

IN-SERVICE AND PRE-SERVICE TEACHERS' LITERACY AND AFFECTIVE DISPOSITION TOWARDS ENGLISH LANGUAGE ASSESSMENT

CLAUDIO DÍAZ LARENAS*¹
Universidad de Concepción (Chile)
ORCID: 0000-0003-2394-2378

JOSÉ ALMONACID REBOLLEDO
Universidad de Concepción (Chile)

GHIREN CANAHUATE HENRÍQUEZ
Universidad de Concepción (Chile)

AMPARO PEIRANO CHANDIA
Universidad de Concepción (Chile)

ABSTRACT: Assessment is an essential tool to measure students' progress. It is paramount for teachers to understand their classroom assessment practices to make informed educational decisions. This study seeks to identify in-service and pre-service teachers' understanding of knowledge and affective dispositions towards language assessment through two instruments: the *English Language Assessment Literacy Inventory (ELALI)* and the *Affective Disposition Scale (ADS)*. 78 in-service teachers and 132 pre-service teachers participated in the study. The findings revealed differences between in-service and pre-service teachers' responses to the *ELALI*, yet overall both groups demonstrated deficiencies. The results of the *ADS* showed that both groups had a similar perception of language assessment. The analysis of the data obtained helped to identify that, although in-service and pre-service teachers have a weak understanding of English language assessments, they regard assessment positively and focus on students' learning process rather than the grading of a final product.

KEYWORDS: language assessment, assessment literacy, affective disposition

* Para correspondencia, dirigirse a: Claudio Díaz Larenas (claudiódiaz@udec.cl).

¹ This paper is in the context of the research grants FONDECYT 1220307 Estudio sobre el diseño de instrumentos de evaluación del idioma inglés: procesos y carga cognitiva, respuesta afectiva y desempeños de candidatos a profesores y FONDECYT 1191021 entitled Estudio correlacional y propuesta de intervención en evaluación del aprendizaje del inglés: las dimensiones cognitiva, afectiva y social del proceso evaluativo del idioma extranjero. Investigador responsable: Dr. Claudio Díaz.

EL CONOCIMIENTO Y DISPOSICIÓN AFECTIVA HACIA LA EVALUACIÓN DEL INGLÉS DE PROFESORES Y PROFESORES EN FORMACIÓN

RESUMEN: La evaluación es una herramienta esencial para medir el progreso de los estudiantes. Es importante que los profesores entiendan sus prácticas evaluativas para poder tomar decisiones informadas. Este estudio busca identificar el conocimiento, y la disposición afectiva de profesores en ejercicio y formación. En el estudio participaron 78 profesores en ejercicio y 132 profesores en formación quienes respondieron dos instrumentos: el Inventario de Alfabetización en Evaluación del Inglés, y la Escala de Disposición Afectiva hacia la Evaluación. Los resultados revelaron que hay diferencias en las respuestas entre ambos grupos en los instrumentos, pero, ambos grupos mostraron deficiencias. En la Escala de Disposición Afectiva las respuestas de ambos grupos fueron similares respecto a la percepción de la evaluación del inglés. Estos resultados concluyen que, aunque los profesores y los profesores en formación tienen un conocimiento insuficiente sobre la evaluación del inglés, ellos sí perciben la evaluación como un proceso positivo y se enfocan en el proceso de aprendizaje, en vez de la calificación del producto final.

PALABRAS CLAVE: evaluación de la lengua, alfabetización en evaluación, disposición afectiva

1. INTRODUCTION

Chile recognizes English as an essential subject among Chilean schools as this foreign language represents a key resource for students' education. The main focus of the English language curriculum provided by the Ministry of Education (2014) is the development of communication skills. To achieve this, the National English Program for primary and secondary education considers different assessment practices and tools at the beginning, middle, and end of the teaching and learning process.

When talking about assessing students' performance, we are in front of one of the most critical responsibilities for classroom teachers. However, many teachers express to feel inadequately prepared for this responsibility, as they often manifest the need to update their assessment practices and techniques and request assistance when making assessment-related decisions.

Classroom assessment practices are based on teacher beliefs, training, and knowledge in educational assessment. Understanding teachers' classroom assessment practices is essential for informed educational decisions about students and their learning process; that is why, identifying teachers' knowledge and their affective disposition is imperative when it comes to assessment.

