

# But who trains the language teacher educator who trains the language teacher? An empirical investigation of Chilean EFL teacher educators' language assessment literacy

Language Testing

1–30

© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/02655322221134218

[journals.sagepub.com/home/ltj](https://journals.sagepub.com/home/ltj)**Salomé Villa Larenas** 

Universidad Alberto Hurtado, Chile

**Tineke Brunfaut** 

Lancaster University, UK

## Abstract

Research has shown that language teachers typically feel underprepared for assessment aspects of their job. One reason may relate to how teacher education programmes prepare future teachers in this area. Research insights into how and to what extent teacher educators train future language teachers in language assessment matters are scarce, however, as are insights into the language assessment literacy (LAL) of the teacher educators themselves. Additionally, while increasingly research insights are available on components that constitute LAL, how such components interrelate is largely unexplored. To help address these research gaps, we investigated the LAL of English as a Foreign Language teacher educators in Chile. Through interviews with 20 teacher educators and analysis of their language assessment materials, five LAL components were identified (language assessment knowledge, conceptions, context, practices, and learning), and two by-products of LAL (language assessor identity and self-efficacy). The components were found to interrelate in a complex manner, which we visualized with a model of concentric oval shapes, depicting how LAL is socially constructed (and re-constructed) from and for the specific context in which teacher educators' practices are immersed. We discuss implications for LAL conceptualisations and for LAL research methodology.

## Keywords

Language assessment literacy, language teachers, second language testing and assessment, teacher education, teacher educators

---

## Corresponding author:

Salomé Villa Larenas, Departamento de Inglés, Universidad Alberto Hurtado, Erasmo Escala 1835, Santiago, Chile 8340539.

Email: [svilla@uahurtado.cl](mailto:svilla@uahurtado.cl)

## **Pero ¿quién prepara al formador de profesores de idiomas?: Una investigación empírica de las competencias en evaluación de idiomas de formadores de docentes de inglés en Chile**

Las investigaciones han demostrado que los profesores de idiomas suelen sentirse poco preparados en la tarea de evaluar los aprendizajes de sus estudiantes. Una razón puede relacionarse con la forma en que los programas de formación inicial docente preparan en esta área. Sin embargo, existe escasa evidencia sobre cómo y en qué medida los formadores de docentes capacitan a los futuros profesores en evaluación de idiomas y sobre sus propias competencias en evaluación de idiomas (LAL, del concepto en inglés *language assessment literacy*). Por otro lado, si bien existe investigación incipiente sobre los componentes que constituyen LAL, la forma en que dichos componentes se interrelacionan está en gran parte inexplorada. Para ayudar a abordar esta brecha, investigamos el LAL de formadores de docentes de inglés como lengua extranjera en Chile. A través de entrevistas a 20 formadores de profesores y el análisis de sus materiales de evaluación de idiomas, se identificaron cinco componentes de LAL (conocimiento, concepciones, contexto, prácticas y aprendizaje sobre la evaluación de idiomas) y dos subproductos de LAL (identidad y autoeficacia como evaluador de idiomas). Los hallazgos del estudio muestran que los componentes se interrelacionan de una manera compleja, que ilustramos con un modelo de forma ovalada concéntrica, que representa cómo LAL se construye (y reconstruye) socialmente desde y para el contexto específico en el que están inmersas las prácticas de los formadores de docentes. Discutimos las implicaciones para las conceptualizaciones de LAL y para la metodología de investigación de LAL.

### **Introduction**

To date, language assessment literacy (LAL) research has largely focused on pre- and in-service language teachers, to understand their needs and develop their LAL (e.g., Hasselgreen et al., 2004). This has shown that language teachers around the world typically feel underprepared for the assessment-related aspects of their job (e.g., Lam, 2015). While there may be various reasons for this, a number of studies have called for insights to start from the earliest stages of the language teaching trajectory, that is, how teacher education programmes prepare future teachers for their professional tasks in language assessment (e.g., Vogt & Tsagari, 2014). A central role is thereby reserved for those conducting the pre-service training, that is, teacher educators. Hadar and Brody (2016), for example, point out that “teacher educators’ role in preparing the next generation of teachers is at the crux of educational innovation and effective schooling” (p. 58). Yet, they are a “neglected” group in the professional and research literature, including in LAL research. Little is known about how and to what extent language teacher educators train future language teachers in language assessment matters, or even what the LAL is of the teacher educators themselves (and thus how well-positioned they are to pass on or instigate any LAL). Therefore, the present study aimed to help fill this gap by investigating the LAL of language teacher educators. Empirically derived insights might inform LAL training and make it more targeted, which may in turn result in more effective training. In practice, the study was situated in Chile, focusing on English teacher education. In the following, we provide background information on this research setting. This is followed by a review of literature on LAL, enriched by relevant work on assessment literacy more

broadly (not specifically language-oriented). The remainder of the article describes and discusses the empirical study we conducted.

## English teacher education in Chile

In Chile, the main route to qualifying as an English teacher involves the successful completion of a 5-year undergraduate programme in English as a Foreign Language Teacher Education (henceforth, EFLTE). Since English is a foreign language in Chile, EFLTE programmes have the dual role of providing training in language pedagogy as well as developing the English proficiency of student-teachers. Therefore, although curricula vary somewhat between universities, EFLTE programmes typically comprise a battery of courses on (1) English language, language acquisition, and linguistics; (2) English teaching methodology; (3) school placements; and (4) other courses in general education and/or social sciences (Barahona, 2016). By the end of the programme, the novice English teachers are expected to have reached CEFR C1 level in English (Council of Europe, 2001) and possess the pedagogic competencies to teach English at Grades 5–12 of the school system.

In terms of language assessment training within Chilean EFLTE programmes, a 2017 systematic review of curricula indicated that only 52% of programmes included language assessment courses (Villa Larenas, 2020). Thus, whether Chilean English teachers have received language assessment training as part of their initial teacher education depends on which university they studied at, with a substantial proportion of teachers entering the profession without any language assessment training whatsoever.

Based on data from 2016, there are 31 accredited EFLTE programmes in Chile, offered by public (45%) and private (55%) universities, located in 11 of the 15 territorial regions (M. Silva, personal communication, January 10, 2017). Approximately 600 teacher educators are thought to work in these programmes, with about 300 of these teaching courses on English teaching methodology, English language acquisition, and language assessment (and others teaching non-language oriented courses). In theory, there are no specific requirements for becoming a teacher educator in Chile, but in practice, due to programme accreditation requirements (Comité Nacional de Acreditación), applicants to the role of language teacher educator in EFLTE programmes need to have (1) an English teaching qualification, (2) a master's degree in English Language Teaching or Linguistics, and (3) teaching experience in the Chilean school system. In terms of expertise in language assessment, since many of these teacher educators originally were English language teachers who went through Chilean EFLTE programmes as students themselves, the language assessment training of many is likely to have been limited, with potential implications for their LAL.

To gain insights into the LAL of Chilean EFL teacher educators, as a crucial stakeholder group for developing and shaping future language teachers' LAL, we conducted interviews with language teacher educators and also analysed assessment materials they use in the teacher training they offer. To ensure a well-grounded research design and to inform data collection and analysis methods, we first turned to the literature on assessment literacy—on one hand, work specifically situated in language testing, and on the other hand, work on teachers in more general educational assessment.

## Assessment literacy: Levels, components, models

The past few decades have seen a surge in scholarly interest in assessment literacy (AL) in the educational assessment arena, followed by *language* assessment literacy (LAL) in language testing. A currently influential definition in the latter field, although not making the language focus explicit, is Fulcher's (2012), which emphasizes skills-based know-how and the ability to understand this within a broader context:

The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order [sic] understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (Fulcher, 2012, p. 125)

Several researchers have argued, however, that different types of assessment stakeholders (e.g., test-takers, policy makers, teachers, test developers) do not require the same amount of knowledge, skills, and abilities regarding assessment. Within language assessment, for example, Pill and Harding (2013) proposed a continuum with five LAL levels, ranging from illiteracy, over nominal, functional, procedural and conceptual literacy, to multidimensional literacy. They suggested, for instance, that functional literacy might be an adequate level for policy makers, whereas language testers are expected to have multidimensional literacy. The labelling of the latter, highest level (multidimensional) is interesting, as it signals a multifaceted nature, with LAL comprising several components. However, Pill and Harding did not clearly specify what these components entail.

Taylor (2013) similarly hypothesized that different stakeholders (are expected to) have different levels of knowledge, skills and abilities in language assessment, and that LAL comprises several components. Taylor conceptualized eight components: knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes, and scores and decision-making. This led Taylor to a profiling approach in order to depict different stakeholder groups' LAL, combining the idea of levels and components: within one stakeholder group, there will be differences in the level of LAL expected depending on the component, and between stakeholder groups, there will be differences within the same component. For example, university administrators may need high LAL regarding scores and decision-making, but low LAL regarding language pedagogy.

