Language Studies: Article

DOI: https://doi.org/10.5007/2175-8026.2024.e96898

DEVELOPING PRE-SERVICE TEACHERS' LANGUAGE ASSESSMENT LITERACY: AN ANALYSIS OF CLASSROOM TASKS

Isadora Teixeira Moraes[1*]
[1]Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

Viviane Bagio Furtoso[2**]
[2]Universidade Estadual de Londrina, Londrina, PR, Brasil

Frank Giraldo[3***]
[3]Universidad de Caldas, Manizales, Caldas, Colombia

Abstract

Language assessment literacy (LAL) comprises the knowledge, skills and principles for sound assessment. In this paper, we investigated classroom tasks from two assessment courses from a state university in Brazil; specifically, which LAL components the tasks prioritized, and which of them promoted LAL the most. Data generation instruments included the courses' calendars, observation protocols and pre-service teachers' answers to questionnaires. In terms of LAL components, tasks that involved knowledge and skills were top-ranked, and in some cases included principles. Findings also indicate pre-service teachers' preference for practical activities; their favorable view of the teaching procedures, and their recognition of the courses' relevance to their professional development.

Keywords: language assessment literacy; teachers' professional development; classroom tasks.

[*] Isadora Teixeira Moraes is a professor in the Languages Faculty at *Universidade Federal de Minas Gerais*. Her main research interests are language assessment literacy and teacher education. Email: isadoramoraes@ufmg.br. ORCID: https://orcid.org/0000-0002-5393-286X.

[**] Viviane Bagio Furtoso holds a PhD in Language Studies from *Universidade Estadual Paulista "Júlio de Mesquita Filho"*. She did her postdoctoral internship at Georgetown University, USA. She is an Associate Professor in the Modern Languages Department at *Universidade Estadual de Londrina*. She develops research in the area of Portuguese for speakers of other languages with a focus in assessment, teacher education, and materials development. Email: viviane@uel.br. ORCID: https://orcid.org/0000-0001-6535-2902.

[***] Frank Giraldo is a professor at the Foreign Languages Department of Universidad de Caldas in Manizales, Colombia. His main academic and research interest is the interface between language assessment literacy (LAL) and teachers' professional development. Email: frank.giraldo@ucaldas.edu.co. ORCID: https://orcid.org/0000-0001-5221-8245.

Esta obra tem licença *Creative Commons*
http://creativecommons.org/licenses/by/4.0

1 Introduction

Language Assessment Literacy (henceforth LAL), an area within the field of language assessment, has been receiving attention particularly when it comes to teachers. LAL refers to the knowledge, skills, and principles that are necessary for sound language assessment in various educational contexts. Authors have argued that teachers need enough LAL to assess language ability professionally (Fulcher, 2012; Inbar-Lourie, 2008; 2017; Taylor, 2013). Thus, LAL is becoming a focus of learning and professional development for language teachers. The need for language teacher education in LAL is sensible, especially because teachers need to promote and report the development of their students' language proficiency, which is one key factor to determine the quality of language learning programs.

In fact, teachers themselves have reported burning needs in the area of language assessment, particularly when it comes to practical matters that are directly pertinent to their classroom practice (Berry, Sheehan & Munro, 2019; Hasselgreen, Carlsen & Helness, 2004; Vogt & Tsagari, 2014). Other studies have shown that LAL programs for teachers should directly address their needs and promote their professional development in language assessment (for example, Baker & Riches, 2017; Montee *et al.*, 2013).

Although conceptual and empirical research on *in-service* teachers' LAL is ongoing, research regarding *pre-service* teachers' LAL and how to teach them is still under development (Giraldo, 2021). Therefore, in this paper, we report an analysis and evaluation of classroom tasks[1] that sought to promote the LAL of a group of pre-service teachers in two language assessment courses of a state university in Brazil. Specifically, we investigated 1) what LAL components were prioritized in the courses, and 2) what classroom tasks promoted LAL the most. We hope that our findings can contribute to the growing discussion on pre-service teachers' education in LAL.

We start this paper by overviewing the meaning of assessment literacy, the general educational term, and then move on to discussing LAL as it pertains to language teachers. This then provides a background for examining the LAL of pre-service teachers, which is the focus of our study. After detailing the theoretical and research context, we explain the methodology for our study, and then we present and discuss our findings. We end the paper with conclusions and implications on teaching language assessment to pre-service teachers.

2 Theoretical Framework

The theoretical framework is divided in three main subsections: first, we explore how the concept of language assessment literacy has been framed in the area; second, we discuss LAL for language teachers in general; and third, we look at the specificities of LAL for pre-service language teachers.

2.1 Language Assessment Literacy

Before delving into LAL, we begin by discussing assessment literacy (henceforth, AL), the more general concept. AL refers to the ability to conscientiously evaluate the results obtained from assessment (Stiggins, 1991). AL for teachers means possessing a repertoire of theoretical and practical competences and knowing how to use them appropriately to improve teaching and learning (Brookhart, 2011). In language education, AL takes on different characteristics, since there are additional competences related to language concepts and theories about language learning and teaching that inform the assessment process, therefore distinguishing AL from LAL (Inbar-Lourie, 2008; Scaramucci, 2016; Taylor, 2013).