The purpose of this study is to identify in-service and pre-service teachers' literacy and affective disposition towards English language assessment. Two instruments were employed to gather the data, one for identifying assessment literacy and another one for affective dispositions towards language assessment.

2. LITERATURE REVIEW

2.1. *Language assessment*

Assessment provides crucial information that can be used for different purposes, including guidance for instructional classroom-based decision-making, school-level accountability for students' achievement, and the supervision and evaluation of educational programs (Coffey *et al.*, 2008). Purpura (2016) identifies language assessment as a broad term referring to a systematic procedure for eliciting test and non-test data (e.g., a teacher checklist of student performance) to make inferences or claims about certain language-related characteristics of an individual. The term assessment refers not only to formal tests such as the Test of English as a Foreign Language (TOEFL) or an end-of-chapter assessment, but also to other methods of obtaining data about a student's language knowledge, skills, and attitudes, such as by observing second language (L2) performance during pair work or by asking learners to report their understandings and uncertainties.

2.2. *Language assessment literacy*

For this study, we used Stiggins' (1991) definition of assessment literacy, in which he notes that educators who possess assessment literacy know what they assess, why they assess, how to assess, what the possible problems with assessment are, and how to prevent them from occurring. Therefore, language assessment literacy can be defined as what a language teacher knows about assessment. Moreover, language assessment literacy is a key factor at the moment of assessing students' performance. As stated by Mertler and Campbell (2005), this is one of the most critical skills of a classroom teacher, since it impacts nearly everything that teachers do. Some aspects of a teacher's job that are impacted by assessment can include: guiding decisions about large-group instruction; developing individualized instructional programs; determining the extent to which instructional objectives have been met; providing information for administrative decisions, such as promotion, retention, or graduation; and providing data for state or federal programs.

According to Fulcher (2012), teachers' assessment literacy involves the knowledge, skills, and abilities to design, develop, and assess any type of tests, including standardized and classroom-based tests. Furthermore, teachers' knowledge of testing processes, principles, and concepts are essential to guide assessment within the ethics and codes of practice. Additionally, assessment literacy is described by Fulcher (2012) as the faculty to study the practices of testing on society, institutions, and individuals at a historical, social, political, and philosophical level. Therefore, developing teachers' assessment literacy helps the assessment and testing processes to be more precise and can improve students' learning.

Inbar-Lourie (2008) states that teachers need to be critical at the time of assessment. They need to understand what skill is being assessed, why and how it is being assessed and they should be able to understand the results obtained and their implications. For language teachers to be assessment literate, they must possess assessment literacy skills

combined with language-specific competencies. This can be achieved only through the appropriate teacher training in assessment (Jeong, 2013).

Taylor (2009) argues that an appropriate level of language assessment literacy needs to be nurtured, not just among researchers, applied linguists or language teachers involved in delivering language education, but much more broadly in the public domain so a better understanding of the function and values of assessment tools and their outcomes are recognized across society. The importance of identifying the degree of language teachers' assessment literacy is of great significance in teachers' performance since assessment literacy can enhance or limit the education process and student achievement (Mertler, 2003).

2.3. Standards for Teacher Competence in Educational Assessment of Students

The Assessment Literacy Inventory (*ALI*) (Mertler & Campbell, 2005) is based on the Standards for Teacher Competence in Educational Assessment of Students developed by the American Federation of Teachers, the National Council on Measurement in Education and the National Education Association in 1990, out of concern that the potential educational benefits of student assessments are fully realized. These standards are intended to guide the pre-service and in-service preparation of educators, the accreditation of preparation programs, and the future certification of all educators.