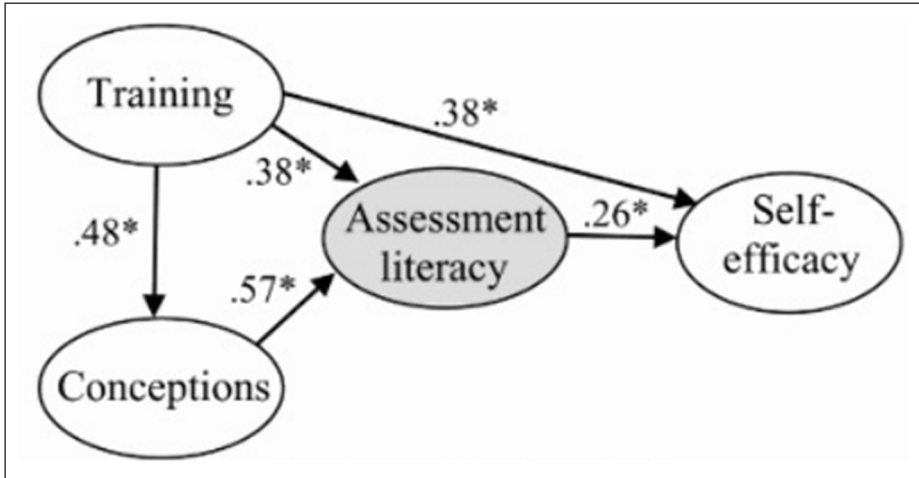
While Pill and Harding's and Taylor's conceptualizations constitute significant contributions to the LAL literature, these were not directly underpinned by empirical research. Empirically-confirmed or -derived conceptualisations of LAL, however, would constitute a convincing, well-grounded basis for LAL development initiatives and are likely to enable suitably targeted and effective LAL training. Recently, a few empirical studies have shed further light on LAL components. Baker and Riches (2018), for instance, investigated the LAL development of teachers and testers participating in language assessment workshops related to a national exam revision project in

Haiti. While Baker and Riches started from Taylor's eight LAL components, they made adaptations to better represent their empirical findings. For example, they collapsed Taylor's components "knowledge of theory" and "principles and concepts" into one component "theoretical and conceptual knowledge," as these were not easily distinguishable in their data. They also created a new component, "collaboration," as they found that co-operation between teachers and assessment specialists facilitated teachers' LAL development.

Another empirically grounded expansion regarding LAL components is Kremmel and Harding's (2020), which also aimed to validate Taylor's hypothesized LAL dimensions. Through an online survey of different stakeholders' self-perceived language assessment knowledge and skills, Kremmel and Harding identified nine LAL components: developing and administering language assessments; assessment in language pedagogy; assessment policy and local practices; personal beliefs and attitudes; statistical and research methods; assessment principles and interpretation; language structure, use and development; washback and preparation; and scoring and rating. These components partly confirmed Taylor's, but also empirically indicated that some of Taylor's hypothesized components were not distinct, while other components needed separating out. As an example of the former, the exploratory factor analysis which Kremmel and Harding ran on their dataset indicated that Taylor's two proposed components of "sociocultural values" and "local practices" rather constituted one component only, which Kremmel and Harding labelled "assessment policy and local practices." An example of the latter is that, as borne out by the factor analysis results, Taylor's proposed "language pedagogy" component needed to be divided into two separate components, which Kremmel and Harding labelled "assessment in language pedagogy" and "washback and preparation."

In line with the above literature, we adopted a componential view to study the LAL of Chilean EFL teacher educators. Furthermore, given the perspective that distinct LAL profiles may be justified for different types of stakeholders, we felt encouraged in our specific focus on the stakeholder group of language teacher educators, which has hardly received attention in LAL research.

However, one issue these recent empirical studies in language testing do not clarify is how various LAL components relate to one another. Here, we found research in educational assessment helpful, where apart from multidimensional depictions of AL, frameworks have also tried to capture the interrelationship among AL components. Two models seemed particularly relevant to our study as they focus on teachers. First, Levy-Vered and Alhija (2015) conducted an AL study with beginning teachers of a wide range of subjects in Israel. By means of structural equation modelling, they investigated the interrelationship between beginning teachers' assessment literacy and their assessment training, self-efficacy, and conceptions of assessment. In this manner, they represented AL as multidimensional rather than through the traditional division by levels, and they recognized internal, personal factors that could affect teacher learning about assessment and how such factors interrelate. Their best-fitting model (see Figure 1) showed the impact of assessment training and conceptions on beginning teachers' assessment literacy, while teachers' self-efficacy was influenced by their assessment literacy (and training) rather than the other way around.



**Figure 1.** Levy-Vered and Alhija's (2015, p. 393) teacher assessment literacy model (\* $p \leq .05$ ).

Another feature of Levy-Vered and Alhija's (2015) model relevant to this study is that it involved the teacher educator as an essential agent in the process of AL development; it measured teacher educators' modelling of assessment practices to their student-teachers as part of the AL model's "training" component. One limitation of Levy-Vered and Alhija's model, however, is its primary focus on teachers' internal factors, while overlooking external contextual aspects such as wider socio-political or cultural factors, which may affect teachers' assessment learning and practices.

The second model from educational assessment relevant to the present study addresses the latter issue, as it considers both teachers' internal and external worlds. This model—Xu and Brown's (2016) *Teacher Assessment Literacy in Practice* framework (TALiP)—is presented in Figure 2.

Informed by a synthesis of the AL literature, Xu and Brown (2016) described AL as consisting of six components—a knowledge base; teacher conceptions of assessment; institutional and sociocultural contexts; teacher assessment literacy in practice; teacher learning; and assessor identity (re)construction—but added that AL is "dependent on a combination of cognitive traits, affective and belief systems, and sociocultural and institutional influences, all of which are central to teacher education" (p. 155). Xu and Brown visually represented these components' relationships with a pyramidal shape, which illustrates their thinking on how teachers become assessment literate. Namely, according to Xu and Brown, the foundation of AL is a knowledge base on assessment. This knowledge base, however, is filtered and interpreted by teachers' own beliefs about assessment and by the assessment conditions in their professional context. In turn, these influence teachers' assessment practices. Ultimately, then, the interrelationship of teachers' assessment knowledge and practices, and their opportunities for learning about assessment, are thought to culminate in the construction of teachers' own identity as assessors.

While ground-breaking in several ways, it should be noted that the TALiP is a theoretical model rather than an empirically-validated visualization of AL components and their

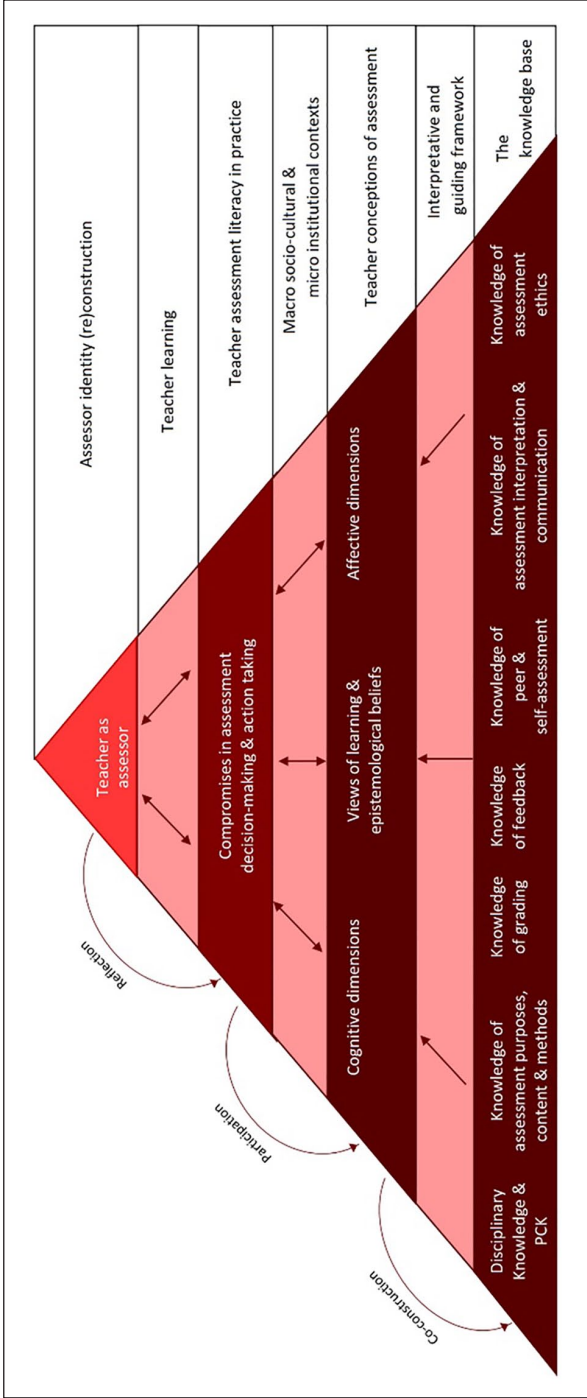


Figure 2. Teacher Assessment Literacy in Practice framework (TALiP) by Xu and Brown (2016, p. 155).