The most accepted characterization considering its constant presence in articles and discussions in the area is that LAL encompasses three components: knowledge, skills and principles (Davies, 2008; Giraldo, 2018). Inbar-Lourie (2008) explains that knowledge refers to the attribute to be assessed, which in this case is language ability, as well as an understanding of language learning theories and strategies. The skills refer to the process of assessment, and are related to familiarity with testing and measurement techniques, the use of a broad range of assessment tools, the development of rubrics and the ability to provide effective and valid feedback to students. Finally, the principles are the rationale behind assessment, that is, the reasons for assessing. They involve matters such as ethics in assessment, recognizing power relations and the impact that assessment has on its stakeholders (i.e., teachers, students, parents, test developers) and society at large.

Importantly, various authors have given attention to the critical dimension in language testing and assessment and have even suggested Critical Language Assessment Literacy (CLAL) as an emerging trend in general LAL (Tajeddin *et al.*, 2022). As these authors have stated, teachers' CLAL refers to the analysis of 1) tests and their social and educational consequences; 2) the design and/or adoption of useful tests and the necessary contextual conditions for them to be so; 3) models of language ability; 4) their own practices, beliefs, and dispositions; and 5) unfair or unethical practices in assessment (Brindley, 2001; Inbar-Lourie, 2008; 2012; Scarino, 2013; Shohamy, 2001; 2022).

Overall, general models of LAL have included the following:

Knowledge: Models of language ability; core concepts (e.g., proficiency; validity; reliability); frameworks for assessment (e.g., norm-referenced); statistical interpretation;

Skills: Assessment construction and evaluation; use of statistics for test analysis and improvement; use of technological and statistical tools;

Principles: Ethics, fairness, transparency, and the impact of assessment on society, which all contribute to CLAL (Tajeddin *et al.*, 2022).

Although these components are defined separately, they are, in fact, intrinsically connected. To illustrate the relationship among LAI's three components, rubric development requires *skills* in doing so, but choosing the criteria would imply *knowledge* of language and teaching concepts. Finally, using the rubric and choosing to share it with students has to do with the *principles* side of LAL.

In the next section, we explore the discussion about the LAL components in relation to a main stakeholder group, namely that of language teachers.

2.2 LAL for Language Teachers

Although LAL's three core components have remained constant in scholarly discussions, specificities within each component are still in refinement. In the case of language teachers, various authors in the last 20 years have offered descriptions of what LAL may involve for these stakeholders (Allan, 2020; Brindley, 2001; Davies, 2008; Inbar-Lourie, 2008; Scarino, 2013; Stabler-Havener, 2018; Taylor, 2013). In Table 1, we synthesize specific components for teachers' LAL. We base our table on Inbar-Lourie (2008), because her seminal paper was one of the first that discussed the possible required LAL knowledge base with a specific focus on assessment courses. We also base this table on Giraldo (2018), because his list provides detailed LAL descriptors for teachers, which are based on extensive literature in the area. The dimensions proposed by the latter author in the last column will also be relevant to the findings of this study, specifically in the analysis of classroom tasks.

Table 1 - LAL for language teachers

Components	Definitions	Dimensions
Knowledge	…refers to: - understanding the attribute to be assessed; - analyzing language models rigorously; - analyzing one's language assessment lifeworlds; - defining proficiency; - understanding core concepts such as validity and reliability; - understanding current theories in language teaching, learning and assessment (e.g., project-based learning, assessment as learning).	I. Awareness of applied linguistics II. Awareness of theory and concepts III. Awareness of own language assessment context
Skills	…are practical aspects that involve: - alignment between objectives, instruction and assessment methods; - development of assessment rubrics; - familiarity with testing/ measurement techniques; - providing feedback to improve students' performance; - employing multiple assessment methods; - implementing technologies for assessment.	IV. Instructional skills V. Design skills for language assessments VI. Skills in educational measurement VII. Technological skills

Principles	...include the rationale behind assessment and require: - considerations about ethics, fairness, democracy, transparency; - recognizing and understanding the historical, political, social and contextual aspects of assessment; - recognizing the power relations surrounding assessment and its impact in stakeholders' lives.	VII. Awareness of and actions towards critical issues in language assessment

Source: the authors, based on Giraldo (2018) and Inbar-Lourie (2008)

Various authors have explained that, rather than seeing the LAL components as separate, they relate to one another (Giraldo, 2018; Kremmel & Harding, 2020; Malone, 2017; Taylor, 2013). For example, Taylor (2013) argues that teachers need high levels of language pedagogy, –which may be considered knowledge and skills–; knowledge of personal beliefs, attitudes, and their assessment context; technical skills for design; and sociocultural values. In sum, particular LAL profiles are needed for particular assessment contexts.

When it comes to the empirical side of LAL, studies have shown that teachers need education mainly on the *knowledge* and *skills* aspects of language assessment (Fulcher, 2012; Kremmel *et al*., 2018; Vogt & Tsagari, 2014; Giraldo & Murcia, 2019). In other words, teachers seem to want a practical approach to language assessment whereby they construct, implement, and evaluate the usefulness of instruments and procedures. Interestingly, the place of *principles* in teachers' LAL remains an issue for ongoing scrutiny (more on this below).