The standards are defined by the American Federation of Teachers, and the National Council on Measurement in Education (1990) as follow:

1. *Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.*
2. *Teachers should be skilled in developing assessment methods appropriate for instructional decisions.*
3. *Teachers should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods.*
4. *Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.*
5. *Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.*
6. *Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.*
7. *Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.*

In Chile, the Ministry of Education in 2014 created the *Estándares Orientadores para Carreras de Pedagogía en Inglés* to guide EFL teacher preparation programs in the training processes of English teachers. When comparing Standards 5, 7, and 8 of the standards set by the Chilean Ministry of Education and the Standards to the

Teacher Competence in Educational Assessment of Students (AFT, NCME, and NEA, 1990), we can establish a relationship between both standards. Standard 5 details that teachers should know how to recognize, design, and apply different assessment methods to obtain information about the students' learning process and to improve the teaching and learning process. Standard 7 is related to the theories of learning a foreign language, which allow teachers to select and apply effective methodological approaches, besides the adequate strategies for the teaching and learning process, and Standard 8 highlights that teachers should be skilled at designing, choosing, and adapting physical and virtual resources for English teaching and learning (MINEDUC, 2014).

2.4. Affective dimension towards language assessment

Kahveci & Orgill (2015) claim that the first step to measure affect is to define the term. Forgas (2001) states that from a research perspective, the affective domain includes a host of psychological constructs and is often described as attitudes, values, beliefs, opinions, interests, and motivation. Kahveci & Orgill (2015) explain that there are several ways to measure the affective domain. Their study was carried out through a questionnaire, which was classified as a self-report method. The use of this self-report method is based on the assumption that the individuals are the ones who know their affective state.

Rahayu & Rahayu (2019) express that when assessing writing, for example, experienced teachers assess based on what they believe about assessment. This implies that the affective dimension is an important factor of the assessment process, as it has an impact on the assessment's quality. Rahayu & Rahayu (2019) pose that teachers' affective factors including attitude, efficacy, and motivation, contribute a lot to how they assess their students. In her paper about the 2014 Teachers College, Columbia University Round Table in the Second Language conference, Stabler-Havener (2014) emphasizes the fact that, even though there is literature on the affective dimension of teaching and learning, little research has been done on how affect influences assessment.

2.5. Empirical studies on language assessment

The *Assessment Literacy Inventory* measures teacher competency levels in the educational assessment dimensions. It was developed by the American Federation of Teachers, the National Council on Measurement in Education and the National Education Association in 1990. A total of 555 teachers and 286 schools across the USA participated in the study, including individuals from large and small school districts in rural, suburban, and urban areas. Results showed that, in general, teachers performed best in the competency area of administering, scoring, and interpreting test results but demonstrated the weakest results concerning the communication of test results. Teachers who had some measurement training scored significantly higher than those who did not, and teachers expressing comfort in interpreting standardized tests scored significantly higher on the inventory than teachers who expressed discomfort.

Mertler (2003) measured and compared pre-service and in-service teachers' assessment literacy. Both groups were surveyed using the *Classroom Assessment Literacy Inventory (CALI)*, which was designed to parallel the standards for teacher competence in the educational assessment of students (AFT, NCME & NEA, 1990). In-service teachers performed highest on Standard 3, *Administering, scoring, and interpreting the results of assessment*, and lowest on Standard 5, *Developing valid grading procedures*. Pre-service teachers performed highest on Standard 1, *Choosing appropriate assessment methods*, and lowest on Standard 5, *Developing valid grading procedures*. Comparisons between the two groups revealed significant differences in five of the seven competency areas, as well as in the total scores. In all cases where significant differences were found, in-service teachers scored higher than their pre-service counterparts.

Yamtim & Wongwanich (2013) carried out a study aimed at investigating the levels of classroom assessment literacy of primary school teachers and suggesting a developmental approach for improving the classroom assessment literacy of primary school teachers. The study sample consisted of 19 primary school teachers at a Thailand school; who completed the *CALI* (Mertler, 2003) and 8 teachers who participated in a focus group discussion. Data were analyzed using descriptive statistics and content analysis. The study findings revealed that most of the teachers had a low level of classroom assessment literacy. The standard that scored the highest mean score was Standard 1, *Choosing an assessment method*, whereas the standard that scored the lowest mean score was Standard 5, *Using assessment to determine levels of learning outcomes*. Such findings reflect the need to improve the classroom assessment literacy of primary school teachers at Thai schools.

