Colombian Applied Linguistics Journal
ISSN: 0123-4641
Facultad de Ciencias y Educación de la Universidad Distrital, Bogotá Colombia

Giraldo, Frank; Murcia, Daniel
Language Assessment Literacy and the Professional Development of Pre-Service Language Teachers1
Colombian Applied Linguistics Journal, vol. 21, no. 2, 2019, July-December, pp. 243-259
Facultad de Ciencias y Educación de la Universidad Distrital, Bogotá Colombia

DOI: https://doi.org/10.14483/22487085.14514

Available in: https://www.redalyc.org/articulo.oa?id=305764540007

RESEARCH ARTICLES

Language Assessment Literacy and the Professional Development of Pre-Service Language Teachers[1]

Literacidad para la evaluación del lenguaje y desarrollo profesional de los profesores de idiomas antes del servicio

Frank Giraldo[2]
Daniel Murcia[3]

Citation/ Para citar este Artículo: Giraldo, F., & Murcia, D. (2019). Language Assessment Literacy and the Professional Development of Pre-Service Language Teachers. *Colomb. Appl. Linguistic. J.*, *21*(2), pp. 243-259.

Received: 23-Feb.-2018 / **Accepted:** 07-Sep.-2019

DOI: https://doi.org/10.14483/22487085.14514

Abstract

Language Assessment Literacy (LAL) research has focused on defining the knowledge, skills, and principles that the stakeholders involved in language assessment activities are required to master. However, there is scarce research on the relationship between LAL and the professional development of language teachers. Therefore, this exploratory action research study examined the impact of a language assessment course on pre-service teachers in a Colombian language teaching programme. Data were collected through questionnaires, interviews, teacher and researcher journals and class observations. The findings show that the course promoted theoretical, technical and operational dimensions in the language assessment design practices of the participants. In addition, it enhanced their LAL and professional development. Consequently, this study contends that the LAL course changed language assessment perceptions radically and encouraged pre-service teachers to design assessments conscientiously, a feature not explicitly stated in LAL research involving this group of stakeholders elsewhere.

Keywords: language assessment, language assessment literacy, language testing, professional development

Resumen

La investigación de la investigación de la Literacidad en Evaluación del Lenguaje (LEL) se ha centrado en la definición de los conocimientos, habilidades y principios que las partes involucradas en las actividades de evaluación del lenguaje deben dominar. Sin embargo, hay poca investigación sobre la relación entre LEL y el desarrollo profesional de los profesores de idiomas. Por lo tanto, este estudio exploratorio de investigación-acción examinó el impacto de un curso de evaluación del lenguaje en los profesores de pre-servicio en un programa de enseñanza del lenguaje colombiano. Los datos se recopilaron mediante cuestionarios, entrevistas, diarios de profesores e investigadores, y observaciones en clase. Los resultados muestran que el curso promovió dimensiones teóricas, técnicas y operativas en las prácticas de diseño de evaluación lingüística de los participantes. Además, mejoró su LEL y su desarrollo

1 This is a classroom-based research study conducted as an independent research process to examine the impact of a new course on language assessment in the Licenciatura en Bilingüismo con énfasis en inglés at Universidad Tecnológica de Pereira. This study was not funded.

2 Universidad de Caldas, Colombia. ORCID https://orcid.org/0000-0001-5221-8245. frank.giraldo@ucaldas.edu.co

3 Universidad Tecnológica de Pereira, Colombia. ORCID https://orcid.org/0000-0002-3146-6116. dmurcia@utp.edu.co

profesional. En consecuencia, este estudio sostiene que el curso para la LEL cambió radicalmente las percepciones de la evaluación del lenguaje y alentó a los profesores en formación a diseñar las evaluaciones de manera concienzuda, una característica que no está explícitamente establecida en la investigación de LEL que involucra a este grupo de interesados en otros lugares.

Palabras clave: evaluación del lenguaje, literacidad en evaluación de lenguas, pruebas de lenguaje, desarrollo profesional

Introduction

In language education, language assessment has been a focus of scholarly work. This focus is necessary given that assessing students' language ability is a key task for language teachers. Information from assessment is used for a variety of purposes, including monitoring progress in and achievement of learning. Specifically, language assessment in the classroom has gained considerable attention from scholars, who agree that it must be sound (Katz, 2013; Rea-Dickins, 2001). Authors, such as Davison and Leung (2009), Fulcher (2012), López and Bernal (2009), have highlighted the need for quality classroom language assessment, arguing that language teachers need to improve their assessment knowledge, skills, principles and practice in language assessment.

Given this background, the notion of Language Assessment Literacy (LAL), which refers to knowledge, skills and principles for the enterprise of language assessment (Davies, 2008; Fulcher, 2012), has become an all-encompassing theoretical framework to research, with special focus being drawn to in-service language teachers. In fact, our literature review reveals that studies on in-service teachers' LAL predominate this field of research (Arias & Maturana, 2005; Kremmel et al., 2017; López & Bernal, 2009; Sellan, 2017; Xun & Brown, 2017). Research has shown that, in general, in-service language teachers do need to improve their knowledge—particularly their design skills—in language assessment.

Consequently, experts have raised a clear call to action for language teacher education programmes to improve pre-service teachers' LAL so that their practices in the field are professional and effective. Although the call has been emphatic (see Herrera & Macias, 2015; Inbar-Lourie, 2017), research with *pre-service* language teachers and their professional development in language assessment has been scarce (but see Restrepo & Jaramillo, 2017). Therefore, this study characterises the impact of a language assessment course on the professional development of pre-service language teachers at a language education programme in a Colombian state university. The researchers conducted an action research study, whose diagnostic stage helped us identify the core topics for the language assessment course under scrutiny. They understood that a course combining theory and practice of language assessment was highly expected. This paper reports the findings from the action-evaluation stage of the action research cycle.