The other dimension of empirical research in LAL is composed of studies which report LAL initiatives (programs, courses, workshops) for language teachers. While the studies are few but ongoing, the trends regarding teacher education through LAL are clear (Giraldo, 2021). LAL programs for teachers are positively impactful on three fronts:

> *The practice of language assessment*: Teachers become skillful at designing assessment instruments for different language skills (Arias, Maturana & Restrepo, 2012; Levi & Inbar-Lourie, 2019; Nier, Donovan & Malone, 2009).
>
> *Awareness of the impact of language assessment*: As teachers study what language assessment implies, they become aware of the positive effect it can have on their students' learning and even their own teaching (Baker & Riches, 2017; Boyd & Donnarumma, 2018; Montee *et al*., 2013; Restrepo, 2020).
>
> *A growing theoretical framework*: In the reported LAL initiatives, teachers learned about, or expanded, core concepts for language assessment as they were engaged in the critique and design of instruments (Kleinsasser, 2005; Kremmel *et al*., 2018; Nier, Donovan & Malone, 2009; O'Loughlin, 2006; Walters, 2010).

From the trends above, we confirm that the focus on teachers' LAL development has been particularly on the knowledge and skills components, rather than on principles. Davies (2008) states that principles are now prominent in language assessment, as this area has societal –and not just educational– impact[2]. Other authors have placed principles such as ethics and fairness as central in LAL descriptions for teachers. For example, in the study by Arias, Maturana and Restrepo (2012), there was particular attention to principles; this inclusion led the teachers in this program to develop democratic and transparent practices through including students' voices in assessment and providing them with clear information on assessment procedures. Unfortunately, we feel there is still limited information on principles to see trends regarding teachers' LAL (Giraldo, 2021; Inbar-Lourie, 2017; 2020).

2.3 LAL for Pre-Service Language Teachers

Existing LAL models and frameworks generalize the LAL of language teachers, which begs the question of whether they apply to pre-service teachers. What is clear, however, is the call for fostering these stakeholders' LAL. Authors claim that providing explicit education can better prepare them for their in-service work (Lam, 2015; López & Bernal, 2009; Restrepo, 2020; Vogt & Tsagari, 2014). It seems, then, that LAL discussions need refinement given the scarce differentiation between pre- and in-service teachers. This differentiation might lead to what authors call the era of language assessment *literacies* (Harding, 2021; Inbar-Lourie, 2017).

In their studies with pre-service teachers, Giraldo and Murcia (2019) planned and taught an assessment course based on stakeholder views, which brought attention to general national policies for assessment. They also found that the course promoted both theoretical and practical dimensions of assessment, especially during task design. In his study, Lam (2015) suggests that practicum courses for pre-service teachers target LAL knowledge, skills, and principles explicitly. This can be done through observation protocols and portfolios in which pre-service teachers can reflect on their LAL. Restrepo (2020) found that, through a learning journal, pre-service teachers conceived and reflected upon all the components of LAL, and therefore developed a broader conception of language assessment concepts and practices. Thus, the common thread among these studies is that the LAL components can be approached through different strategies and within different stages of LAL planning or development, e.g., when conducting needs analysis, designing instruments, discussing concepts explicitly or reflecting upon them. Given that the research arena is welcoming discussion (Inbar-Lourie, 2013; 2017), in this paper we hope to contribute to further problematizing the LAL of pre-service language teachers.

Moving beyond a discussion about the core knowledge, skills and principles of LAL, we seek to investigate the ways in which LAL is developed in instructional scenarios (Scarino, 2013). Thus, the main purpose of our paper is to report an

analysis of a series of classroom tasks which aimed to promote LAL among pre-service teachers in two language assessment courses of a state university in Brazil. Our analysis led us to elucidate which LAL components the classroom tasks prioritized and which tasks seemed to have promoted LAL the most. Against this background, we now explain the methodology to reach our findings, followed by relevant discussions.

3 Methodology

This study was conducted at a state university in Brazil in two elective assessment courses carried out in the first and second semesters of 2019. These courses' target audience were 3rd and 4th year students of both English and Spanish teacher education programs, with a total of 24 students in the first course, and 38 in the second. These pre-service teachers had already begun their teaching practicum and had not had specific classes or courses in language assessment before; they were mostly female and from the English teacher education program. Classes were given in Portuguese by three teacher educators: two PhD students, who, at the time, were fulfilling internship requirements, and the teacher educator responsible for the courses.

3.1 Description of courses

The first course focused on assessment of written production, and teacher educators worked with the pre-service teachers on the characteristics of specific textual genres (e.g., e-mail, poster, invitation) and how these are represented in assessment rubrics. Other activities involved analyses of activities in textbooks, as well as the evaluation of written productions of proficiency exam participants. This was followed by discussions on the grades given by the teachers in comparison to those of raters, for example, and in relation to the assessment criteria. There were also discussions of theoretical papers, and teacher educators provided pre-service teachers with both collective and individual feedback on their work (e.g., rubrics, critical reviews, didactic sequences), requesting activities to be redone when necessary. Checklists and rubrics were used to stimulate peer feedback and self-assessment in order to promote learner autonomy.

The second course was about the assessment of oral production. Pre-service teachers were familiarized with the assessment criteria commonly used for oral performance. Then, they practiced their judging and grading skills with recordings of proficiency exam participants, as well as with a simulation done in class. Again, there was an attempt to promote learner autonomy by means of self- and peer-assessment checklists. Pre-service teachers were provided with both collective and individual feedback and required to redo activities if these did not meet the agreed upon criteria. In order to stimulate critical thinking, they were asked to write critical reviews of papers so these could be discussed in class.