Muhammad & Bardakçı (2019) also applied the *CALI* in their study aimed at understanding the assessment literacy levels of Iraqi English language teachers. Data were collected from 101 teachers working at secondary and preparatory schools in Iraq. On average, Iraqi English language teachers were able to answer only 15 correct responses out of 35 questions. Teachers scored lowest on questions regarding Standard 7, *Recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information*, and highest on questions regarding Standard 4, *Using assessment results when making educational decisions*. Although 77% of teachers believe they had been adequately trained for assessment, the results revealed that the teachers' level of assessment literacy was low.

3. METHODOLOGY

3.1. Type of study

The purpose of this non-experimental, quantitative, and cross-sectional research study is to identify in-service and pre-service teachers' literacy and affective disposition towards language assessment.

3.2. Research aims

This research was designed with two main purposes:

- To compare in-service and pre-service teachers' literacy towards language assessment.
- To compare in-service and pre-service teachers' affective disposition towards language assessment.

3.3. Participants

A total of 210 participants contributed to this study, including 78 in-service teachers and 132 pre-service teachers. All 78 in-service teachers answered the *Assessment Literacy Inventory (ALI)*, and 48 answered the *Assessment Disposition Scale (ADS)*. In the case of pre-service teachers, 102 answered the *ALI* and 112 answered the *ADS*.

In-service teachers ranged in age from 20 to 60 years, while the age range among pre-service teachers was from 20 to 42 years. Five percent of in-service teachers and two percent of pre-service teachers did not answer any of the instruments.

3.4. Instruments

The inventory used in this research to measure in-service and pre-service teachers' assessment literacy is called the *Assessment Literacy Inventory* which was adapted from the *Assessment Literacy Inventory* developed by Mertler and Campbell in 2005 and was adjusted from the *Teacher Assessment Literacy Questionnaire* by Plake and Impara in 1993. The inventory was developed based on the *Standards for teacher competence in educational assessment of students*. It consists of 35 questions adapted to the Chilean context, divided into five scenarios. Each scenario contains seven items based on one of the standards for teacher competence in educational assessment of students. The alignment of the standards per item is the following:

- Standard 1: Items 1, 8, 15, 22, 29
- Standard 2: items 2, 9, 16, 23, 30
- Standard 3: items 3, 10, 17, 24, 31
- Standard 4: items 4, 11, 18, 25, 32
- Standard 5: items 5, 12, 19, 26, 33
- Standard 6: items 6, 13, 20, 27, 32
- Standard 7: items 7, 14, 21, 28, 35

The *Affective Disposition Scale* was used to measure in-service and pre-service teachers' perceptions towards English assessment in the Chilean context. The *ADS* is

a type of semantic differential scale created to measure the connotative meaning of an issue on any topic. Rosenberg and Navarro (2018) define the semantic differential scale as a technique for measuring people's attitudes towards nearly anything. This type of scale uses a standardized set of bipolar adjectives on which participants rate an issue or object.

The *ADS* consists of 33 items regarding English language assessment. Using the scale illustrated in Figure 1, participants have to place an "x" on the icon which more accurately represents their feeling toward each item.

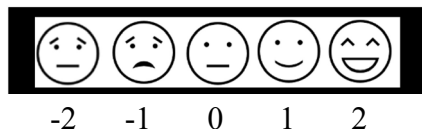


Figure 1. *Affective disposition scale rating*

Each icon is labeled with numbers going from -2 to 2 and, depending on the items, the scale is graded by a set of bipolar adjectives that represent a continuum that moves from a negative to a positive perception: *False to True; Indifferent to Shared; Insignificant to Important; Little to Much; Inefficient to Efficient; Traditional to Innovative; Complex to Simple; Inadequate to Suitable and Unproductive to Productive*.

3.5. Procedure

We distributed two instruments plus a consent form at the same time to in-service and pre-service teachers in person or by e-mail. They had to return the forms in a physical or digital format.

Each instrument asked for participants' personal information and had an introductory paragraph which explained to in-service and pre-service teachers the purpose of the instrument, what was expected of them and the benefit of their answers for the study. A total of 78 in-service teachers and 132 pre-service teachers completed and returned the surveys.

3.6. Data analysis

The analysis of in-service and pre-service teachers' responses to each instrument included comparisons of means scores and standard deviations between groups. Additionally, we applied the t-test to identify statistical differences among in-service and pre-service teachers' responses.