interconnections. Its pyramidal shape, for example, might be questionable, and seems to conflict with Xu and Brown's own statement that their AL framework is cyclical in nature. This shape conveys the ideas of hierarchical importance and foundational conditions, that is, the knowledge base as the component with the most "weight" in teachers' AL and as a pre-condition for the development of assessment literacy. Yet, in reality, this might not be the case, especially in contexts with little/no assessment training (as seems to be the case in our study's research context), and thus, there may be a very small knowledge base and little to develop assessment literacy from (assuming Xu and Brown's visualization). Thus, the relationship between assessment knowledge and practice might not be hierarchical, as suggested by the pyramid. Empirical investigations of the TALiP's components and their hypothesized relationships are therefore needed. It should also be noted that the TALiP was designed for the context of educational assessment (i.e., AL); it is unknown to what extent it transfers to the specific context of assessing languages (i.e., LAL). Regardless of these challenges, the TALiP's hypothesized, wide-ranging components and their interrelationships seemed a useful starting point to try and characterize the LAL of language teacher educators.

### **This study: Aims and research questions**

The above literature review shows that, within the field of language testing, we have already gained valuable conceptual and empirical insights into LAL in general and on types of LAL components more specifically. We know less, however, about how LAL components interrelate. In the broader AL literature, researchers have specifically started thinking about such componential interrelationships. Xu and Brown's (2016) TALiP framework is a notable example for our study, given its focus on teachers and consideration of both teacher-internal and -external factors, but it is theory-based rather than empirically-derived. The present study, therefore, seemed a meaningful opportunity to explore the interrelationship of components for the specific area of *language* assessment, and to do so in an *empirical* manner, thus filling gaps in existing LAL and AL research, respectively.

Additionally, it should be noted that the TALiP was designed with specific reference to teachers as assessment stakeholders—indeed, a key stakeholder group, and the main focus of most LAL research too. However, while the present study was similarly situated in the teaching context, its focus on *language teacher educators* offered an opportunity to broaden our insights into wider stakeholder groups and shed light on a distinct subgroup in the teaching profession. Indeed, those who train language teachers (and are assumed to help develop teachers' LAL) constitute an underrepresented population in current research.

Consequently, our study aimed to address the following research questions to characterize the LAL of EFL teacher educators:

RQ1: What components constitute the language assessment literacy of Chilean EFL teacher educators?

RQ2: What are the relationships between the components?



## Method

To answer these research questions, we interviewed Chilean EFL teacher educators and analysed assessment materials they used in their courses and which they had brought along to the interviews. We now provide more details on our study's methodology.

### *Participants*

Out of 31 Chilean universities with EFLTE programmes, 10 universities were invited for participation, based on their accredited programme status, reputation, and ease of access. Six universities (four public and two private) agreed, representing three of the 11 regions with accredited EFLTE programmes. From these, 20 language teacher educators volunteered to participate in the study. 60% were female, and 40% were male. Ninety percent were Chilean, Spanish-L1 speakers, 10% were English-L1 speakers. Ninety percent were qualified English language teachers, 10% were linguists; all held master's degrees. Eighty percent had more than five years' experience as an EFL teacher educator. In their teacher education programme, at the time of this study, they taught (a combination of) English language acquisition (14 teacher educators), language teaching methodology (4), language testing (3), and/or coordinated teaching practicums (3) or managed the programme (6). Nevertheless, all had taught English language acquisition in their EFLTE programmes in previous years. In the following, participants (P) are represented with a randomly allocated number (P1–P20).

### *Data collection*

To gain insights into the nature of the language teacher educators' LAL, interviews were conducted. These lasted 45 to 90 minutes and were audio-recorded. Participants were given a choice of interview language, with 16 preferring English and four Spanish (translations from Spanish are marked below with \*).

First, in a semi-structured phase, participants were asked questions regarding their professional background and working contexts, their language assessment knowledge and practices, their beliefs and feelings about language assessment, their previous training in language assessment, their self-perceptions as language assessors, how language assessment was approached/taught in their institutions, and what they taught about language assessment in their contexts and how. The interview guide (see Supplementary File 1) was inspired by the TALiP components, but also left room for expansion on topics interviewees might deem relevant.

For a second, unstructured interview phase, participants were asked to bring along (language) assessment materials from the courses they taught (18 out of 20 shared materials). The materials acted as a prompt to elicit conversation on topics and issues related to the participants' language assessment practices which had not naturally emerged in the first part of the interview, for example, because the teacher educator had not remembered or because we had not anticipated it. In practice, the teacher educators ended up talking about their materials' purpose, development, uses, and perceived quality, but also—significantly—about issues concerning institutional factors, assessment culture, and the language teacher educator community. Another purpose for asking to bring along materials was that this allowed us to

analyse the materials' design and content afterwards, to gain additional insights into teacher educators' language assessment knowledge and practices as observed in these "artefacts."

We would like to note that the study was restricted to formal, marked assessments. Although any teacher educators' assessment practices may go beyond formal assessment tasks, and may include formative evaluations such as live corrective feedback while teaching, for practical and logistic reasons, data collection was limited to "opportunities" of assessment that were more visibly and explicitly recorded as such (see Hill & McNamara, 2012).

### *Data analysis*

The interview recordings were transcribed and then analysed thematically in Atlas.ti. To develop a coding scheme, we started off from the six TALiP components and tried these out on the dataset. This indicated the need for adaptations and additions to more comprehensively capture the information and insights provided by teacher educators during the interviews. Through an iterative process of coding scheme development and piloting, we arrived at a final coding scheme (see Supplementary File 2) comprising six high-level codes with multiple subcodes each.

The teacher educators' language assessment materials were analysed using the same coding scheme, but restricted to the high-level code "language assessment knowledge" and its subcodes only, as this captures the relevant and observable information for materials analysis.

Twenty-five percent of the data was double-coded by another, independent coder—a teacher educator trained in language teaching and assessment, and familiar with the Chilean EFLTE context. The inter-coder agreement was .80—categorized "good" by Mackey and Gass (2016).

For full methodological details, and extensive examples of the findings, see Villa Larenas (2020).

## **Results**

In the following, we present the findings of our analyses according to the high-level codes of our coding scheme. As will be argued, this provides empirical evidence that the LAL of EFL teacher educators in Chile can be characterized through five components: language assessment knowledge, conceptions of language assessment, context, language assessment practices, and language assessment learning (RQ1). As we describe these components, specific connections between them will transpire (RQ2). Additionally, the data revealed insights into language teacher educators' identity as language assessors and self-efficacy regarding language assessment. However, as we will explain, the findings suggest that these (identity and self-efficacy) are the result of LAL development rather than components of LAL themselves.

### *Language assessment knowledge*

The dataset indicated that the language teacher educators possessed some knowledge about language assessment, albeit patchy in some crucial areas. Some of this knowledge

transpired from their interview responses and some was displayed through a range of language assessment practices evidenced in their assessment materials. Following Inbar-Lourie (2008), we describe this in terms of teacher educators' knowledge of *what* to assess and *how* to assess. Notably, however, the teacher educators reported that they had not received specific language assessment training in their careers—at most, educational assessment training. They also mentioned that the knowledge that they possessed most likely came largely from instances of collaboration in their work contexts, as will be discussed below.

*Knowledge of what to assess.* Knowledge of *what* to assess concerns “the description of the trait to be assessed” (Inbar-Lourie, 2017, p. 264). The data suggested that the EFL teacher educators' knowledge of linguistic constructs is limited. Their language assessment practices, as described in their responses and seen in their materials, showed a lack of consideration of language constructs while developing language assessment tasks. When discussing their materials, most teacher educators were doubtful about what linguistic skills their tests were targeting, especially for the receptive skills. Furthermore, instances of construct-irrelevance and construct-underrepresentation were observed in their tests. For example, P5's reading comprehension test included an item that read: “Read the target sentence and provide a similar sentence based on the text.” Essentially, the item required paraphrasing an individual sentence, which could be done without reading the rest of the text and which shifted the construct to writing.