Against this backdrop, the research questions that underlay the study are as follows:

1. What LAL components were prioritized in the courses?

2. What classroom tasks promoted LAL the most?

We relied on a post-positivist approach to research, whereby the ideas and opinions of research participants are analyzed in their environment (Cohen, Manion & Morrison, 1998; Mackey & Gass, 2005). Specifically, we resorted to a case study research in which we used qualitative methods of data generation[3], namely class observations, document analysis, and a questionnaire for pre-service teachers.

Case studies enable detailed contextualization and in-depth insights into the specific research context, highlighting the challenges of the learning process, and therefore allowing for a comprehensive end-product (Mackey & Gass, 2005). They shed light into the complexities of specific contexts and specific groups, and since we were focusing on particular courses from a given context and the LAL particularities of this context, the case study approach seemed the most appropriate for our research goals.

3.2 Data generation and analysis procedures

After receiving students' signatures in the signed consent forms, following ethical research procedures[4], the first stage of data generation involved classroom observation. During this stage, observation protocols were used to enable note-taking on the tasks that took place during the courses. These were later compared to the courses' calendars to make sure all tasks carried out had been listed. Then, for stage two, these tasks were included in questionnaires given to pre-service teachers at the end of each course containing both open- and close-ended questions. Teacher educators received feedback from 20 out of 24 students from the first course, and 29 out of 38 from the second, totaling 49 questionnaire answers[5]. Our goal with the questionnaires was to have pre-service teachers evaluate, through Likert-type items, which tasks had enabled most learning, make overall comments about these tasks and point out the positive and negative aspects of the courses.

As for data analysis, pre-service teachers' answers to the Likert-type (close-ended) items were analyzed quantitatively, with tasks being ranked from top to bottom based on students' choice of the first option on the scale (*the activity/procedure helped a lot/enabled a lot of learning*). The open-ended answers were coded qualitatively through content analysis, whereby the content of the corpus is separated in preliminary codes, then goes through a process of peer revision and recoding until the researchers are left with final categories (Schreier, 2012). For instance, the categories *class management*, *class discussions* and *procedures* that arose from the comments ended up classified under the more general theme

teaching procedures. Similarly, *recognition of learning* and *relevance of courses* all fell under the theme *reflection*. These themes, then, became our research findings, which we describe and discuss below.

4 Findings and discussion

During the observation stage, we recognized a total of 30 classroom tasks, including learning activities (those mainly performed by students) and teaching procedures (those mainly performed by the teacher educators). We start our findings by presenting a list of the six tasks that ranked at the top and the six that ranked at the bottom. These were chosen because they represented the most interesting data to our analyses, as will be shown in the following sections.

In Table 2, A refers to 'activities' and P refers to 'procedures'. WP stands for the elective in assessment of Written Production, and OP stands for elective in assessment of Oral Performance. The percentages represent the number of pre-service teachers (henceforth, PSTs) who chose the first option on the Likert-type scale (*the activity/procedure helped a lot/enabled a lot of learning*)[6]. In this section, we give the participants in our study pseudonyms in accordance with ethical implications in research involving human beings.

Table 2 - Task ranking

Task rank	%	Course	A/P	Top-ranked tasks
01	95%	WP	A	Creation of rubrics
02	90%	WP	A	Analysis and assessment of a written production task from a proficiency exam
03	80%	WP	P	Providing collective feedback to learning tasks
04	79%	OP	P	Using materials (e.g., rubrics) from standardized tests for discussions about classroom-based assessment (e.g., language concepts, descriptors)
		OP	A	Assessment of proficiency exam participants' oral production based on (audio/video) recording
05	72%	OP	A	Assessment of proficiency exam participants' oral production based on a simulation performed in class
06	70%	WP	P	Discussing theoretical concepts through practical activities (e.g., discussing of language concepts through proficiency exams rubrics)
		WP	P	Adapting learning and assessment activities to a language in use perspective

Task rank	%	Course	A/P	Bottom-ranked tasks
13	48%	OP	P	Collective feedback on oral presentations guided by a monitoring sheet
		OP	P	Peer-assessment of written production (critical review) guided by a checklist
14	44%	OP	A	Discussion about the use of a self-assessment sheet to assess online-based oral productions
		OP	P	Providing collective feedback to a written production activity
		OP	A	Guided reading of theoretical papers
15	41%	OP	A	Oral presentation about proficiency exams in foreign languages
		OP	A	Peer-assessment/collective feedback on oral presentations guided by monitoring sheet
16	34%	OP	A	Peer-assessment on written production activity guided by checklist
		OP	A	Review of learning activities
17	20%	WP	A	Non-guided reading of a theoretical paper
18	13%	OP	A	Production of critical reviews

Source: the authors

Overall, pre-service teachers favored more hands-on tasks over theoretical ones, since the top student-led tasks were all of a practical nature, and some of the bottom tasks are mainly theoretical (for example, reading papers and writing critical reviews). Among the least evaluated tasks are also a few involving oral presentations about proficiency exams (e.g., tasks ranked 15) and peer-assessment (tasks ranked 15 and 16).