4. RESULTS

The analysis will be done per instrument. We will group the three highest and the three lowest mean scores of the results obtained in the *ALI* and *ADS*, from both pre-service and in-service teachers. Additionally, we will analyze through semantic differential scale graphs, the contrast of in-service and pre-service teachers' affective reactions towards each item's answers.

4.1. The Assessment Literacy Inventory

4.1.1. Scores from in-service teachers

The results obtained per standard can be observed in Figure 2. The findings show that, on average, in-service teachers responded 42% of the inventory correctly.

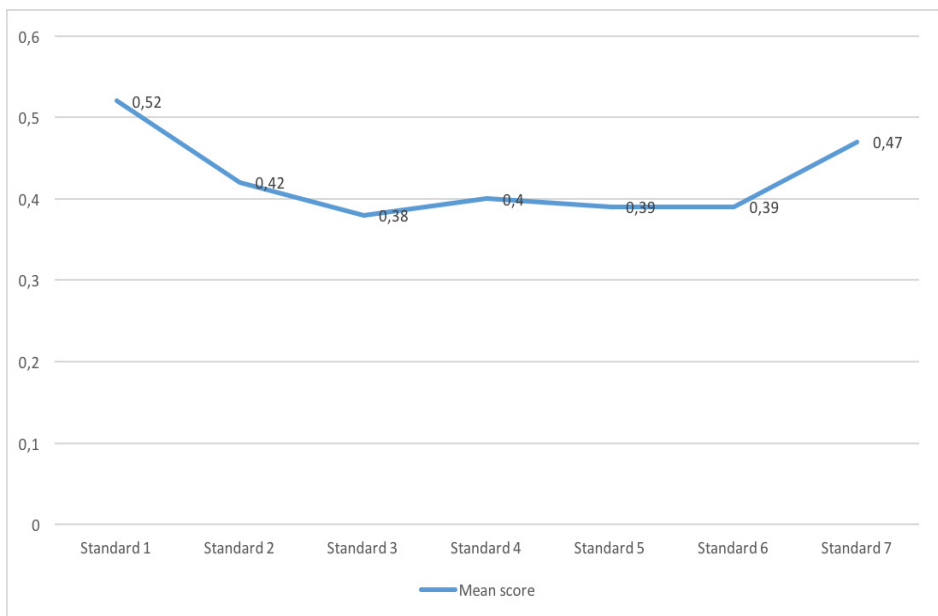


Figure 2. In-service teachers' results per standard

In-service teachers' three highest standards' mean scores were on Standard 1 (52%), Standard 2 (42%), and Standard 7 (47%) as explained below:

- "Standard 1: Teachers should be skilled in choosing assessment methods appropriate for instructional decisions" (MS: 0.52 / SD: 0.21).
- "Standard 2: Teachers should be skilled in developing assessment methods appropriate for instructional decisions" (MS: 0.42 / SD: 0.21).

- “Standard 7: Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information” (MS: 0.47 / SD: 0.18).

One way to interpret these results is that, as in-service teachers are supposed to have experience with assessments, especially in choosing and developing assessment methods appropriate for instructional decisions, they are therefore more aware of assessment methods and can differentiate between right and wrong when it comes to assessment. The standard deviations of the three standards are low, suggesting that teachers agree on the same response. However, for Standard 7 and 2 the results show that teachers in fact agree on the wrong response, suggesting their understanding of these standards is weak.

Scores for in-service teachers were the lowest for Standard 3 (38%), Standard 5 (39%) and Standard 6 (39%), as described below:

- “Standard 3: The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods” (MS: 0.33/SD: 0.23).
- “Standard 5: Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments” (MS: 0.39/SD: 0.21).
- “Standard 6: Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators” (MS: 0.39/SD: 0.23).

These findings suggest that in-service teachers are more concerned with grading than analyzing the implications of the data obtained from the assessment methods. As a result, in-service teachers are not being able to communicate effectively the outcomes to students, parents, lay audiences or other educators. Additionally, in-service teachers are not skilled in developing valid pupil grading procedures and assessments. Although there is not a large range in answers, since the standard deviations are low, the overall percentage of achievement is low because the answers are not the expected ones.