*Knowledge of how to assess.* Knowledge of *how* to assess concerns insights into theories and research on assessment procedures and methods—understanding of language assessment principles, types and methods, design, evaluation of instruments, and scoring and grading. The teacher educators shared the following language assessment materials: 12 speaking tasks, nine end-of-unit quizzes, seven tests (proficiency and achievement), four projects, three writing tasks, and one extensive reading test. This showed that the teacher educators are familiar with and use a range of language assessment types in their practices. This aligns with research on language teachers (e.g., Berry et al., 2019), which reported their familiarity with and/or use of several assessment techniques. The materials also revealed that language teacher educators sometimes simply adopt existing tests, but also adapt tests from other sources (the Internet, international exams, colleagues) or develop their own instruments. In doing so, the materials demonstrated that they used different types of items, with typically more than one item type being included within one test. The most commonly used item types were gap-fill, long- and short-constructed response, matching, sequencing, and sentence completion. Others found were as follows: true-false, multiple-choice, chart completion, and sentence correction/rewriting/ordering.

Regarding how to design language assessments, the teacher educators stated that this was often done based on agreements reached during team meetings, without development of test specifications. Instead, a strong influence of international language examinations was seen on test development, which, in some cases, served as test specifications for teacher educators' own instrument development. Alternatively, the teacher educators simply used existing instruments. For example, P6 reported using tasks from Cambridge English's Preliminary English Test (PET):

- P6: We design only the grammar and writing [tasks]. And the listening and the reading [. . .] we take them from these PET books. [. . .]
- I: So, you select listening and reading sections from this book?
- P6: One reading and one listening.

As for the teacher educators' knowledge of how to evaluate language tests, the interviews showed a lack of knowledge on evaluating assessment instruments pre- and post-administration. The teacher educators reported that, at most, they conducted peer proofreading pre-administration or they held an informal team conversation to decide on grading when post-administration results were not as expected. This resonates with Vogt and Tzagari (2014), who found that language teachers "do not seem to be in a position to critically evaluate their tests: establish reliability and validity or do statistical analysis of the results to gauge [their] quality" (p. 385). In this study, P5 for example stated,

I just usually reflect, when I see the results, on "ok, this was useful," or "this was too hard," or maybe, "I should have less exercises of this type."

Only those three teacher educators in the present study who taught language testing courses said they performed systematic test result analyses to model language assessment practices to their student-teachers, which led to the evaluation of these teacher educators' own tests.

Regarding knowledge of how to score and grade, first, some of the teacher educators' materials included scoring information. The teacher educators explained in the interviews that their scoring system was typically based on how difficult they judged an item to be (more points for more difficult items). However, this "level of difficulty" rationale for maximum score allocation was not consistent across the dataset; the materials analysis found no clear criteria for point allocations. For example, Figure 3 shows four different sections from P6's midterm test, in which items were assigned either 0.5 or 1 point. However, the rationale underlying the point allocation was unclear: Sections I and II seemed of similar difficulty, but were assigned 0.5 and 1 point per item, respectively; Section VI required student-teachers to paraphrase a sentence, which seems a more difficult task than the one in V, which requests correcting a sentence, but both sections were assigned 1 point per item. Only the teacher educators who taught language testing courses consistently provided scoring information and went beyond the perceived difficulty rationale when allocating scores.

Second, the teacher educators showed understanding of rating instruments since all their productive tests included rating tools, typically analytic rubrics. Third, all participants valued the opportunity to provide performance feedback to their student-teachers; however, they explained that this feedback mainly focused on language accuracy; not on the targeted language constructs.

In sum, EFL teacher educators' language assessment knowledge was partial and did not result from prior, formal language assessment training. Instead, the dataset indicated that language teacher educators have *practice-derived language assessment knowledge*, formed on the job.

<p><b>I. FILL IN THE GAPS WITH A PREPOSITION AND THE RIGHT FORM OF THE VERB: 6 pts (0,5 each)</b></p> <p>1. Tom is afraid _____ (visit) the doctor.</p>
<p><b>II. Complete with Present Perfect Simple or Continuous. 7 pts (1 each)</b></p> <p>1. I _____ (not see) my best friend in years so I _____ (wait) for this moment for a long time.</p>
<p><b>V. Find and correct the mistake in these sentences. 6pts (1 each)</b></p> <p>1. What people really like about Valdivia? _____</p>
<p><b>VI. Rewrite these sentences using gerunds: (4 pts., 1 each)</b></p> <p>1. To ride a horse can be quite pleasant (if you know how to do it)</p>

**Figure 3.** Example of score distribution in midterm test.

### *Conceptions of language assessment*

The teacher educators' conceptions of language assessment were explored in terms of affective aspects, understood as "emotional inclinations that teachers have about various aspects and uses of assessment," and cognitive aspects, understood as "what teachers believe is true and false about assessment" (Xu & Brown, 2016, p. 156).

*Affective aspects.* The teacher educators described mostly negative emotions about language assessment during the interviews:

I don't think I like [assessing] that much. Assessing is very difficult because, in the end, it's always a subjective thing. (P4)

They expressed that such emotions came from negative previous assessment experiences in their student lives, which they described as "unpleasant" (P4) and "negative" (P13), such as perceived unfair results due to lack of scoring instruments (P4), or public sharing of scores resulting in bullying (P13). Other interviewees expressed negative emotions because they did not like how language assessment was conducted in their programme:

I don't like [assessment] by the way, I have to say it, I hate it actually. I don't believe very much in the way they do it here. (P9)

However, the data also indicated that these negative perceptions can change with language assessment learning. Namely, a shift towards more positive perceptions was expressed by teacher educators who had had more opportunities for learning about language assessment theory through their graduate studies or through teaching language testing courses on their teacher education programme:

[After my master's,] I changed that [negative] view and I think that assessment is positive. It's necessary. It's part of what we do all the time, we assess. (P14\*)

This reflects other studies' findings (e.g., López & Bernal, 2009) that perceptions of (language) assessment are influenced by training, i.e., more positive attitudes towards assessment are associated with more assessment knowledge.

*Cognitive aspects.* Overall, the teacher educators indicated favouring formative over summative assessment. P9 said,

I like formative. It's not that I don't believe in summative tests. It's that I want to see the product, the bits and pieces.

They also felt a strong connection between the process of teaching and assessing languages. P13 stated,

The things that come to my mind when you ask [about assessment] are not about assessment. They're about teaching. And the teaching is assessment.

Notably, they thought that their beliefs about language assessment critically impacted their decisions about what and how to assess. However, they also emphasized that their beliefs required balancing with the contextual demands at the moment of assessing, as elaborated below.

## Context

The dataset showed that teachers educators' present and past educational and professional contexts influenced their language assessment knowledge and their conceptions of language assessment, and also affected their language assessment practices.

*Context and language assessment knowledge.* The teacher educators argued that their study and work contexts—past and present—influenced their language assessment knowledge. First, regarding past context, the training that the teacher educators received influenced their opportunities of formal access to principled language assessment knowledge. The teacher educators reported lacking access to language assessment training during their own teacher education. This limited access to language assessment learning might, for

example, have contributed to the teacher educators' limited understanding/awareness of the language constructs assessed in their tests, as reported above.

Seventeen teacher educators reported receiving training only in educational assessment during their own teacher training, but not specifically catered for language assessment. This training consisted of a single course for all teacher education programmes at their universities, covering aspects of assessment generic to all disciplines. The teacher educators described it as mostly theoretical and disconnected from their language teaching practice. P12\* said,

[The course] was too much theory without practice [. . .], it didn't have anything to do with English. So, I can't tell you that I learned a lot in that subject.

The interviewees felt it was "not simple" (P7) to transfer their understandings of educational assessment from these courses to their language assessment practice. This finding suggests that the educational assessment training alone which many language teacher educators receive is not sufficient. This echoes Levi and Inbar-Lourie's (2019) conclusions, for a study on language teachers: although language teachers can successfully transfer some generic educational measurement knowledge gained through AL courses to their practices, "the acquisition and development of meaningful language teaching-learning-assessment literacy requires complementary language assessment training" (p. 13).

Second, the teacher educators' present professional context was also seen to affect their access to language assessment knowledge. The interview data showed that their learning about language assessment mainly occurred on the job, through collaboration with colleagues. For instance, P10 said,

[W]e get together, and we think of what, how we might assess [. . .] and then we check each other's work. Some of the teachers have great ideas, and I have used them in my instruments as well.

This corroborates the earlier observation that teacher educators gain some insights into language assessment through their practices.

*Context and conceptions of assessment.* We also found that the teacher educators' past and present educational and professional contexts shaped their conceptions of language assessment. Regarding their past context, teacher educators' negative prior experiences affected the way they perceived language assessment now. This aligns with existing research (e.g., Crossman, 2007) which found that teachers' emotions and perspectives on assessment, whether positive or negative, partly arose from their past assessment experiences.