Regarding the results in Table 2, to answer the research questions we will first discuss tasks and how they related to LAL components; then, we will look at pre-service teachers' perspective on which tasks promoted LAL the most.

4.1 Major focuses on LAL that the tasks had

In reference to Giraldo's (2018)'s descriptors, it is possible to see that the top tasks were mostly skills-based, but they also included other LAL components. For instance, the creation of rubrics involved all three components in LAL: pre-service teachers had to articulate the nature and purpose of rubrics (this is referred to as knowledge in LAL, e.g., descriptor 18 from Giraldo's (2018) list), design the rubrics considering their characteristics (skills, e.g., item 45), and be able to explain scoring decisions in transparent and non-biased ways (principles, e.g., descriptor 57), therefore also igniting CLAL. The other top activities (2,

4 and 5) also required the use of rubrics and involved a combination of items from the three LAL components. For example, 'assessment of proficiency exam participants' oral production based on a simulation performed in class' required pre-service teachers to reflect on the different purposes of language assessment, such as placement, achievement and proficiency (knowledge, e.g., descriptor 12). Teachers also had to provide feedback on participants' performance (skills, e.g., descriptor 27) that was strictly related to the assessment criteria decided beforehand, not other aspects or bias towards the student (fairness as a principle in CLAL, e.g., descriptor 58).

As for the top teaching procedures (3, 4 and 6), pre-service teachers valued collective feedback delivery, which involved, from the point of view of teacher educators, understanding the importance of feedback (knowledge, e.g., descriptor 19), providing students with feedback on their performance (skills, e.g., descriptor 27), and making use of transparent assessment processes (principles, e.g., descriptor 66). Based on procedures 4 and 6, tasks involving discussions *about* theory (knowledge) led by the teacher educators and based on practical examples and tasks (skills) also had a positive impact.

The bottom tasks are closely related to knowledge only, and involve aspects such as reading and reflecting upon language learning theories and their impact on assessment practices (knowledge, e.g., descriptor 1). Because of their low ranking, we noticed a trend: it seemed that the more knowledge-centered the tasks were, the lower they were evaluated. Tasks that involved at least two components appeared to be more engaging. Furthermore, tasks related to performing peer-assessment, which demand a more active stance from the learner, were also amongst the lowest ranked. This might indicate that these pre-service teachers are not used to providing peer-feedback in their assessment process, more so if this was not part of their previous assessment-related experiences.

In short, the courses focused on language learning theories and their impact on assessment processes, discussing assessment concepts (e.g., washback, fairness, validity) though practical examples; they promoted mainly the practice of instructional skills (e.g., designing rubrics and instruments, providing feedback) and referred back to the principles constantly, not just by discussing the importance of ethics and transparency but by allowing pre-service teachers to participate in their own assessment process in an ethical way (e.g., by sharing/building together the courses' assessment criteria).

Within the qualitative data generated from the questionnaires, three main trends were identified: pre-service teachers' preference for practical activities, especially those applicable to their teaching contexts, over theoretical ones; their favorable opinion regarding teaching procedures and the educators' assessment practices; and their reflection on their learning process and recognition of the courses' relevance to their professional development.

4.2 Hands-on tasks for promoting knowledge and skills in LAL

The first trend we present, which is an expansion of the quantitative findings, is pre-service teachers' preference for practical activities over theoretical ones. This can be evidenced by some of the best evaluated tasks, which were 'creation of rubrics' (1), 'analysis and assessment of a written production task from a proficiency exam' (2), 'assessment of proficiency exam participants' oral production based on (audio/video) recording' (4), 'assessment of proficiency exam participants' oral production based on a simulation performed in class' (5), and 'discussing theoretical concepts through practical activities' (6).

More specifically, pre-service teachers reinforced the importance of rubrics (which were also present in tasks ranked 2, 4, 5 and 6) in the space for comments: "the analysis and creation of rubrics was essential" (PST Clara); "the creation of rubrics helped a lot and I will definitely use them again in the future" (PST Elisa); "the rubrics were really helpful" (PST Daniela). Pre-service teachers seemed to recognize the usefulness of rubrics for advancing their LAL and also their applicability to current or future teaching contexts.

The emphasis given to rubrics indicates that they could be an important tool in developing LAL not only from a technical point of view, but also from a theoretical one. Our participants' preference for practical over theoretical activities resonates with Berry, Sheehan and Munro's (2019) research with in-service teachers, in which they also reported, in relation to professional development needs, a stronger interest in practical aspects of assessment over theoretical ones.

Pre-service teachers' preference for practical tasks over theoretical ones might be a result of recognizing the tasks' usefulness to their work or internship contexts. PST Fernanda, for instance, thought the tasks "were very effective because they deal with what we [...] have to handle in the real world, in practice, in the classroom." PST Clara would have liked more activities related to classroom assessment as opposed to proficiency exams feedback, and PST Nadia suggests the creation of rubrics for their own teaching or internship contexts. This could indicate that teachers saw a connection between the tasks and their lifeworlds (Scarino, 2013), which may not be the case with theory-only tasks.