4.1.2. Scores from pre-service teachers

Figure 3 shows the results of pre-service teachers per standard. The findings show that pre-service teachers answered 40% of the inventory correctly.

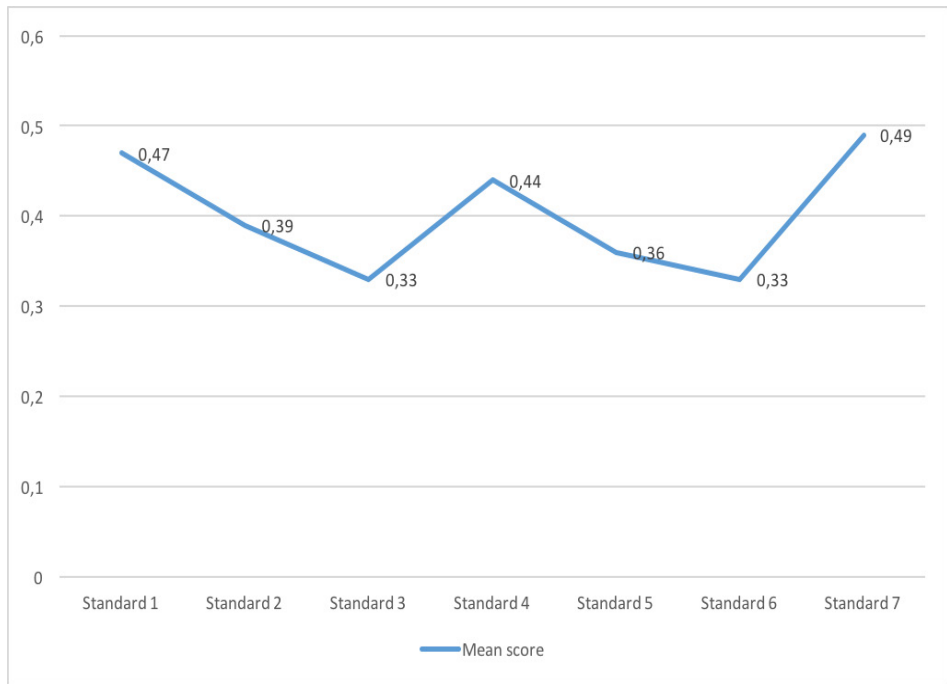


Figure 3. *Pre-service teachers results per standard*

The three highest mean scores were for Standard 1 (47%), Standard 4 (42%), and Standard 7 (49%) as described below:

- “Standard 1: Teachers should be skilled in choosing assessment methods appropriate for instructional decisions” (MS: 0.42/SD: 0.22).
- “Standard 4: Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement” (MS: 0.44/SD: 0.22).
- “Standard 7: Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information” (MS: 0.49/SD: 0.22).

These results suggest that although pre-service teachers are not inserted in the educational system yet, they are able to discern better any inappropriate assessment methods as they are still familiar with theories of assessment because they are still revising these topics at university. Pre-service teachers scored very low on questions related to some of the standards, implying that they still do not have the necessary practical knowledge about assessment. The associated standard deviations are also small, meaning pre-service teachers tended to give similar answers. The lowest mean

scores of pre-service teachers were for Standard 3 (33%), Standard 5 (36%), and Standard 6 (33%) as described below:

- “Standard 3: The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods” (MS: 0.33/SD: 0.23).
- “Standard 5: Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments” (MS: 0.36/SD: 0.22).
- “Standard 6: Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators” (MS: 0.33/SD: 0.20).

These low results may be explained by the lack of pre-service teachers’ experience in the development, administration, and analysis of assessment procedures. The standard deviation is low for each standard, meaning that most of the participants agreed on the same answer.

In brief, the resulting data show that in-service and pre-service teachers achieved less than 50% of the inventory, which means that their knowledge towards English assessment is deficient. This is supported by the t-student test which resulted in $t=1,16$, $p>.05$; therefore, there is no substantial difference between in-service and pre-service teachers’ knowledge of assessment, since both groups show similarly weak literacy towards language assessment.