Literature Review

In general terms, LAL refers to knowledge, skills and principles for language assessment. This kind of literacy involves different stakeholders, key among them being language teachers (Taylor, 2013). According to Giraldo's (2018a) review, LAL for teachers includes knowledge of applied linguistics issues such as communicative approaches to language assessment, second language acquisition, concepts such as validity and reliability, and knowledge of own assessment contexts; skills include instructional skills such as improving teaching based on assessment data, designing quality assessments for language skills, among others; and finally, principles include professional practice through fairness, transparency and ethics in language assessment. Research in LAL has shown that although these three major components have not changed, the nature of each component for different people involved in LAL is still a matter of examination (Inbar-Lourie, 2013; 2017; Taylor, 2013). Although experts are welcoming research in LAL, the existing literature has focused on the areas that we review next.

Research and Conceptual Insights on Language Assessment Literacy

On the one hand, authors have identified how language testing textbooks and courses foster LAL. This line of research has suggested that both sources of LAL have stayed on a rather theoretical side (Brown & Bailey, 2008; Davies, 2008). Particularly, Davies (2008) has made the call that the three fundamental components of LAL (that is knowledge, skills and principles) need to be looked as complementary rather than in isolation. Another issue that has been identified in textbooks and courses that promote LAL is that the social dimension of language assessment (the power language assessment can have on people; e.g. scores used for acceptance to universities) and its uses have been understudied (Jin, 2010). Therefore, scholars have called for the incorporation of not only theory but also practice and a critical stance towards what language assessment involves. In other words, as Davies (2008) implies, practitioners should view LAL as the interplay between knowledge, skills and principles.

On the other hand, research and discussions in LAL have focused on the specific LAL that different stakeholders should have. Taylor (2013), for example, has proposed differential profiles for test writers, university administrators, professional language testers and classroom teachers. In the case of teachers, Taylor suggests that language pedagogy, their contexts of teaching, including beliefs and practices, and technical skills should be predominant in LAL for this group —for a review of specifics in LAL for language teachers, see Giraldo (2018a). As explained, these profiles are gaining momentum in the language testing field, which means research is welcome and encouraged.

In the case of language teachers, LAL research has had two related foci. First, there has been a predominantly diagnostic approach to LAL among teachers, with studies reporting that these stakeholders need to improve their LAL skills across the board (Fulcher 2012; Vogt & Tsagari, 2014), though with a special desire to *design* language assessments (Kremmel et al., 2017; Yan et al., 2017). Second, LAL research with classroom teachers has targeted their beliefs and practices (Giraldo, 2018b; López & Bernal, 2009; McNamara & Hill, 2011; Tsagari & Vogt, 2017). The trends in these studies include the belief that language assessment is an important dimension of language teachers' practices; the frequent use of traditional methods such as tests and quizzes; a mismatch between beliefs and practices; clear sequences for doing assessment in the classroom (planning, presenting, executing and evaluating assessments); and difficulties such as lack of time for doing quality language assessment. Put together, these studies support Scarino's (2013) argument that teachers' contexts for language assessment should contribute to the meaning of LAL.

One last discernible trend in LAL involving teachers has shown that they can indeed improve their LAL when engaged in professional development opportunities. For example, in Walters' (2010) study, ESL teachers used reversed engineering to arrive at test specifications to critique the vagueness in standards for language learning. In Arias, Maturana, and Restrepo's (2012) study, language teachers improved their language assessment practices and, by using thorough rubrics, made them more rigorous, transparent, principled, fair and democratic. These two studies show that well-planned programmes for language teachers foster different dimensions of their LAL, including theory and practice.

Overall, a predominant focus on in-service teachers suggests that their LAL needs to be further developed. Indeed, language teachers are a key group of stakeholders who need LAL for professional development and to impact teaching contexts positively (Giraldo, 2018a; Inbar-Lourie, 2017). The clear need among *in-service* teachers may provide a strong rationale to foster high levels of LAL at earlier stages of professional development, namely *pre-service* teacher training. Therefore, a clear gap in the LAL literature has emerged; that is, research observing LAL among pre-service language teachers is still in its infancy.

In Restrepo and Jaramillo's (2017) preliminary findings, pre-service teachers showed evolving awareness of what language assessment is and what language constructs mean. In a diagnostic

Giraldo, F., & Murcia, D. (2019) • Colomb. Appl. Linguist. J.
Printed ISSN 0123-4641 Online ISSN 2248-7085 • July - December 2019. Vol. 21 • Number 2 pp. 243-259.

study with pre-service teachers and language teacher educators, Giraldo and Murcia (2018) compiled a list of core themes to design a language assessment course for pre-service teachers. Giraldo and Murcia's (2018) findings show that both groups of stakeholders expected a course that combined theory and practice in critical ways. It is against this conceptual and research background that this study hopes to contribute to the language testing field, and especially language teacher education, by targeting the LAL of pre-service teachers and their overall professional development.

Since learning requires well-guided assessment, language teachers should pursue the enhancement of their LAL. However, different authors have insisted that there is scarce attention directed to the role of assessment practices in the development of Colombian language teachers' profiles and how these practices are taught, learned and developed in their cognition (Herrera & Macías, 2015; López & Bernal, 2009; Restrepo & Jaramillo, 2017). The narrowed importance given to language assessment procedures is evident in official Colombian language assessment documents. Although there have been some efforts by the Ministry of National Education to design and publish instructional materials for language assessment (e.g. the Suggested English Curriculum), in-service and pre-service language teachers should still be supported by the interpretation and implementation of such assessment materials. Assessing students' language ability may be problematic if teachers are not familiar with the knowledge, skills and principles that embrace the assessment universe. Therefore, to foster LAL for language teachers, this study observed the impact of this construct on the professional development of pre-service teachers enrolled in a language assessment course. The research process was informed by these questions:

1.) How is Language Assessment Literacy (LAL) developed in pre-service language teachers during a language assessment course?