The fact that PSTs ranked theory-only tasks so low allows us to suggest that they need to be approached in a more familiar way to students in order to maximize LAL learning; in other words, theory needs to be made more concrete, and this may be achieved through practical tasks. For instance, an abstract concept such as *construct validity* might be better understood when pre-service teachers are faced with a task that involves this concept. Examples of theory-only tasks ranked low in our study were 'non-guided reading of a theoretical paper' (17) and 'production of critical reviews' (18). PST Olga, for example, reports that she "would have liked to have had closer monitoring on the writing of the critical review, considering some students had never written one before." Despite the fact that the teacher educators used a monitoring sheet

for the production of reviews and provided collective and individual feedback, it seems that the introduction of an unfamiliar textual genre deviated students' attention from assessment-related issues.

The finding above reiterates what Giraldo and Murcia (2019) recommend in their research study: practice-based tasks combining components of LAL are useful for language assessment courses because they trigger conscientious design across the board. In fact, models of LAL have remarked that LAL development requires attention to all three components of the craft (Fulcher, 2012; Giraldo, 2018; Taylor, 2013). Also, as we report in the literature review, professional development programs for teachers' LAL seem most positively impactful when hands-on tasks are central for learning (for example, see Arias *et al.*, 2012; Baker & Riches, 2017; Kremmel *et al.*, 2018). The impact of practical tasks may occur because teachers use their LAL skills, utilize theory conscientiously, and become aware of malpractice in language assessment. Thus, based on the results in our study, it is possible to highlight the power that hands-on tasks can have for pre-service teacher education in LAL.

4.3 Creating learning spaces for LAL development

As for the teaching procedures, the second trend in data characterization, pre-service teachers highlighted mainly the classroom discussions and the assessment procedures carried out in the courses. According to them, classroom debates were moments that enabled "the sharing of experiences and examples" (PST Fernanda) and "a deeper explanation of a topic, broadening perspectives" (PST Tatiana). PST Daniela mentioned that the courses had "great dynamics and great teachers, who were very didactic and open to debate." Teaching procedures ranked 4 and 6, namely 'using materials (e.g., rubrics) from standardized tests for discussions about classroom-based assessment (e.g., language concepts, descriptors)' and 'discussing theoretical concepts through practical activities' reinforce pre-service teachers' recognition of the discussions' potential in promoting their LAL. Furthermore, both of these procedures support the previous trend, in which pre-service teachers show they preferred practical tasks over theoretical ones.

As we stressed in the previous section, as a way to better engage pre-service teachers, theoretical concepts could be mostly delivered through discussions that arise from practical activities, especially considering teachers have reported they need LAL for more operational purposes rather than reflective ones (Giraldo, 2021). Furthermore, Kleinsasser (2005) and O'Loughlin (2006) found that a learner-centered approach to teaching about assessment, rather than a content-centered one, is more productive in terms of student engagement and involvement. Based on our finding, we then feel the role of the teacher educator is to ignite discussions in which students see theory of language assessment through the lens of practice.

Pre-service teachers also valued teacher educators' own assessment practices in the courses. PST Gisele stated, for example, that "experiencing meta-assessment

enabled me to learn a lot;" likewise, PST Bianca highlighted that the forms of assessment contributed to her learning. PST Fernanda reported that she learned "with two of the best teachers I ever had - they were more worried about students' *understanding and learning than their grades* (our emphasis)." This comment reiterates that the courses' focus was learning much more than grades or results, which is in line with the foremost goal of assessment: to promote student learning. PST Fernanda also mentioned that if the courses were regular, to her,

> It would definitely help educate teachers who are more aware and even more human - who understand students and try to cooperate with them. I feel like it would help mitigate the hierarchy between teacher and student. The course seems to follow a perspective of cooperation between teacher and student, and I find that essential, since no one is superior or inferior in a classroom: everyone can (and probably will) learn from one another and help one another. I think the course follows that perspective (PST Fernanda).

PST Fernanda's testimonial provides evidence that teacher educators' non-authoritarian stance helped build a cooperative learning environment, an aspect that was important in fostering LAL. Additionally, her comment reflects transparency in assessment, which is a principle in the LAL discussion (see Giraldo, 2018, for example).

The comments above also resonate with Restrepo's (2020) suggestion that teacher educators should "model sound assessment practices, and […] implement tools intended to monitor student teachers' actual understandings of what language assessment implies" (p. 49). Similarly, the Brazilian Educational Guidelines for Teacher Education (Brasil, 2002) argue that

> because teacher education takes place in a context similar to the one in which pre-service teachers will work in the future, there needs to be consistency between what is done in teacher education programs and what is expected of these teachers[7] (p. 2).

It seems that this consistency by course instructors between assessment procedures and assessment concepts enabled student learning, as it was one of the aspects brought up in the comments. For instance, teacher educators not only stressed the importance of sharing and building together assessment criteria, but they also promoted a space for that within the assessment courses.

We have reported that principles such as fairness and ethics in language assessment are not fully discussed when it comes to teacher education in LAL. Based on our findings, we suggest that principles can be instilled in a language assessment course by providing PSTs with opportunities to participate in course assessment, problematize assessment through their in-progress LAL, and, as explained in this section, aiming for transparency in assessment procedures. We acknowledge, however, that this is not yet a robust approach to the issue of

principles in LAL. More research is needed, as authors have argued (Giraldo, 2021; Inbar-Lourie, 2017; 2020).