Regarding the teacher educators' present context, half of the participants' responses suggested a clash between teacher educators' beliefs about assessment and their university's internal assessment culture. P9's quote above, on affective aspects, illustrates this. This incongruence seemed to affect the teacher educators' language assessment practices as they reported having to adjust to the demands of their programmes.



*Context and language assessment practices.* Contextual variables also appeared to influence the teacher educators' language assessment decision-making and practices. At a macro-sociocultural level, national policies were reported to influence the types and designs of language assessments adopted by the teacher educators. For example, in Chile, CEFR C1 is the required language proficiency level for graduating English teacher trainees (Ministerio de Educación-República de Chile, 2014). Consequently, EFLTE programmes started using international examinations (e.g., IELTS, Aptis) to certify student-teachers' English proficiency level when graduating. Eight participants explained that they had been instructed by their programme administrators to use international exams as final exams and/or develop tests with the same format as those exams. The teacher educators referred to the latter as "standardized tests," although these were usually self-developed tests lacking any validation or standardization procedures. The interviewees reported, however, that there was a conflict between such international/standardized tests and their teaching practices since these tests often do not align with what and how they teach. For example, P2 complained,

There is also the whole limitation of a standardized test, and what it measures. Ok, it's fine, it measures skills, but the students [. . .] complain: "but why do we have this test that has none of the contents that we saw during the semester?"

At a micro-institutional level, the internal assessment culture of the teacher educators' workplace affected their language assessment decision-making. Each programme's assessment culture is shared informally in meetings—by word of mouth—and indirectly—through language assessment materials that are passed on from year to year and which act as templates for assessment design, indicating the skills targeted and item types. The teacher educators explained that this internal assessment culture sets the boundaries for their language assessment decision-making, determining the nature and types of assessments they are expected to develop. For example, P7 explained that he would like to create new types of speaking tasks but that he has to follow the task formats used by his institution in previous years.

### *Language assessment practices*

We found that a combination of factors influenced the language assessment practices of the teacher educators: their language assessment knowledge, professional context, and conceptions of language assessment had an impact on their decision-making around language assessments.

*Practices influenced by knowledge.* The teacher educators' language assessment practices were seen to be affected by their (perceived) limited language assessment knowledge caused by lack of training. The teacher educators reported that they tended to "borrow" language assessment materials and make "instinctive" decisions in their assessment practices. While they sometimes developed language assessment materials from scratch themselves, they preferred using or adapting existing "ready-made" language assessment materials retrieved from other sources (e.g., the Internet, external exams, their colleagues):

I take them from the Internet. I usually do a mix of things that I think could be useful because I feel that I lack training in that. (P2\*)

Such “borrowing” practices were observed even when the assessment materials were not the most suitable ones for the context. P2\* admitted,

I didn't like [this test]. The truth is that I used the template of a test which a colleague [. . .] shared with me. It's only grammar. I didn't like it very much, but well . . .

The teacher educators explained that their “borrowing” practices were largely motivated by time constraints and because they felt underprepared in language assessment and not very confident when developing their own materials. Berry et al. (2019) made similar observations regarding teachers' language assessment practices.

Teacher educators' feelings of underpreparedness also led them to making “instinctive” decisions on language assessments in their job. Five participants stated that, due to their lack of formal language assessment training, they made rather intuitive practical decisions about (language) assessment, as illustrated by P3:

It's just by instinct, basically, because it's not something that I've been taught how to do. So, is it too general? Is it too specific? Is it too detailed? [. . .] I really don't know.

*Practices influenced by the context.* The dataset revealed that the teacher educators' language assessment practices were furthermore influenced by the assessment culture at their institutions. While a variety of assessment materials was observed across the dataset, little variation was found in the assessment tasks of individual teacher educators or within individual courses. Howley et al. (2013) pointed out that “teachers incorporate into their practice those strategies that make sense in their context and within their own professional frameworks for decision making” (p. 33). Accordingly, the teacher educators in this study tended to repeat the same task types over the years and stick to those that matched their institution's internal assessment culture. For example, P4 explained that their listening tests always included two item types—multiple-choice and open question—out of “tradition” as the “tests [were] always like that.”

The impact of teacher educators' working contexts on their practices also transpired from their self-developed language assessment instruments. These mostly comprised discrete-point items. This practice can mainly be ascribed to the fact that the teacher educators were mirroring the international exams used at their institutions.

Additionally, the context sometimes influenced the skills targeted to be assessed. For example, in three cases, teacher educators' test tasks focused largely on grammar. They explained that their institution's internal assessment culture was largely characterized by a traditional grammar-oriented approach, viewing grammar as one of the most important linguistic aspects that future English teachers should master. Thus, these teacher educators' assessment practices reflected their institutional context.

*Practices influenced by conceptions of assessment.* Teacher educators' beliefs appeared influential when making language assessment decisions. For example, many felt that

assessing writing was time-consuming and thus they avoided testing it. However, three participants who attached great importance to assessing writing still included it in their practices, despite knowing they would struggle with time constraints. P6 said,

[Assessing writing] is something very important for us. But it implies a lot of work. [. . .] It's difficult because it takes [a lot of] time [. . .]. But we still do it all the time.

At the same time, the interview data indicated that the influence of teacher educators' beliefs on their assessment decision-making was conditioned by the level of autonomy they have to make such decisions in their institutional context. Prior studies (e.g., James & Pedder, 2006) similarly argued that the context exerts an influence on how teacher beliefs are materialized. Xu and Brown (2016) maintained that

[t]he tighter the boundaries, the less space there is for professional autonomy. Tensions arise for teachers when they have less autonomy [. . .] and can arise because of incongruence between their conceptions and the boundaries imposed upon them within their context. (p. 157)

In this study, when contextual boundaries were tight, teacher educators adjusted their practices to the educational demands imposed on them. For example, most teacher educators held negative perceptions on summative assessment and the "standardized" tests used in their programmes. However, their assessment materials and interview responses indicated that they predominantly created summative tests, due to their institution's internal language assessment culture, through the use of international examinations. In sum, this study found that language teacher educators' beliefs set the scene for their language assessment decision-making, provided they have autonomy to make decisions in their language assessment practice.

### *Language assessment learning*

As mentioned earlier, the teacher educators reported that due to lack of formal language assessment training, they reverted to other ways of learning about language assessment. While seven participants said they made efforts to do self-study, the main way in which the teacher educators reported to learn about language assessment was on the job, especially by collaborating with colleagues when deciding on and developing assessments. In some cases, junior colleagues teamed up with more experienced ones who acted as mentors. For example, P5 said, "[t]he first teacher I worked with, she taught me all about how the tests work." In other cases, groups of teacher educators worked together on assessments. They felt that such cooperation instances offered them opportunities for reflection on their own assessment practices (through sharing ideas for test development, receiving feedback on test design, co-constructing tasks, discussing marking, etc.) and constituted one of the richest spaces for learning about language assessment. This corroborates findings from studies with language teachers which concluded that collaboration in professional development facilitates teachers' assessment learning (e.g., Baker & Riches, 2018; Harsch et al., 2021).

### *Language assessor identity*

With reference to language teachers, Burns and Richards (2009) defined teacher identity as “how individuals see themselves and how they enact their roles within different settings” (p. 5). Accordingly, language assessor identity in this study is understood as how language teacher educators see themselves as language assessors and how they enact their roles as assessors in their contexts.

The interviews revealed that most teacher educators (17) did not identify themselves as language assessors, but rather as teachers and teacher trainers, although they need to regularly evaluate their student-teachers’ language ability and/or convey (language assessment) knowledge and practices to student-teachers. For example, when explicitly and repeatedly asked about their strengths as assessors, 12 teacher educators talked about their role as teachers or teacher educators rather than as assessors. Furthermore, three expressed that they had “no strengths at all” (P9) as assessors. Only the three teacher educators who taught language testing courses self-identified as language assessors, describing their strengths as assessors or expressing the importance of their roles in enacting assessment practices in their contexts. For example, P12\* said,

The main thing is that I understand assessment as assessment for learning. And as such I use it as a learning tool an important learning tool, a motivational tool.

The language assessor self-identification of these three teacher educators is interesting, given that they had similar backgrounds to the other teacher educators, lacking prior language assessment training (with the exception of one). But they had been assigned the programme’s language testing course and thus had the opportunity/needed to learn about it in that manner.

Overall, the findings suggest that EFL teacher educators’ language assessor identity rather is the *result* of the nature of their LAL (which was richer in those who teach language testing courses) instead of constituting a component of LAL itself. Thus, language assessor identity can be argued to follow from LAL.

### *Self-efficacy*

Initially, we explored confidence regarding one’s own language assessment practices as a subtheme of conceptions of language assessment. However, we found that confidence featured as a constant theme in the interviews, and thus substantiated more careful consideration. We were also strengthened in this decision by the findings from Levy-Vered and Alhija’s (2015) study described earlier.