2.) What factors from the language assessment course have an impact on the pre-service teachers' LAL?

3.) What instructional recommendations could be derived from this study?

Methodology

Context

As part of a curricular reform triggered by the adjustments in teacher education programmes in Colombia (education decree 18583 of 2017), Universidad Tecnológica de Pereira (Colombia) [*Technological University of Pereira*] renamed its language teaching programme to Bilingualism and English Language Teaching, which implied adding courses to the curriculum to contribute to pre-service teachers' professional development.

This study emerged during the aforementioned transition and took place in the subject *Classroom Language Assessment Course* (the CLAC), which students take in semester eight of the ten-semester programme. This course was integrated into the curriculum in order to respond to students' and professors' needs to meet the demands of the language education field. The CLAC is taught every week for four hours and is conducted in English and Spanish. The first-course cohort started in the second term of 2017. Its creation was triggered by the results of a previous diagnostic study that was designed to build the CLAC syllabus based on stakeholders' views. Table 1 synthesises the findings from the diagnostic stage, which were fully reported in Giraldo and Murcia (2018).

Research design

To interpret the factors and draw instructional recommendations from the impact of the CLAC on the professional development of pre-service language teachers, we adopted an anti-positivistic approach (Burrell & Morgan, 1979), which is framed under qualitative research through a collaborative action research methodology (Burns, 1999). Under this collaborative inquiry, we embraced the problem as a dialogical exploration to elucidate major trends in data connecting LAL and professional development.

Table 1. Findings from the Diagnostic Stage.

1. Overall emphasis on the practice of language assessment	2. Close Connection between Theory in Language Education and Practice in Language Assessment	3. Other Issues for Training in Language Assessment
Need to learn how to design assessment instruments to assess language skills.	Need to address theoretical and practical dimensions of language assessment.	Need to explore the relations among language assessment and broader contexts of language education as task-based language teaching, CLIL and bilingual assessment.

Participants

The participants (N=33) in the action-evaluation stage were pre-service language teachers of the programme. Their ages ranged from 17 to 25 years old, and 26 of them signed an informed consent agreeing to participate in the research process. They had a B1 level (Council of Europe, 2001) in English, according to the institutional proficiency test taken by all pre-service teachers in the seventh semester. These pre-service teachers had already accomplished 70% of their curriculum and had been exposed to language assessment in different courses; however, this training was done superficially (i.e. in modules, not entire courses).

The other participants were the course instructor, who acted as a teacher-researcher, and a non-participant researcher. The former was responsible for guiding the CLAC, reflecting upon its development, and collecting and analysing research data. Since this was a collaborative action research, the second researcher complemented and contributed to the study by collecting, analysing and reflecting upon the data. Both researchers have been part of the language teaching programme for more than seven years and have been active participants in language assessment in this context.

Data collection and analysis

We set a matrix for research procedures parallel to the sixteen weeks in the CLAC as well as in the action research cycles. We administered two questionnaires in two different cycles; i.e. one after the first period of the course (week 5) and the other at the end of the second period (week 9) (see Appendix A for the questionnaire). In this instrument, the pre-service teachers manifested their views about language assessment ('the before and after'), their changes in teacher cognition during the course, and recommendations for the class.

To access personal opinions concerning the CLAC and how it impacted their professional development, we conducted two semi-structured interviews; i.e. one in the middle and one at the end of the course (see Appendix B for interview protocol). We also developed two class observations during the first and last cycles. The observations, conducted by the second researcher, recorded pre-service teachers' discussions about the act of designing language assessments, the CLAC environment and instructional decisions which occurred in the course. In addition, both researchers wrote journal entries which evaluated the stages of the action research cycles. The teacher-researcher wrote sixteen entries (one for every week in the course), while the non-participant observer wrote four during the course (see Appendix C for guiding prompts in the journals).

The overarching approach to data analysis was Grounded Theory (Birks & Mills, 2011; Glaser & Strauss, 1967) as we sought to identify the interrelationships among the perceived variables, which were labelled using three different coding notes (i.e. open, axial and selective) so as to categorise the phenomena that occurred in the study. *Open* codings involved a first category of trends that repeated itself across data in all instruments; *axial* codings were grouped in related open codings; and

Giraldo, F., & Murcia, D. (2019) • Colomb. Appl. Linguist. J.
Printed ISSN 0123-4641 Online ISSN 2248-7085 • July - December 2019. Vol. 21 • Number 2 pp. 243-259.

finally, *selective* codings were derived from axial codings and represented iterative and prominent trends in the data.

We implemented each instrument with planned intervals so that we could accurately compile data from the different cycles of the action research process and throughout the CLAC (see Figure 1). Meanwhile, we filled our researchers' journals across sixteen sessions and synthesised the course instructor entries in the journals periodically in notes that were later used as open codings. The rest of the instruments were respectively transcribed, written up and safely stored with backup copies. We coded data from the instruments independently and then developed a dialogical exploration by comparing the results. An inter-rater agreement resulted, on average, in 80% of the cases. We discussed divergent codes until a consensus was reached. We also triangulated the four instruments and coded: Open (stage 1), axial (stage 2), and selective coding (stage 3).