4.2. The Affective Disposition Scale

4.2.1. Scores from in-service teachers

Figure 4 below illustrates the mean scores per item obtained by in-service teachers of English on the *ADS*. This analysis demonstrates that this group shows a favorable affective disposition towards language assessment as, on average, teachers scored above 3.0 on 64% of the items.

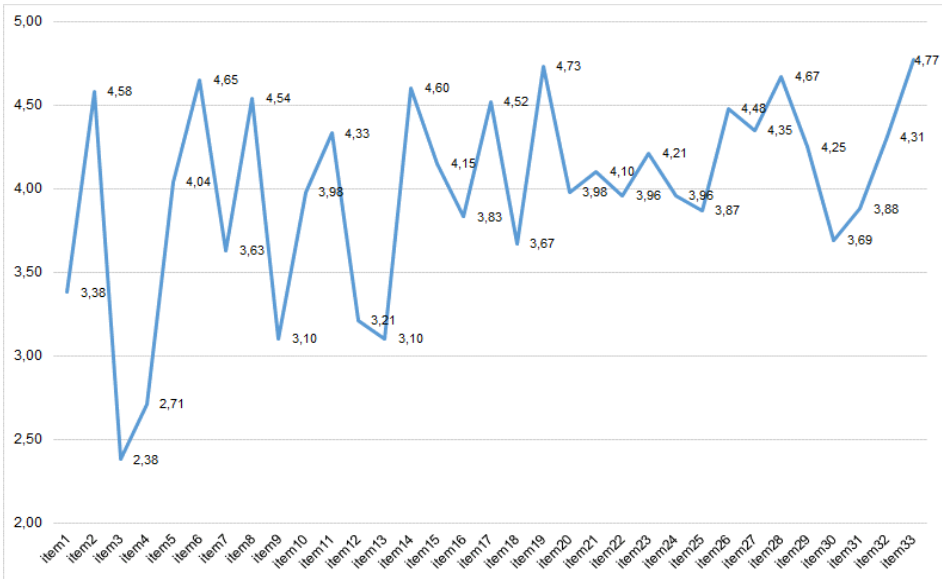


Figure 4. *In-service teachers' mean scores in the ADS towards English assessment*

The three items with the highest mean scores for in-service teachers included item 19, item 33, and item 28 as detailed below:

- “19 (I share the perception that) English assessment should allow students to show, create or produce something meaningful” (MS:4.73/SD:0.61).
- “28 (I share the perception that) English assessment should not be used as a way to punish students who do not fulfill their academic responsibilities (not bringing homework/materials, not paying attention during lessons, among others)” (MS: 4.67/SD: 0.91).
- “33 (I share the perception that) English assessment should give room for the correction of mistakes” (MS: 4.77/SD:0.47).

These results indicate that most in-service teachers shared the perception that English assessment is a process that should allow students to create something meaningful for them and should permit teachers to give students feedback and correct mistakes rather than assign grades based on students' behavior or the final product they submit. The standard deviations of these items are low, reinforcing the idea that a fair number of in-service teachers share these perceptions.

Items 3, 4 and 9 are the ones with the lowest mean scores as described below:

- “3 (I feel indifferent towards the perception that) English assessment formats should be multiple choice with an assigned time for answering” (MS: 2.38/SD: 0.98).
- “4 (I feel indifferent towards the perception that) English assessment formats should be free and without time restrictions” (MS: 2.71/SD: 1.17).
- “9 (I do not agree with the perception that) English assessment should not focus on students giving correct answers” (MS: 3.10/SD: 1.15).

It can be interpreted from these results that a small group of in-service teachers perceived language assessment as simply the final stage of the teaching and learning process, with a focus only on how correct the final product is. Although item 9 is within the three items with the lowest mean scores, it has a mean score above 3.0, which indicates that in-service teachers showed a favorable perception towards this item. The standard deviations for these items are low, implying that an important number of in-service teachers did not share the same perceptions described on items 3 and 4, and that they perceived English assessment should not only focus on students giving correct answers.

4.2.2. Scores from pre-service teachers

Figure 5 represents the results of pre-service teachers. This group shows results that are similar to those of in-service teachers with a mean score of 3.0 or higher on 70% of the items, indicating a favorable affective disposition towards language assessment.