4.4 Overall impact on pre-service teachers' professional development

The third main trend is related to pre-service teachers' reflection on their learning and recognition of the importance of the courses' content to their practicum. PST Aurora stated the following about the course: "I was able to better understand the assessment process, and this knowledge helped me understand my student better as well." PST Gisele had a similar perspective: "I learned immensely with the course […] it is one of the few in which I really recognized that I learned and was able to apply this knowledge efficiently." To PST Karina, "this course gave me a good notion of concepts that were not clear to me but were inserted in my teaching practice." Finally, PST Leila expressed her views on the tasks: they "facilitated my understanding of crucial concepts related to assessment of oral performance." The common thread among these comments is pre-service teachers' recognition of their learning mainly because they thought the concepts discussed and practices carried out were useful.

Therefore, by recognizing their learning (and the usefulness of the activities to their practicum), pre-service teachers also became aware of the importance of such courses to their professional development. Some authors (Furtoso, 2008; Lam, 2015; Stiggins, 1991) have reiterated the need for assessment courses in teacher education curricula, and this is reflected in pre-service teachers' answers.

They commented that the courses can "add a lot personally and professionally" (PST Mirela) and that they touched upon content they considered relevant for their education (PST Leila). To PST Fernanda, "in this course I had access to information that I consider essential and very important in teacher education" and she believes the courses should be regular and not elective.

As suggested by Giraldo (2021), "LAL may be a catalyst of reflection in professional development," and that seems to have been the case in the courses considering pre-service teachers' comments. Being able to reflect upon (and modify) one's own practice to better suit students' needs is a crucial skill for any teacher, and this reflection seems to have been promoted as a result of participation in the assessment courses. In other words, in our study LAL development also fostered reflection regarding language teachers' teaching and assessment practices (Giraldo, 2021).

Research studies with pre-service teachers have clearly indicated the overall impact language assessment courses can have. The studies have shown that PSTs see direct applicability to their realities as practitioners (Giraldo & Murcia, 2019; Jaramillo & Gil, 2019; Restrepo, 2020) and, overall, become more critical (e.g., by analyzing to what extent an assessment process is transparent) towards the task of language assessment and conscious of their roles in this dimension of their profession. In our study, it is worth highlighting that, despite being elective,

the two courses impacted PSTs' LAL, which then seemed to have translated into better preparation for their in-service work.

5 Final remarks

Conceptual and empirical discussions vis-à-vis in-service teachers' LAL are ongoing. However, research on pre-service teachers' LAL is beginning to emerge. Given this gap, our paper seeks to contribute to LAL discussions for these stakeholders. Thus, in the present paper we reported how classroom tasks prioritized different LAL components in language assessment courses for pre-service teachers; we also showed and discussed how these participants evaluated such tasks for their LAL development.

In synthesis, the top-ranked tasks prioritized the skills side of LAL, with knowledge and principles taking on a smaller role. This was possible thanks to PSTs critiquing and designing rubrics for assessment. Specifically, PSTs developed their knowledge of language concepts, theories and design skills. Regarding principles in LAL, we found some evidence to suggest that they can be approached through promoting discussions about the rationale behind assessment decisions (e.g., reasons for assigning a certain grade) and engaging PSTs in transparent assessment practices.

As for task evaluation, our findings indicate that practical tasks foster knowledge, skills, and, to a lesser extent, principles in LAL. This happened as PSTs were engaged in designing rubrics, grading oral and written productions, explaining and questioning grading choices, and analyzing language concepts and rubrics from proficiency exams, to name a few.

Theory-only tasks in language assessment courses for these PSTs may not be as useful: they may not find them relevant to their current teaching and, therefore, they do not see a connection to their lifeworlds (Scarino, 2013). Thus, we suggest that models of LAL may be best approached from a praxis-based lens, and one in which teacher educators and PSTs discuss pertinent practical, theoretical, and ethical issues. When this occurs, as it did in our study, PSTs may feel the positive overall impact of assessment on their professional development.

In terms of research limitations, we did not seek to identify how students conceptualize assessment and whether their view on assessment changed after the courses. However, future research into LAL development can investigate that, besides looking into effective classroom strategies, therefore providing richer data to substantiate findings.

We hope that our paper may fuel more LAL discussions and actions elsewhere. In the end, initiatives for developing pre-service teachers' LAL will ultimately have, we hope, a positive effect on the students whose language they will assess as in-service teachers.

Notes

1. After Fulcher's (2020) suggestion of using tasks for promoting stakeholders' LAL, we use 'classroom tasks' as an umbrella term to refer to 'learning activities' (those mainly performed by students) and 'teaching procedures' (those mainly performed by teacher educators).
2. Tests for immigration, citizenship, medicine practice, etc.
3. We use 'data generation' as opposed to 'data collection' because we understand that our data were researcher(s)-generated, i.e., "produced from their sources using qualitative research methods" (Garnham, 2008, p. 192) and did not exist naturally before the research - which would then entail data collection (Merriam, 2009).
4. Reference number at Plataforma Brasil: CAAE 19612319.1.0000.5231.
5. This does not imply 49 research participants, since some pre-service teachers took both courses and answered both questionnaires. We were unable to correlate their identity(ies) because some of them chose to respond anonymously to the questionnaire(s).
6. Instruments and data were written in Portuguese and translated into English by the authors as required.
7. In the original: "o preparo do professor, por ocorrer em lugar similar àquele em que vai atuar, demanda consistência entre o que faz na formação e o que dele se espera".