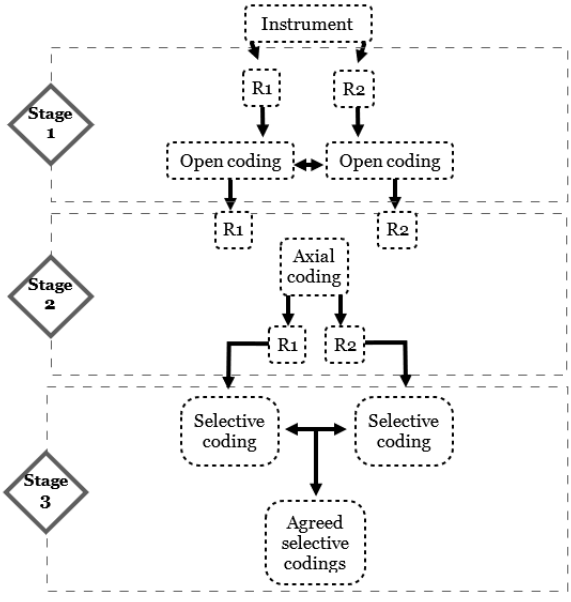

Figure 1. Stages of Analysis of Instruments: Open, Axial and Selective Coding with Inter-rater Agreement (R1= Researcher 1; R2= Researcher 2)

For triangulation purposes, we analysed the four instruments in the three independent stages. For the open coding stage (1 in Figure 1), each researcher explored the instruments while considering trends in data and/or particularities that addressed the questions of the study. This analysis provided a list of open codings that we compared before moving on to the following stage. We derived axial codings (2 in Figure 1) from the consensus of the open codings list and then came up with the same list of axial coding from data. Once again, each researcher analysed the list of axial codings to reach independent selective codings. Further dialogical exploration took place, which led to the agreements for the creation of selective codings (3 in Figure 1). The process was repeated throughout all the instruments and each researcher analysed all the selective codings independently to make a final compendium of categories (see Figure 2), which are later presented as the findings of this study.

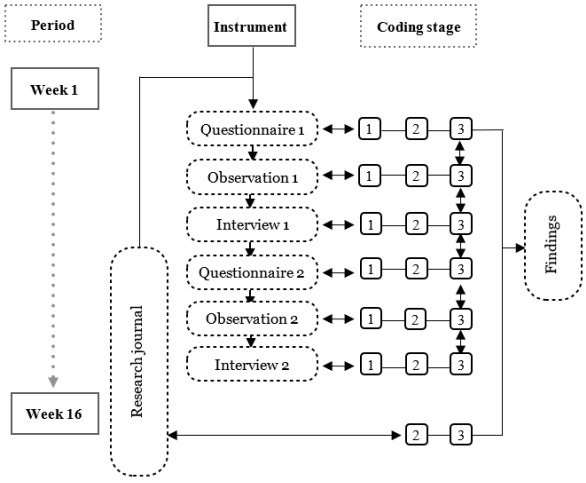

Figure 2. Period, Instruments and Coding Stages.

Pedagogical intervention

Since the CLAC was part of a curriculum for educating pre-service language teachers, the course was based on specific strategies to teach LAL to these students. During the first month in the CLAC, students were presented with an overview of the fundamental theoretical issues in language assessment (e.g. validity and reliability), and they studied this theory through sample assessments that they analysed in class. Some of the assessments they critiqued had design problems so as to allow them use theory to provide sound criticism. A major task during this phase was a report based on

interviews with state school English teachers, who provided information about why, how and what they assessed in their contexts. The interview helped the students to see language assessment in practice and compare real scenarios with what they learned in class.

During the second phase of the course, which lasted two months, students learned how to design items and tasks for language assessment. For example, they followed strict specifications for writing multiple-choice assessments for reading and listening and rubrics for speaking and writing. Throughout this phase, students exchanged the assessments they designed and received feedback on how to improve them.

During the last month of the course, students read about and discussed language assessment in the Colombian context. To facilitate this, they explored the general guidelines for assessment, provided by the national Ministry of Education, and the standards for language learning in the country. In addition, classroom discussions were encouraged and implemented throughout the course. During these discussions, students had the chance to ask follow-up questions from readings and critique language assessment against practical, theoretical and even social contexts (e.g. the influence of standardised testing in language teaching).

Findings

The findings we present are divided into two core sections. In the first part, we present data and analyses to describe the impact that the CLAC had on the pre-service teachers' LAL. In this part, we focus on two specific impacts: A change in the conception of language assessment and the development of a critical theoretical framework. In the second part, the findings include the factors that, according to our analysis, generated such impacts. The most prominent factors were the design of language assessments, multimodal materials for test design, forces shaping design and the overall classroom atmosphere during the CLAC. We also include data from four instruments to substantiate our findings and provide corresponding interpretations.

Impact of the Classroom Language Assessment Course on Pre-Service Teachers

As the data below indicate, it became apparent that the pre-service teachers underwent a radical change towards their conception of what language assessment implied. The data frequently shows that, before the course, the participants thought that language assessment was about grades and/or tests. However, during the course, the participants developed an intricate view of language assessment. They repeatedly stated that language assessment is more than just a grade or test—it involves a continuous process where factors (e.g. clear constructs) are involved. Therefore, in terms of their professional development, the pre-service teachers developed a broader perspective of language assessment.

The sample data below illustrate how their perceptions changed. The first sample comes from Questionnaire # 1, which was administered during the fifth week of the course after —students had studied and contextualised fundamental theoretical and conceptual issues in language assessment.

> Questionnaire#1-Student4
> 4. Before the course, I thought language assessment…
> *Was summative assessment. To design a test and provide a grade.*
> 5. Now, I think language assessment …
> *Is a long process (system) that embraces a number of considerations to have in mind as mentioned above* [principles for language assessment, e.g. construct validity] *to successfully measure students' proficiency level and foster improvement on their language ability and also on course objectives and assessment methodologies into the classroom.*

The change in conception was not only related to the theory of language assessment from an abstract perspective but also in practice. The second data sample is taken from Questionnaire # 2, which was administered during a module in which students were designing language assessments. Still, the simplistic and rather uncritical view of

Giraldo, F., & Murcia, D. (2019) • Colomb. Appl. Linguist. J.
Printed ISSN 0123-4641 Online ISSN 2248-7085 • July - December 2019. Vol. 21 • Number 2 pp. 243-259.

language assessment changed into something more complex.

