language Teachers,Colombia. https://doi.org/10.15446/profile.v20n1.62089 .

29. -Harris, D .(1969). Testing English as a Second Language. New York: mc.Graw Hill Book Company.

30. -Hatipoglu, C .(2010). Summative Evaluation of an English Language Testing and Evaluation Course For FutureEnglish Language Teachers in Turkey. Vol 13. Middle East Technical University.Turkey.

31. -Herrera, L, & Macías, D. (2015). A call for language assess-ment literacy in the education and development of Teachers of English as a foreign language. Colombian Applied Linguistics Journal, 17(2), 302-312. https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09.

32. -Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. TESOL Journal, 5(1),8-1 1.

33. -Inbar-Lourie, O. (2008). Constructing a language assess-ment knowledge base: A focus on language assessment courses. Language Testing, 25(3), 385-402. https://doi.org/10.1177/0265532208090158.

34. -Jin, Y.(2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China.

35. -Jones, R. (1979). Performance Testing of Second Language Proficiency. In E. Briere & F.Hinofotis (Eds.), Concepts in Language Testing (pp. 50-57). Washington, DC: TESOL.

36. Lam, R. (2015) 'Language assessment training in Hong Kong: Implications for language assessment literacy': Language Testing. 32,(2) pp. 169-197.

37. -Lewy, A. (1990). Formative and summative evaluation. In Walberg, H. & Haertel, G. (Eds.), The International Encyclopedia of Educational Evaluation, 26-28.

38. -Linn, R.(2000). Assessments and Accountability.ER Online 29(2): 4-14.

39. -Malone, M. (2011). Assessment literacy for language educators. *CALDigest,* October, pp. 1-2.

40. -Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. Language Testing, 30(3), 329–344 .

41. -Michelle L. Stabler-Havener. (2018) defining, conceptualizing, problematizing, and

42. -Millind ,S .Tullu. (2019) . Writing the title and abstract for a research paper .doi: 10.4103/sja.SJA_685_18 .

43.   -Mohajan,H .(2017) . Two Criteria for Good Measurements in Research: Validity and Reliability . Annals of Spiru Haret University, 17(3): 58-82

44. -Murad ,I . (2014). A Critical Review Study Conducted on Two Academic ArticlesPublished in the Educational Field: From a Research Prospective.University of Zakho .Kurdistan Region. Iraq.

45. -Newfields, T. (2006). Teacher development and assessment literacy. Authentic Communication: Proceedings of the 5thAnnual JALT Pan-SIG Conference. May 13-14, 2006. Shizuoka, Japan: Tokai University College of Marine Science.

46. -Noll, V. H. (1955). Requirements in educational measurement for prospective teachers. School and Society, 80, 88-90.

47. -Parveen ,H, Showkat, N. (2017). Research Ethics. Aligarh Muslim university. India.

48. -Pham,L.(2018). Qualitative approach to research, a review of advantages and disadvantages of three paradigms: positivism, interpretivism and critical inquiry. SouthAustralia : University of Adelaide.

49. -Pierce, L. V. (2002). Performance-Based Assessment: Promoting Achievement for Language Learners. Center for Applied Linguists(ERIC/CLL News Bulletin), 26(1), 13.

50.   -Popham, W. (2009). Assessment literacy for teachers: Faddish or fundamental? Theory Into Practice, 48(1), 4-11. https://doi.org/10.1080/00405840802577536.

51.   -Popham, W.J. (1987). The Merits of Measurement-driven Instruction. PhiDelta Kappan 68, 679-82.

52. -Purpura, J. E., Brown, J. D., & Schoonen, R. (2015). Improving the validity of quantitative measures in applied linguistics research. Language Learning, 65(S1), 37–75.

53. -Richard, JC & Schmidt, R. (2002). Longman Dictionary of Language Teaching and Applied Linguistics.London: Pearson Education.

54. -Sanders, W & Horn, S.(1995).EducationalAssessmentReassessed: TheUsefulness of Standardized andAlternativeMeasures of StudentAchievement as Indicators for Assessment of Educational Outcomes.Education Policy Analysis Archives 3(6): 14-23.

55. -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.

56. -Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice, 32*(2), 118-126.

57. -Schaffer,w.Lissitz,R.(1987). Measurement Training for School Personnel Recommendations and Reality. https://doi.org/10.1177/002248718703800312.

58. -Stiggins, R. (1995). Assessment literacy for the 21st century. Phi Delta Kappan, 77(3), 238-245.

59. -Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan, 69,* 363-368.

60. -Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72*, 534-539.

61. -Stiggins, Richard J.(1993). Teacher Training In Assessment: Overcoming The Neglect . Teacher Training in Measurement and Assessment Skills. Assessment Training Institute. University of Nebraska – Lincoln .

62. -Susan M.(2019). Brookhart Enterprises LLC educational measurement.

63. -Taras .M.(2005). Assessment – Summative and Formative – Some Theoritical Reflections. British Journal of educational Studies, University of Sunderland.

64. -Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language testing, 30*(3), 403-412.

65. -Tosuncuoglu,I .(2018). Importance of assessment in ELT. Journal of Education and Training Studies .Vol. 6, No.9;. doi:10.11114/jets.v6i9.3443 .

66. -Xu,Y. &Brown,G. University English teacher assessment literacy: A survey-test report from China & New Zealand . Papers in Language Testing and Assessment Vol. 6, Issue 1, 2017.

67. -Zaim.M.(2019) .Developing Instruments for Evaluating Validity, Practicality, And Effectiveness of The Authentic Assessment for Speaking Skill at Junior High School. Atlantis Press.vol 267.