All teacher educators expressed some degree of insecurity about their language assessment practices:

I don’t feel very confident. (P1)

I don’t know if I’m doing it right. (P4)

I think that I don't know much [about language assessment]. I feel like, in terms of theory, I know very little. (P5)

This echoes the lack of confidence regarding assessment found in other studies with language teachers (e.g., Berry et al., 2019).

Confidence in language assessment practices resonates with the concept of self-efficacy, which has been defined as the “beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (Bandura, 1997, p. 3). Teacher educators stated that their low levels of self-efficacy were due to their lack of theoretical language assessment knowledge because of the deficient training they had received. Conversely, they felt that opportunities for learning through collaboration helped increase their levels of confidence in their assessment practices:

[A]ll I know about language assessment comes from working with [colleagues] [. . .] I think I'm much more confident with designing and applying rubrics now than I was before. (P3)

This reflects the findings of other studies which associated collaboration with self-efficacy in the context of teachers' assessment practices (e.g., Ciampa & Gallagher, 2016).

This study's findings therefore suggest that *language assessment self-efficacy* is a concept that is relevant to teacher educators' LAL, yet it is influenced by LAL development.

## Discussion

This study aimed to characterize the LAL of EFL teacher educators by exploring what components constitute their LAL and how these interrelate.

### *LAL components*

The empirical findings presented above suggest that the LAL of language teacher educators in Chile can be characterized through five components: language assessment knowledge, conceptions of language assessment, context, language assessment practices, and language assessment learning (RQ1). These matched five of the six TALiP components theorized by Xu and Brown (2016) and comprised the teacher-internal components from Levy-Vered and Alhija (2015). The novelty in our study is that these components were applied to the specific area of language assessment (as opposed to educational assessment in the TALiP) and the stakeholder group of teacher educators (as opposed to classroom teachers). The sixth TALiP component—assessor identity—also featured in our dataset, but not so much as something that formed part and parcel of language teacher educators' LAL and characterized it, but rather as something that followed from their LAL. Thus, it does not seem to be a LAL component as such.

As discussed in the literature review, to date, within the field of language assessment, most descriptions of LAL components remain theoretical. Recent empirically-informed descriptions, however, are Baker and Riches' (2018) update of Taylor's (2013) LAL heuristic, and the emerging components from Kremmel and Harding's (2020) LAL

**Table 1.** Language assessment literacy components identified in recent empirical studies.

This study	Baker & Riches (2018)	Kremmel & Harding (2020)
Language assessment knowledge ( <i>what, how, why</i> )	<ul style="list-style-type: none"> <li>• Theoretical and conceptual knowledge</li> <li>• Language pedagogy</li> </ul>	<ul style="list-style-type: none"> <li>• Assessment in language pedagogy</li> <li>• Assessment principles and score interpretation</li> <li>• Language structure, use, and development</li> <li>• Washback and preparation</li> <li>• Scoring and rating</li> <li>• Assessment policy and local practices</li> </ul>
Contextual variables ( <i>past: training &amp; previous experiences</i> ) ( <i>current: macro- &amp; micro-context</i> )	<ul style="list-style-type: none"> <li>• Awareness of local practices</li> <li>• Sociocultural values</li> </ul>	<ul style="list-style-type: none"> <li>• Personal beliefs and attitudes</li> </ul>
Conceptions of language assessment ( <i>emotions &amp; beliefs</i> )	<ul style="list-style-type: none"> <li>• Awareness of personal beliefs/attitudes</li> </ul>	<ul style="list-style-type: none"> <li>• Personal beliefs and attitudes</li> </ul>
Language assessment practices	<ul style="list-style-type: none"> <li>• Decision-making</li> <li>• Task performance</li> </ul>	<ul style="list-style-type: none"> <li>• Developing and administering language assessments</li> <li>• Statistical and research methods</li> </ul>

survey. It therefore seems meaningful to establish any commonalities or differences between these LAL studies’ findings and the present study’s, as any emerging patterns can provide insight into potential generalizability of the findings. Table 1 shows the result of our mapping exercise of LAL components across the three studies.

Table 1 shows clear overlap between the three studies’ components, although the present study’s components appear to act as an empirically-grounded synthesis or “umbrella” conceptualisations of LAL components. The components proposed by Baker and Riches, and Kremmel and Harding, might be seen as subcomponents of the “umbrella” components identified in this study.

A notable difference between the studies is found in the last of this study’s components—language assessment learning. This component was not identified by Kremmel and Harding; however, it should be kept in mind that their study involved a needs analysis survey and also targeted stakeholders outside the language teaching profession. Language assessment learning was not part of their declared construct, and LAL development was not part of the focus of their study. In the case of Baker and Riches, they proposed a component labelled “collaboration,” which can be argued to be one type of opportunity for learning about language assessment. Their study’s methodology might explain this more “niche” component; it was based on a series of workshops which explored LAL development in “practice,” and collaboration in the workshop context constituted an important factor for language assessment learning.

The lack of a language assessment learning component in Kremmel and Harding's study, versus the identification of such a component in Baker and Riches' and in this study suggests, however, that LAL research could benefit from research methodologies which explore the stakeholders' language assessment *practices* by means of instruments other than surveys (see also Levi & Inbar-Lourie, 2019).

### *Relationship between LAL components*

An important contribution of the present study is that it also aimed to establish how the different LAL components connect or influence each other (RQ2), which was not explored in prior empirical studies. The findings indicated that Chilean EFL teacher educators' LAL is not a linear process with a literacy progression that starts from an assessment knowledge base, as was suggested by Xu and Brown (2016). In reality, such a "base" may be inexistent; indeed, the teacher educators in this study had not received language assessment knowledge training during their own degree studies and instead developed any language assessment knowledge while on the job. Similar observations were made in Berry et al.'s (2019) study regarding language teachers' assessment learning. Furthermore, the present study revealed several complex and often simultaneous interactions between LAL components. For example, the teacher educators' language assessment practices and decision-making were influenced by factors in their professional context, their language assessment knowledge, their conceptions of language assessment, and the balance and interactions between these components.

Figure 4 illustrates the LAL components and their interrelationships as they were found for the group of teacher educators in this study. The findings are visualized in a model of concentric oval shape that embraces the five components which were found to constitute Chilean EFL teacher educators' LAL: context, language assessment knowledge, conceptions of language assessment, language assessment practices, and language assessment learning. Outside of the ovals, two concepts are depicted which were found to be products or outcomes of LAL rather than components of it—language assessor identity and language assessment self-efficacy. This is clearly distinct from the pyramidal visualization and component interrelationships of Xu and Brown's (2016) TALiP framework, which does not accurately capture the interrelationships observed in our study with teacher educators in language education. We now describe our empirically-based model in more detail, pointing out differences and similarities with existing models and studies.

*Outer oval: Context component.* Xu and Brown (2016) put forward "context of assessment" as an AL component, and also argued that "micro- and macro-contextual variables exert an influence on teachers' assessment practices" (p. 157). This study showed that contextual variables have an influence on various dimensions of teacher educators' LAL. Xu and Brown, however, had "simply" represented context as one layer (in-between two others) in their TALiP pyramid, which seems an underestimation of this component's varied and extensive impact in empirical reality. Therefore, we depicted the context component as surrounding the other four LAL components in the oval shape, to represent the wide-ranging effect of contextual factors. This also resembles Baker and Riches' (2018)