> Questionnaire#2-Student11
> 4. Before the course, I thought designing language assessments…
> As I was not familiar with the process of designing assessments, I thought it was just a matter of finding a test and adapting it according to the skill that was going to be assessed.
> 5. Now, I think designing language assessments…
> Is a carefully act of creating and writing an important part of assessing students understanding of course content, their level of competency in applying what they are learning and check what students can do with a language and their language ability.

Such changes in perception may be attributed to a series of converging factors in the CLAC. The participants had the chance to contrast their *then and now* experiences in language assessment, so bringing their evolving teacher cognition to the forefront, through prompts as simple as those in the samples above, may have triggered deep reflections. Additionally, as it shall be explained later, the CLAC used problem-based learning as a core methodology to enable students could see language assessment in action through problems posed by the course instructor. In other words, they did not just see the theory of language assessment from an abstract perspective. Further, the CLAC included discussions that usually led to reflections that viewed language assessment as an intricate practice in language education. In Arias et al. (2012), the in-service teachers were engaged in critical tasks that helped them re-conceptualise language assessment and see it more critically. Just like our study, Arias et al.'s study also engaged in-service teachers in careful examination of what language assessment implies and how to design thorough assessments. In Restrepo and Jaramillo (2017), the pre-service teachers, through learning journals, showed intricate views of language assessment. The collective converging findings in these studies show that direct training in and reflection on language assessment leads to heightened awareness of what language assessment is for *both* pre- and in-service teachers' professional development.

Critical Use of a Theoretical Framework in Language Assessment

Through the constant practice of assessment design, the CLAC created an opportunity to assimilate and recycle conceptual issues in language assessment, which became part of the pre-service teachers' theoretical framework. Course sessions provided segments to elucidate language assessment at its practical level, and this triggered discussions about theoretical concepts such as principles (e.g. reliability). In these discussions, the data below, from journal entries and interviews, revealed that the pre-service teachers assimilated and controlled core conceptual issues in language assessment. In addition, the participants referred critically to these issues with frequent use of the metalanguage related to the field. This phenomenon was repeated throughout several sessions of the CLAC, which impacted pre-service teachers' theoretical standpoints about language assessment. Consequently, as part of their professional development, they enhanced their capacity to interpret language assessments in depth.

The following sample from an interview shows how the pre-service teachers used theory critically: *Student 2* uses terminology related to LAL to critique her previous simplistic conceptions of language assessment procedures.

> *Interview#1-Student 2*
> [While designing] I realized that I have to follow the construct and the purpose of the assessment, because it's not only to give an assessment and that's it. It's, it's also to take into account why do you do the assessment, what answers you want to collect, or what information you want to gather.

Since LAL theory was studied, tested and critiqued during classroom sequences, pre-service teachers felt that they were well acquainted with LAL concepts and used them as part of their academic discourse. For instance, the previous sample represents critical reflection manifested through the use of the term 'construct' in which the pre-service teacher expresses how the CLAC shaped her new conceptual positions regarding language assessment.

From the following data extract, theoretical knowledge helps to interpret assessment

instruments through refined LAL-related terms in *Observation 1*. We include ethnographic data, which further presents the use of metalanguage when a pre-service teacher developed a theoretical analysis derived from a test sample.

> Observation#1
> It was used a sample of a test to develop a contrastive analysis which proved what learners have picked up from [...] the course. Once again, students used a repertoire of terminology linked to the field of LAL [...] In the analysis, a participant says: 'Since the quiz is not reliable, then the grade is not valid.

This extract shows how the CLAC triggered analysis when using real samples of language assessment in which the pre-service teachers referred to phenomena with terminology studied in the course (e.g. reliability and score validity in the extract above). When evaluating assessment products with concepts such as 'reliable' and 'valid', the pre-service teachers valued specialised terms. This suggests that the course had shaped their theoretical knowledge as well as promoted their analytical skills. This is also expressed in the study of Arias, Maturana and Restrepo (2012), whose studies show conceptual coherence in the academic discourse of participants when constructing and appropriating terminology for language assessment. Consequently, as also evidenced in this study, pre-service teachers felt empowered and used the metalanguage of the area to project their interpretations of language assessments, assuming the conceptual weight that each of the terms carries. As in Walters' (2010) study, the CLAC was implemented as a formal training that aided the articulation of conceptual aspects to the design of assessments that were more theoretically solid with metalinguistic sophistication. These design practices led to the enhancement of the LAL theoretical framework of the students.

Factors that Contributed to Pre-Service Teachers' Professional Development through Language Assessment

Designing Language Assessments

The most prominent factor that helped the participants in our study to develop professionally was the act of designing language assessments. Data across instruments revealed that while engaged in designing assessments, the pre-service teachers were conscientious regarding the decisions they made for their designs. Furthermore, during design tasks, the pre-service teachers were persistently analytical towards what they constructed, as they even kept contextual factors (e.g. potential students) in mind during the development of assessments. It then seems that the design of language assessments was not a rudimentary activity, but rather an exercise which involved theoretical, technical and operational dimensions. This combination of factors, we believe, had a direct impact on the pre-service teachers' practice of language assessment.

The data extracts below show pre-service teachers and researchers perspectives on designing language assessments, and they illustrate the rather intricate process pre-service teachers went through as they designed instruments for collecting information about language ability.

> Interview#2-Student4
> In the design during the course, the instruments I designed were corrected with all the theory we saw. That means, what was wrong with what I designed? What else do I need to learn? What else do I need to include? What should I avoid? How can I make it more authentic, more valid [...] for students but at the same time more meaningful. So, I think this was what helped me the most in my professional development, because when designing future assessment instruments for my students, I will have in mind everything I learned, which will spare me common mistakes that I made when I had not learned about these assessment theories.