References

Allan, D. (2020). New Directions Webinar - Teacher Education & Training for Language Assessment [Webinar]. British Council. https://www.youtube.com/watch?v=qQGZygnTAJ4&feature=emb_logo

Arias, C. I., Maturana, L., & Restrepo, M. I. (2012). Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas [Evaluation in foreign language learning: Towards fair and democratic practices]. *Lenguaje, 40*(1), 99–126. https://doi.org/10.25100/ lenguaje.v40i1.4945

Baker, B. A., & Riches, C. (2017). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing, 35*(4), 557–581. https://doi.org/10.1177/0265532217716732

Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, *73*(2), 113–123. https://doi.org/10.1093/elt/ccy055

Boyd, E., & Donnarumma, D. (2018). Assessment literacy for teachers: A pilot study investigating the challenges, benefits and impact of assessment literacy training. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 105–126). Springer. https://doi.org/10.1007/978-3-319-77177-9_7

Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, C. MacNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge University Press.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3-12.

Cohen, L., Manion, L., & Morrison, K. (1998). *Research methods in education*. London, UK: Routledge.

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

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113-132. https://doi.org/10.1080/15434303.2011.642041.

Fulcher, G. (2020). Operationalizing assessment literacy. In D. Tsagari (Ed.) Language Assessment Literacy: From theory to practice (pp. 8 -28). Cambridge Scholars.

Furtoso, V. A. B. (2008). Interface entre avaliação e ensino-aprendizagem: Desafios na formação de professores. In A. B. A. B. Durão, O. G. Andrade, & S. Reis (Eds.), *Reflexões sobre o ensino das línguas estrangeiras* (pp. 127–158). Londrina: UEL.

Garnham, B. (2008). Data generation. In Given, L. M. (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 192-193). SAGE publications.

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development, 20*(1), 179–195. https://doi.org/10.15446/profile.v20n1.62089

Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development, 23*(2), 265–279. https://doi.org/10.15446/profile.v23n2.90533

Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal, 21*(2), 243–259. https://doi.org/10.14483/22487085.14514

Harding, L. (2021). Language assessment: Current trends, future challenges. SALT Virtual Lectures Series. https://www.youtube.com/watch?v=UljPM8H9ywI

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs* (Report: Part one – general findings). http://www.ealta.eu.org/documents/resources/ survey-report-pt1.pdf

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

Inbar-Lourie, O. (2012). Language assessment literacy. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). John Wiley & Sons. https://doi.org/10.1002/9781405198431.wbeal0605

Inbar-Lourie, O. (2013). Guest Editorial to the special issue on language assessment literacy. *Language Testing*, *30*(3), 301–307. https://doi.org/10.1177/0265532213480126

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (3rd ed., pp. 257–268). Springer. https://doi. org/10.1007/978-3-319-02261-1_19

Inbar-Lourie, O. (2020). LAL Reading Group Taylor 2013. International Language Testing Association. https://www.iltaonline.com/members/group_content_view.asp?group=219621&id=900692.

Jaramillo, J., & Gil, A. M. (2019). Pre-service English language teachers' use of reflective journals in an assessment and testing course. *Funlam Journal of Students' Research*, (4), 210–218. https://doi. org/10.21501/25007858.3010

Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, *20*(3), 77–102.

Kremmel, B., Eberharter, K., Holzknecht, F, & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173–194). Springer. https://doi.org/10.1007/978-3-319-77177-9_10

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197. https://doi.org/10.1177/0265532214554321.

Levi, T., & Inbar-Lourie, O. (2019). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, *17*(2), 168–182. https:// doi.org/10.1080/15434303.2019.1692347

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.

Mackey, A., & Gass, S. (2016). *Second language research methodology and design* (2nd ed.). London: Lawrence Erlbaum Associates, Inc.

Merriam, S. B., & Tisdell, E. J. (2009). *Qualitative research: A guide to design and implementation* (2nd ed.). John Wiley & Sons.

Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, *13*, 1–31.

Nier, V. C., Donovan, A. E., & Malone, M. E. (2009). Increasing assessment literacy among LCTL instructors through blended learning. *Journal of the National Council of Less Commonly Taught Languages*, *7*, 95–118.

O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, *1*, 71–85.

Restrepo, E. M. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of ELT Research*, *17*(1), 38–52.

Scaramucci, M. V. R. (2016). Letramento em avaliação (em contexto de línguas): Contribuições para a Linguística Aplicada, educação e sociedade. [Language assessment literacy: Contributions to Applied Linguistics, education and society]. In Jordão, C. M. (Ed.) *A Linguística Aplicada no Brasil: rumos e passagens* (pp. 141-165). Pontes Editores.

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

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Routledge.

Shohamy, E. (2022). Critical language testing, multilingualism and social justice. *TESOL Quarterly*, *56*(4), 1445-1457. https://doi.org/10.1002/tesq.3185

Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL*, *18*(1), 1–22.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534–539.

Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing*, 39(4), 649–678. https://doi.org/10.1177/02655322211057040

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, *30*(3), 403–412. https://doi. org/10.1177/0265532213480338

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374–402. https://doi. org/10.1080/15434303.2014.960046

Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, *7*(4), 317–342. https://doi.org/10.1080/15434303.2010.516042

Submission date: 20/10/2023
Acceptance date: 15/10/2024