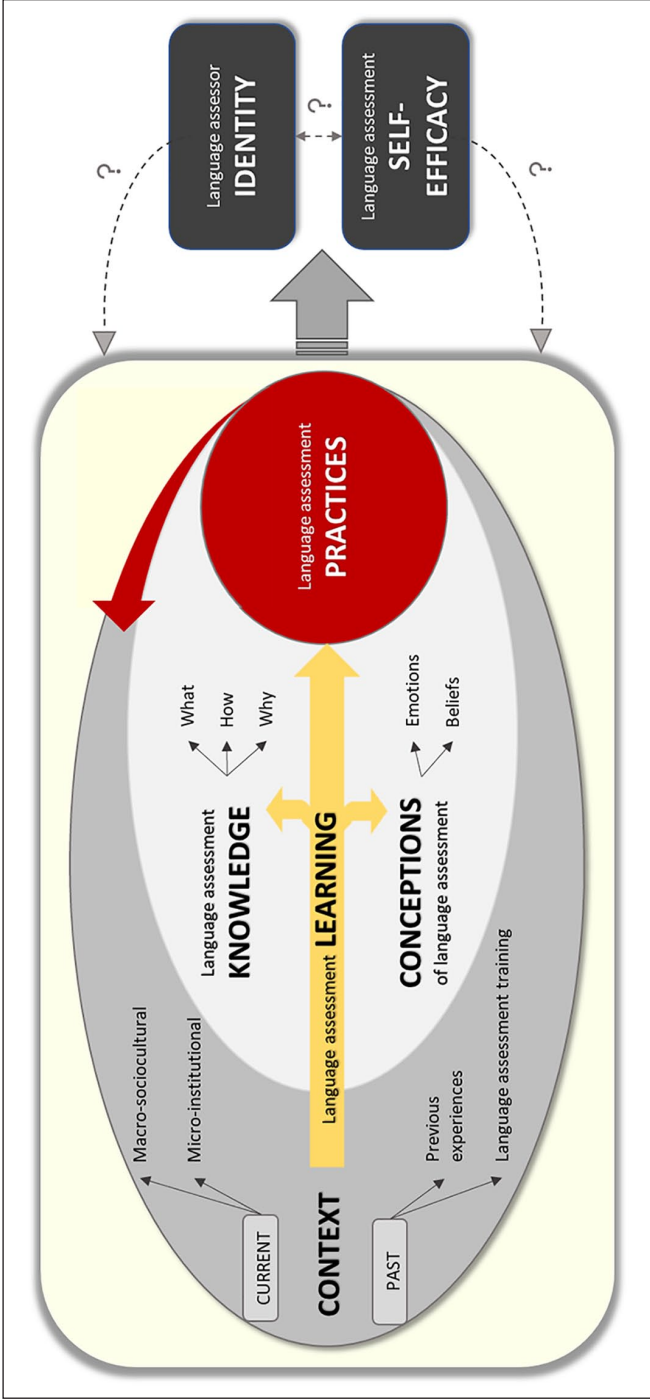


Figure 4. EFL teacher educators' language assessment literacy: components and interrelationships.

decision to place their “sociocultural values” component (which can be seen as a sub-component of context) as surrounding their updated LAL profile heuristic to “symbolize how it informs all the other elements” (p. 574).

Our findings also identified different types of contextual factors playing a role in teacher educators’ LAL. We reflected this insight in the model by distinguishing between *past* and *current* contextual factors. *Past context* comprises (1) the *language assessment training* stakeholders have received through formal language assessment courses, workshops, or professional development programmes and (2) *previous experiences* with (language) assessment which stakeholders have had. *Current context* involves the subcategories suggested by Xu and Brown—macro-sociocultural and micro-institutional factors—which emerge from the socio-political and educational settings in which stakeholders work.

#### *Middle oval: Language assessment knowledge and conceptions components*

*Language assessment knowledge.* The nature of what it is to “know” is an ongoing debate in the philosophical community; epistemologists have proposed a difference between knowing about something (declarative knowledge; knowledge-that) and knowing how to do something (procedural knowledge; knowledge-how) (e.g., Ryle, 1945). Our findings align with this, as they identified knowledge about “what” and “how” to assess to be part of the language assessment knowledge component. Additionally, we found that language assessment knowledge emerges from interactions immersed in the sociocultural and political context in which assessment is deployed, and that it is informed by that context. Indeed, Scarino (2013) argued that “[n]ot only do teachers need to understand the conceptual bases of different approaches, they also need to relate such knowledge to their professional practice in their particular context” (p. 310). Therefore, teaching professionals should be aware of the rationale and role of language assessment in their communities, and understand national policies and how these impact institutional and individual decisions in their practices. Inbar-Lourie (2008) called this type of knowledge the “why” of language assessment.

The practice-derived language assessment knowledge which was observed in the present study conflicts with Xu and Brown’s conceptualization of knowledge as “the basis of all the other components” (p. 155). Our results suggest that language assessment knowledge is instead a complex, dynamic and evolving component which is not only acquired through formal training but also socially constructed through informal ways of learning occurring in stakeholders’ practices.

Therefore, to reflect the knowledge component’s dynamic nature and the fact that it is influenced by the context and in turn exerts an influence on the beliefs and practices of teaching professionals, we placed it in a middle oval.

*Conceptions of language assessment.* Scarino (2013) argues that it is also important in teachers’ development of LAL to consider “the ‘inner’ world of teachers and their personal frameworks of knowledge and understanding and the way these shape their conceptualizations, interpretations, decisions and judgments in assessment” (p. 316). This inner world of beliefs, emotions, etc. was also observed in our data and is represented by the component *conceptions of language assessment*.

Similar to what has been argued for teachers (e.g., Berry et al., 2019), this study found that teacher educators' previous experiences with assessment as students influenced how they feel about language assessment. Unfortunately, these experiences, and consequently teacher educators' emotions about language assessment, were mostly negative. We also found that teacher educators' beliefs influenced their language assessment practices and helped them balance the constraints exerted by their context. However, this influence also depended greatly on the teacher educators' level of autonomy to make decisions in their contexts. Therefore, we placed the conceptions component in the middle oval to reflect its relationship to other components: (1) being affected by the context (previous experiences, training) and (2) influencing language assessment practices depending on level of autonomy for decision-making.

*Knowledge and conceptions relationship.* Xu and Brown (2016) argued that conceptions of assessment filter the knowledge acquired through training: "teachers tend to adopt new knowledge, ideas, and strategies of assessment that are congruent with their conceptions of assessment, while rejecting those that are not" (p. 156). However, teacher educators' lack of language assessment training meant that it was not possible to verify this claim in our study. But to not overlook the plausible influence of conceptions on knowledge, we have assumed this relationship in the model. Our data did reveal, however, that language assessment knowledge shapes teacher educators' conceptions of language assessment, which is not something Xu and Brown had hypothesized. Language assessment learning on the job was seen to have a positive impact on teacher educators' beliefs and emotions about language assessment. This confirms Levy-Vered and Alhija's (2015) and López and Bernal's (2009) findings that training in assessment has a positive effect on teachers' conceptions of assessment, and in turn, a positive impact on teachers' AL.

Based on this, we positioned the language assessment knowledge and conceptions components within the same oval and depict their relationship as reciprocal. Nevertheless, future research needs to confirm whether there is an influence of conceptions on knowledge.

*Inner circle: Language assessment practices component.* Xu and Brown (2016) hypothesized that teachers' assessment practices are the result of compromises they make to reconcile tensions between their own beliefs and external factors. This study confirmed this as we found that teacher educators' language assessment practices resulted from their language assessment decision-making, balancing their language assessment knowledge, conceptions, and the influences of contextual variables. This supports recent conceptualizations of (L)AL which go beyond traditional conceptions centred on knowledge, skills and principles, by embracing the individual's conceptions of assessment (e.g., Levy-Vered & Alhija, 2015) and their sociocultural and political contexts (e.g., Willis et al., 2013). Language assessment practices therefore seem to be the ultimate manifestation of LAL. Additionally, the findings indicated that teacher educators' language assessment practices influence the micro-institutional context where they work; their practices influenced those of their colleagues by sharing and co-constructing language testing materials in teams. Consequently, we added an arrow from the language assessment practices component to the context component of the model.

Taken together, the above findings regarding language assessment practices indicate that it forms a central component of LAL, influenced by and interconnected with all other components. Therefore, we have depicted it as an inner circle in the model.

*Crosswise arrow: Language assessment learning component.* This study showed that learning through collaboration influences teacher educators' language assessment knowledge, conceptions, and practices. Nevertheless, the relationship and positioning of the language assessment learning component in the TALiP framework seems inadequate. Xu and Brown (2016) described teacher learning as a necessary component for advancement in assessment literacy; thus, they placed it high up in the pyramid. However, in this study's context in which language assessment training was deficient (as also found in other studies, e.g., Berry et al., 2019; Tsagari & Vogt, 2017), teacher educators developed their language assessment knowledge through learning on the job. In reality, teaching professionals in several contexts around the world might need to resort to their communities of practice to learn about language assessment in the first place. Therefore, our findings suggest that teacher learning is often a starting point for acquiring language assessment knowledge; not just "the impetus for advancing" (Xu & Brown, 2016, p. 157) in the process of becoming assessment literate.

With this in mind, the language assessment learning component is illustrated as a crosswise arrow in Figure 4, starting from the context and influencing the practices, with two more arrows showing its effect on knowledge and conceptions.

*By-products of LAL: Language assessor identity and self-efficacy.* Besides the five components found to constitute the LAL of EFL teacher educators, two other concepts surfaced from the data: language assessor identity and language assessment self-efficacy. Different from the other five, these two emerged as a result of the combination of the other five components, i.e., as a result of the LAL development of the teacher educators.

In the case of *language assessor identity*, while Xu and Brown (2016) considered assessor identity to be "the ultimate goal of TALiP" (p. 158), this study found that it is a sense of ownership of language assessment knowledge and practices which leads teacher educators to perceive themselves as language assessors. This was seen in teacher educators who teach language testing courses and talked about their identity as language assessors and positioned themselves in a particular assessment paradigm.