As explained by Student 4, design process not only triggered the use of theory but also a growing awareness for contextualising language assessment instruments. The following samples also show the presence of theoretical, technical and contextual factors that shaped the construction of instruments (e.g. rubrics), as explained by Student 1 and described in Entry 6 of the teacher-researcher journal.

Giraldo, F., & Murcia, D. (2019) • Colomb. Appl. Linguist. J.
Printed ISSN 0123-4641 Online ISSN 2248-7085 • July - December 2019. Vol. 21 • Number 2 pp. 243-259.

Questionnaire # 2–Student 1
3. When designing language assessments, I should…
Have clear objectives and a clear construct to assess, establish a rubric, take into account the level, context, age, interest, knowledge of the students and design a reliable language assessment.

Journal Entry # 6
Students could show me how much thought should be put into designing a reading assessment. Among the things to consider, they highlighted:
- the construct
- students' proficiency level
- support for students in the test (like examples)
- it is important to follow guidelines for item-writing.

In a related fashion, we identified a particular impact of materials on designing language assessments. The data showed that the exercise of constructing test items and tasks relies heavily on multimodality, which not only requires paper-based resources but also technological resources. When available, these materials empowered design; when not, design efforts seemed to be fruitless. We present the data below as evidence to suggest that the conscientious design of language assessments is driven by instructional guidelines, theory, context and a variety of materials._

Journal entries #7 and #11
7: *Designing assessments, at least initially, needs a lot of explicit instruction on what to do and what not to do. For example, making lots of emphasis on the construct and avoiding writing vague descriptors. Design requires that students be ready for it so they don't waste time and, rather, use materials efficiently.*
11: *Finding the right content for a CLIL assessment is key and a constraint. As they were looking for material, they kept the construct in mind, which was also an outstanding thing to see already in their LA system.*

Observation#2
Students should be aware of the need of technological efficacy to have all the materials ready for a design session. (i.e. all theoretical foundations ready to be reviewed, all input for adaptation ready, audios, images and videos, etc.).

Highly structured design tasks were the reason for which pre-service teachers in our study showed heightened awareness during the construction of language assessments. Usually, the course provides guidelines for writing items and tasks, namely using Colombian standards for learning English, technical considerations for item writing (e.g., length of options in a multiple-choice item), and the pre-service teachers are guided to consider purposes and constructs for design. As far as our literature review is concerned, we did not find various studies showing what actually happens as language teachers design assessments. However, in Walters (2010), the participants constructed test specifications as they analysed and critiqued standards of ESL learning, arguably a conscientious activity on its own. In Arias et al.'s (2012) study, the language teachers became increasingly critical towards their own instruments. However, this study does not report any information on the impact of using multimodal materials during the design stage of the instruments they used for assessing language.

As for the influence of materials on the pre-service teachers' design of language assessments, we consider that the use of resources was not a simple matter to include in design but a determining factor to succeed in writing language assessment items and tasks. Regarding this finding, we did not find any study investigating the impact of materials on the design of language assessments. Therefore, our study is probably pioneering in language assessment research conducted with pre-service language teachers in our context.

To conclude, data show an overall positive impact on the pre-service teachers' design of language assessments as they perceived it as a complex and demanding task. The act of design seemed to have impacted their professional development at the theoretical, technical, practical, contextual and critical dimensions of language assessment.

Contextual Forces that Shape Language Assessment

Another factor that contributed to the pre-service teachers' professional development involved contextual forces shaping their design and theoretical framework for language assessment. By contextual factors we mean those including, for example, participants' own teaching contexts and their students. Through data analysis, we noticed that these forces not only influenced the particularities in the design of instruments for language assessment but also allowed the pre-service teachers to connect theory from the CLAC to teaching-learning contexts. It appeared that in our LAL process, the pre-service teachers' professional development was not limited to classroom contexts—during the course, there was a burning need to connect the CLAC to students' lifeworlds.

On the one hand, the questionnaire data below show how contextual forces led the pre-service teachers to design language assessments *vis-à-vis* factors beyond the CLAC. On the other hand, the interview extracts show how contextual factors (type of school and type of instrument) influenced the participants' theoretical framework.

> Questionnaire#2-Student1
> Has the second part of the course had any impact on your professional development as a language teacher? Y/N, why?
> *Yes, the second part has taught me and made me aware of several aspects when designing a test; for instance, stem considerations to avoid misunderstandings as well as clear instructions. Besides, the type of input taking into account students cognitive level and interests.*

> Interview#1-Student9
> *Now I'm teaching English in a private school, so I in some classes, no, in some, in the exam and in a quiz, I implement a criteria for the speaking part, eh, so this guided so much in order to what know skills I'm going to assess, I'm going to assess, to follow the construct, the purpose, eh yes.*

The CLAC engaged the pre-service teachers in analysing standards for learning English in Colombia, reading about general assessment policies (i.e., Decree 1290), interviewing in-service teachers and designing language assessments for their practicum courses. Therefore, we believe the influence of these external forces was a product of being engaged in the CLAC. Otherwise, they would not have come together. In addition, we interpret this finding in light of Scarino's (2013) views towards LAL for language teachers. The author argues that language teachers' lifeworlds—their experiences, contexts and beliefs—shape and are fundamental in developing their LAL. In summary, the pre-service teachers in our study enhanced their LAL because they delved into the practice of language assessment within specific contexts for language education.

The CLAC's Atmosphere for Learning about Language Assessment

The last factor that frequently, and perhaps not surprisingly, emerged from the data and that impacted professional development was the CLAC's atmosphere for fostering LAL among the pre-service teachers. Particularly, we present data to illustrate that the CLAC was based on problem-based learning, used realistic language assessment samples, engaged the participants in teamwork tasks, and provided opportunities for interactive learning through teacher-led discussions and peer feedback exercises. This instructional approach seemed to be conducive to learning about LAL, which in turn contributed to these pre-service teachers' professional development.

The data samples below show evidence of three different but converging instructional decisions in the CLAC. The extract from the class observation provides information about teacher-led discussions for problem-based learning through the analysis of sample assessments. Further, the interview data show how the pre-service teachers worked in teams to design language assessments and its corresponding impact. Lastly, the journal entry describes the use of peer feedback exercises and how they fostered LAL.

> Observation #2
> At the beginning of the session, students held a discussion with the reflection questions proposed by

Giraldo, F., & Murcia, D. (2019) • Colomb. Appl. Linguist. J.
Printed ISSN 0123-4641 Online ISSN 2248-7085 • July - December 2019. Vol. 21 • Number 2 pp. 243-259.

the professor. The prompting question written on the board was: 'What's going on.'?.' This type of question directed the conversation of the class to make connections between the aspects dealt in class (Be it a theoretical explanation or a debate about the multiple conceptions of a term). When the connection between factors was not explicit, the professor started to make some relations to the students 'For example, el Reto. Are you using it to build the design?'

Interview#2-Student12
It was very useful for the three of us that designed the instruments because everyone […] learns something from the other, right? So, I have one way to design something, but my partner has other way that when you mix them you, we have a good product. It was helpful, it was very helpful to work in pairs or in trios…

Journal Entry #7
The assessment evaluation activity proved very successful (this is becoming a trend). When students have the chance to evaluate each other's work on assessment design, I have noticed that they confirm their learning and issues arise, which I as a teacher can address. For example, today, a couple of true-false reading assessments had all statements with true as the key. This led me to remind them that the statements in a true-false assessment need to be balanced or at least have both true and false keys.

Based on the above data from our study, we believe that engineering critical, reflective and practical learning tasks in the CLAC fostered LAL among these pre-service teachers. What is more, the instructional strategies used for the CLAC were not chosen randomly but reflected our sensitive decision-making from the diagnostic stage of our study (See Giraldo & Murcia, 2018), where we concluded that theory, practice and reflection in language assessment had to be included in the course. Therefore, we argue that the needs assessment before designing the CLAC was pivotal in bringing about enriching experiences in language assessment to cultivate professional development. Additionally, our findings are similar to those in Arias et al.'s (2012) and Walters' (2010) studies in that all the three studies engaged participants in discussions, teamwork and critique of language assessment issues.

Unfortunately, studies researching training in language testing for language teachers have shown a rather theoretical view of training (Brown & Bailey, 2008; Jin, 2010). However, the findings of our study, Arias et al., (2012), and Walters (2010) imply that, in fostering LAL among teachers, more is needed than mere theoretical input. In conclusion, our study suggests that to promote the professional development of the pre-service teachers we studied, the CLAC utilised key strategies for developing LAL, which include: a clear and strong connection between our diagnostic and action-evaluation stages; an instructional approach focused on problematising language assessment; a combination between theory and practice; and contextual, critical scenarios to exercise the practice of language assessment.

Conclusions

With the present study, we sought to describe and examine the development of LAL in pre-service language teachers enrolled in a language assessment course. Pedagogically speaking, the study, framed as an action research, helped to cultivate the Language Assessment Literacy of the participating pre-service teachers. Therefore, its purpose was to contribute to LAL discussions by observing an under-researched group of stakeholders in language assessment. In synthesis, the CLAC helped pre-service teachers to develop LAL on two main fronts. First, their perceptions about language assessment evolved from limited views (that of equating language assessment to a test and/or grade) to an intricate and professionalising process-oriented endeavour it indeed is. Second, the CLAC allowed these stakeholders to mature a theoretical framework that they constantly used to critically discuss and do language assessment.

As for the factors that led to these two overarching results in LAL development, it became apparent that the act of designing language assessments empowered the pre-service teachers to use theory in increasingly conscientious ways. Interestingly, we found evidence to suggest that materials used in designing language assessments were a pivotal factor that influences the pre-service teachers' enterprise of design. Also, our data show

that they did not design assessments in the abstract but rather configured a network of forces, external to the CLAC, to ensure their products would be of high quality. The last factor that led to increased awareness and action in language assessment was the way the CLAC was engineered on solid needs analysis data and taught with engaging, critical strategies aimed at furthering LAL.

Implications

The first recommendation we have for the field of teacher education, particularly regarding training in language assessment, is to conduct a thorough multi-stakeholder needs analysis with interested stakeholders. The fact that we gathered course expectations from students and professors helped us propose a language assessment course that made sense. We also suggest that contents for language assessment courses be prioritised. In our case, the diagnostic stage taught us that design had to be a fundamental dimension of the course, which was then reflected on our overall findings —design tasks for training in language assessment are powerful. This leads us to our third recommendation.

Language assessment courses for pre-service teachers should emphasise highly structured design tasks because they trigger conscientious decisions fuelled by seasoned theoretical frameworks. We are confident that we have gathered valid empirical data to argue for a design-based type of course and encourage further studies in other teacher education contexts. Lastly, and in line with the spirit of action research in classroom contexts, we argue that the use of contextual problem-based tasks and the promotion of an interactive atmosphere are conducive to learning in language assessment courses for pre-service teachers.

More research for the development of language teachers should always be welcome so that our practices as educators can evolve. Particularly, and given the rather scarce research to date, we invite researchers to study how pre-service language teachers develop professionally through LAL. Additionally, we believe it may be enriching to learn from how LAL programmes impact in-service teachers. In doing so, we are collectively aggregating findings to help language teachers assess language ability professionally.