As for *language assessment self-efficacy*, we observed a generalized sense of insecurity in teacher educators regarding their language assessment knowledge and practices, mainly due to their perceived low LAL. Prior research on assessment self-efficacy has identified variables which influence it. For example, Hartell et al. (2015) found a relationship between the extent of assessment training and teachers' self-efficacy levels, i.e., the more prepared teachers felt, the higher was their sense of self-efficacy regarding assessment. Similarly, as mentioned earlier, Levy-Vered and Alhija (2015) found that levels of AL predict levels of self-efficacy: "teachers' feeling of assessment self-efficacy is an outcome of their assessment ability and of the knowledge and skills they acquired through training and their conceptions of assessment" (p. 394). In line with this, in this study instances of language assessment learning were reported to directly affect the teacher educators' self-efficacy.

The data suggests therefore that language assessor identity and language assessment self-efficacy are not components of LAL itself, but rather by-products of it, i.e., they seem to depend on the LAL development of stakeholders. Consequently, we positioned them outside of the LAL oval. Whether the relationship is reciprocal—identity and self-efficacy in turn determining teacher educators' LAL—was not observable in our findings, unfortunately. Similarly, the connection between the two by-products themselves (identity and self-efficacy) could not be established from our data. Further research needs to shed light on these matters; thus, we represent these uncertainties with dotted lines and question marks in Figure 4.

## Conclusion

This study set out to characterize the nature of language teacher educators' language assessment literacy (LAL). By means of interviews and course materials, it identified five LAL components (language assessment knowledge, conceptions, context, practices, and learning) and two by-products (language assessor identity and self-efficacy). It also established the relationships between the LAL components in a complex dynamic construct of LAL in practice; to our knowledge, our study constitutes the first empirically-grounded description of this. We visualized our findings with a model of concentric oval shape which depicts Chilean EFL teacher educators' LAL components and their interrelationships (Figure 4).

Our findings suggest a sociocultural construct for LAL, consisting of the interrelationships of the five components, whereby language teacher educators' current and past context, and their relationships with their peers are key factors in their language assessment literacy. LAL was found to be a complex system which is socially constructed (and re-constructed) from and for the specific context in which stakeholders' practices are immersed. This aligns with contemporary views of assessment literacy in which AL is seen as a negotiation between stakeholders' inner and outer worlds.

Our findings invite us to rethink how language teaching professionals' LAL is understood and studied. They suggest the need to go beyond describing LAL through levels or needs and call for a characterization of LAL through an understanding of connections between LAL components in teaching professionals' realities. A rich exploration of their contextual working conditions, local assessment culture, access to language assessment knowledge, and assessment belief system appears necessary to understand their LAL and develop suitable LAL training programmes. Consequently, the findings invite LAL research to go beyond survey-only or interview-only methodologies, and employ mixed-methods approaches with, for example, observations, document analyses, or ethnography. Our interviewees, for instance, remembered and elaborated on their practices, beliefs, and contextual constraints only when going through the assessment materials they brought along.

We acknowledge, of course, that our visual model was abstracted from a study on one particular setting (Chile) and a specific stakeholder group (language teacher educators). Studies using similar methods in other settings would be highly welcome, as well as research that explores the model for other types of stakeholders such as language teachers more generally. The fact that our componential results align with those of various

other LAL studies, focussing on teachers, is promising for the potential wider applicability of our model, though. Future research may also aim to fill some gaps in the present study's details, such as whether teacher educators' conceptions of language assessment filter their learning of language assessment knowledge, or what the interrelationships between language assessment self-efficacy and language assessor identity are, and their potential effects on teacher educators' LAL.

A number of practical implications can also be drawn from this study's insights into the nature of Chilean teacher educators' LAL (which may be valuable beyond this geographical setting too). First, our findings indicate that there is a demand and need for the LAL development of Chilean teacher educators, and thus initiatives in this area would be welcome. Ideally, public educational policy would recognize, facilitate, instigate, and reward this type of professional development. Secondly, given the importance of the context in LAL, any LAL development programmes should be localized, i.e. take into account the sociocultural and political realities of teacher educators' working contexts and be contextualized for the local language assessment needs (the assessments' focus, purpose, scale, stakes, policy regulations, etc.). Finally, it is recommended that LAL development is a collaborative and longitudinal undertaking, supported through relevant and practical EFLTE management structures and guidance, and instilling communities of practice, co-construction of assessments, and space for dialogic reflection on language assessment practices.

### Declaration of conflicting interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Agencia Nacional de Investigación y Desarrollo of Chile with a 2015 BecasChile grant. It was also part-supported by the British Council with an 2017 Assessment Research Award, and by The International Research Foundation for English Language Education with a 2018 Doctoral Dissertation Grant.

### ORCID iDs

Salomé Villa Larenas  <https://orcid.org/0000-0002-3563-7517>

Tineke Brunfaut  <https://orcid.org/0000-0001-8018-8004>

### Supplemental material

Supplemental material for this article is available online.

### References

- Baker, B. A., & Riches, C. (2018). 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>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.



- Barahona, M. (2016). *English language teacher education in Chile: A cultural historical activity theory perspective*. Routledge.
- 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>
- Burns, A., & Richards, J. (Eds.). (2009). Introduction: Second language teacher education. In A. Burns & J. Richards (Eds.), *Cambridge guide to second language teacher education* (pp. 1–8). Cambridge University Press.
- Ciampa, K., & Gallagher, T. (2016). Teacher collaborative inquiry in the context of literacy education: Examining the effects on teacher self-efficacy, instructional and assessment practices. *Teachers and Teaching*, 22(7), 858–878. <https://doi.org/10.1080/13540602.2016.1185821>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research and Development*, 26(3), 313–327.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Hadar, L., & Brody, D. (2016). Professional development for teacher educators in the communal context: Factors which promote and hinder learning. In B. de Wever, R. Vanderlinde, M. Tuytens & A. Aelterman (Eds.), *Professional learning in education challenges for teacher educators, teachers and student teachers* (pp. 57–83). Academia Press.
- Harsch, C., Seyferth, S., & Villa Larenas, S. (2021). Evaluating a collaborative and responsive project to develop language assessment literacy. *Language Learning in Higher Education*, 11(2), 311–342. <https://doi.org/10.1515/cercles-2021-2020>
- Hartell, E., Gumaelius, L., & Svärth, J. (2015). Investigating technology teachers' self-efficacy on assessment. *International Journal of Technology and Design Education*, 25(3), 321–337. <https://doi.org/10.1007/s10798-014-9285-9>
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs: General findings* (Needs Assessment No. 1). University of Bergen.
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420. <https://doi.org/10.1177/0265532211428317>
- Howley, M. D., Howley, A., Henning, J. E., Gilla, M. B., & Weade, G. (2013). Intersecting domains of assessment knowledge: School typologies based on interviews with secondary teachers. *Educational Assessment*, 18(1), 26–48. <https://doi.org/10.1080/10627197.2013.761527>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <http://doi.org/10.1177/0265532208090158>
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (3rd ed., pp. 257–270). Springer.
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, 17(2), 109–138. <https://doi.org/10.1080/09585170600792712>
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120. <https://doi.org/10.1080/15434303.2019.1674855>
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <http://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>
- Levy-Vered, A., & Alhija, F. N. A. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation*, 21(5–6), 378–406. <http://doi.org/10.1080/13803611.2015.1117980>
- López, A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language Assessment. *Profile*, 11(2), 55–70. <https://revistas.unal.edu.co/index.php/profile/article/view/11442>
- Mackey, A., & Gass, S. M. (2016). *Second language research: Methodology and design* (2nd ed.). Taylor & Francis.
- MINEDUC, Ministerio de Educación-República de Chile [Ministry of Education – Republic of Chile]. (2014). *Estándares orientadores para carreras de Pedagogía en Inglés* [Guiding standards for EFL teacher education programmes]. <https://bibliotecadigital.mineduc.cl/handle/20.500.12365/2237>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 23(3), 381–402. <http://doi.org/10.1177/0265532213480337>
- Ryle, G. (1945). Knowing how and knowing that: The presidential address. *Proceedings of the Aristotelian Society*, 46, 1–16. <http://www.jstor.org/stable/4544405>
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 7(3), 309–328. <https://doi.org/10.1177/0265532213480128>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 7(3), 403–413. <http://doi.org/10.1177/0265532213480338>
- 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.
- Villa Larenas, S. (2020). *An investigation of the language assessment literacy of teacher educators in Chile: Knowledge, practices, learning, beliefs, and context* [Unpublished doctoral dissertation, Lancaster University].
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <http://doi.org/10.1080/15434303.2014.960046>
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *Australian Educational Researcher*, 40(2), 241–256. <https://doi.org/10.1007/s13384-013-0089-9>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <http://doi.org/10.1016/j.tate.2016.05.010>