References

Arias, C. I., & Maturana, L. M. (2005). Evaluación en lenguas extranjeras: Discursos y prácticas. *IKALA, Revista de Lenguaje y Cultura*, *10*(16), 63-91.

Arias, C., Maturana, L., & Restrepo, M. (2012). Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas. *Lenguaje*, *40*(1), 99-126.

Birks, M., & Mills, J. (2011). *Grounded theory: A practical guide*. London, England: Sage Publications.

Brown, J. D., & Bailey, K. (2008). Language testing courses: What are they in 2007? *Language Testing*, *25*(3), 349-383.

Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life.* London, England: Routledge.

Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge, England: Cambridge University Press.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, and assessment*. Cambridge: Cambridge University Press.

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, *25*(3), 327-347.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393-415.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113-132.

Giraldo, F. (2018a). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, *20*(1), 179-195.

Giraldo, F. & Murcia, D. (2018). Language assessment literacy for pre-service teachers: Course expectations from different stakeholders. *GiST: Education and Learning Research Journal, 16*, 56-77.

Giraldo, F. (2018b). A diagnostic study on teachers' beliefs and practices in foreign language assessment. *ÍKALA: Revista de Lenguaje y Cultura*, *23*(1), 25-44.

Glasser, B. J. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Jersey: Aldine Transaction.

Herrera, L., & Macías, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, *17*(2), 302-312.

Inbar-Lourie, O. (2013). Guest Editorial to the special issue on language assessment literacy. *Language Testing*, *30*(3), 301-307.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, S. May, & I. Or (Eds.), *Language Testing and Assessment (third edition), Encyclopedia of Language and Education* (pp. 257-268). Cham, Switzerland: Springer.

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, *27*(4), 555-584.

Katz, A. (2013). Assessment in second language classrooms. In M. Celce-Murcia, D. Brinton, & M. Snow (Eds.), *Teaching English as a second or foreign language: Fourth edition*, (pp. 320-339). Boston, USA: Heinle Cengage Learning.

Kremmel, B., Eberharter, K., Holzknecht, F., & Konrad, E. (2017). *Enhancing language assessment literacy through teacher involvement in high-stakes test development*. Paper presented at the language testing research colloquium, Bogotá, Colombia.

López, A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teachers' 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.

Ministerio de Educación Nacional, Colombia. (April 16th, 2009). Por el cual se reglamenta la evaluación del aprendizaje y promoción de los estudiantes de los niveles de educación básica y media. [Decreto 1290 de 2009]. Retrieved from https://www.mineducacion.gov.co/1759/w3-article-187765.html?_noredirect=1

Ministerio de Educación Nacional, Colombia. (September 15th, 2017). Por la cual se ajustan las características específicas de calidad de los programas de Licenciatura para la obtención, renovación o modificación del registro calificado. [Resolución 18583 de 2017]. Retrieved from https://www.usbcali.edu.co/sites/default/files/resolucion_final_18583_de_2017deroga_2041.pdf

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, *18*(4), 429-462.

Restrepo, E. & Jaramillo, D. (2017, July). *Preservice teachers' language assessment literacy development*. Paper presented at the language testing research colloquium, Bogotá, Colombia.

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.

Sellan, R. (2017). Developing assessment literacy in Singapore: How teachers broaden English language learning by expanding assessment constructs. *Papers in Language Testing and Assessment*, *6*(1), 64-87.

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

Tsagari, D. & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, *6*(1), 41-63.

Vogt, K. & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374-402.

Walters, F. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, *7*(4), 317-342.

Xu, Y. & Brown, G. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, *6*(1), 133-158.

Yan, X., Fan, J., & Zhang, C. (2017). *Understanding language assessment literacy profiles of different stakeholder groups in China: The importance of contextual and experiential factors*. Paper presented at the language testing research colloquium, Bogotá, Colombia.

Appendix A

Questionnaire for Cycle One and Cycle Two of the CLAC

Cycle one:

Please provide candid answers to the questions below.

1. Has the course had any impact on your professional development as a language teacher? Y/N, why?

Complete the following statements based on what you have experienced in the course:

2. When assessing language, I should…

3. When assessing language, I should not…

4. Before the course, I thought language assessment…

5. Now, I think language assessment…

6. What recommendations do you have for the course?

Cycle two:

This present questionnaire asks you about the second part of the course, in which you have designed listening, reading, speaking and writing assessments. Please provide candid answers to the questions below.

1. Has the second part of the course had any impact on your professional development as a language teacher? Y/N, why?

2. When designing language assessments, I should…

3. When designing language assessments, I should not…

4. Before the course, I thought designing language assessments…

5. Now, I think designing language assessments…

6. What recommendations do you have based on the second part of the course?

Appendix B

Interview Protocol

Ice-breaking question:

What do you think about the second and third periods of the course?

1. In your opinion which factors contributed to your professional development in the course?

Probe:

What about the practical aspect had an impact on your professional development?

2. From the practical view, have you used any language assessment knowledge and skills in your practice as a teacher?

Probes:

Tell me about the design process that you experienced in the course. How did you do it? Did you have any challenges? What was effective? What was it like to co-design a test?

3. What can you say about the classroom tasks presented in the course? Examples: small group discussions, whole group discussions, the interview you conducted, analysing assessment examples, etc.

4. Since the CLAC is going to continue, what recommendations do you have for the course?

Appendix C

Prompts for Writing Journal Entries

Teacher-Researcher's Journal:

What went well during this lesson?

What did not go so well?

Conclusions and lessons learned

Non-participant Observer's Journal:

Action research cycle (implementation-action stage) objectives

General: Analyse information from students, tutor and teacher researcher to determine what kind of impact the course is having on pre-service teachers.

Specific: Derive broad instructional recommendations for the language assessment course.