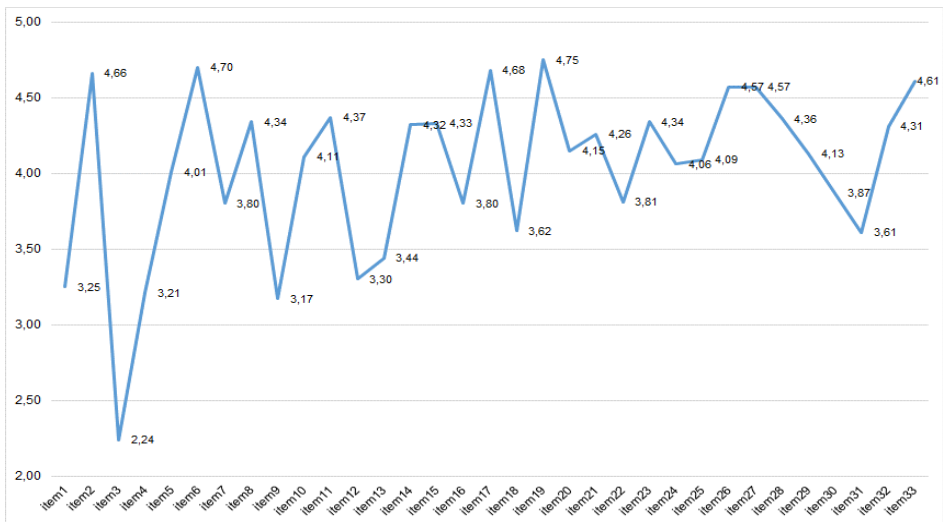


Figure 5. Pre-service teachers' mean scores in the ADS towards English assessment

The three items with the highest mean scores included item 2, item 17 and item 19 as explained below:

- “2 (I perceive as true that) English assessment should be a continuous, long-term process” (MS: 4.66/SD: 0.57).
- “17 (I perceive as important that) English assessment should favor students’ self-motivation to learn the language” (MS: 4.68/SD: 0.58).
- “19 (I share the perception that) English assessment should allow students to show, create or produce something meaningful” (MS: 4.75/SD: 0.48).

These outcomes suggest that pre-service teachers shared the perception that language assessment should focus on the creation of something meaningful instead of using the language without a clear purpose. Additionally, pre-service teachers found students’ self-motivation as a paramount asset at the moment of learning the language and believed that language assessment should be regarded as a continuous, long-term process. The standard deviations for these items are low, showing that pre-service teachers hold a favorable affective disposition towards language assessment.

The lowest mean scores were associated to items 3, 4 and 9 as described below.

- “3 (I feel indifferent towards the perception that) English assessment formats should be multiple choice with an assigned time for answering” (MS: 2.24/SD: 0.98).
- “4 (I share the perception that) English assessment formats should be free and without time restrictions” (MS: 3.21/SD: 1.08).
- “9 (I do not agree with the perception that) English assessment should not focus on students giving correct answers” (MS: 3.10/SD: 1.19).

These scores indicate that the only item in which pre-service teachers expressed an unfavorable perception towards language assessment was item 3, meaning that a reduced number of participants regarded language assessment as standardized testing. The mean scores above 3.0 on items 4 and 9 suggest that pre-service teachers understood language assessment as a guided process focused on communication rather than a final accurate product. Given the small standard deviations for each item, it appears most pre-service teachers agree on the answers for the items.

While most of the responses between in-service and pre-service teachers varied, teachers did share a similar idea of what English assessment should and should not be like. Both groups perceived English assessment as an opportunity for students to create something that has a meaning for them, where self-motivation and guidance are paramount characteristics of the long-term process of language acquisition. Participants rejected the idea that English assessment should only focus on students’ final products and their accuracy. The only difference between in-service and pre-service teachers was on the affective disposition demonstrated on item 4. The results showed that in-service teachers were indifferent about whether or not English assessments should

be free and without time restrictions, whereas pre-service teachers shared a positive perception towards this level of flexibility. However, these differences do not represent a big discrepancy among both groups when talking about their affective disposition towards English language assessment.

Figure 6 below shows in-service and pre-service teachers' mean scores and affective reactions towards assessment for the first 11 items of the *ADS*.

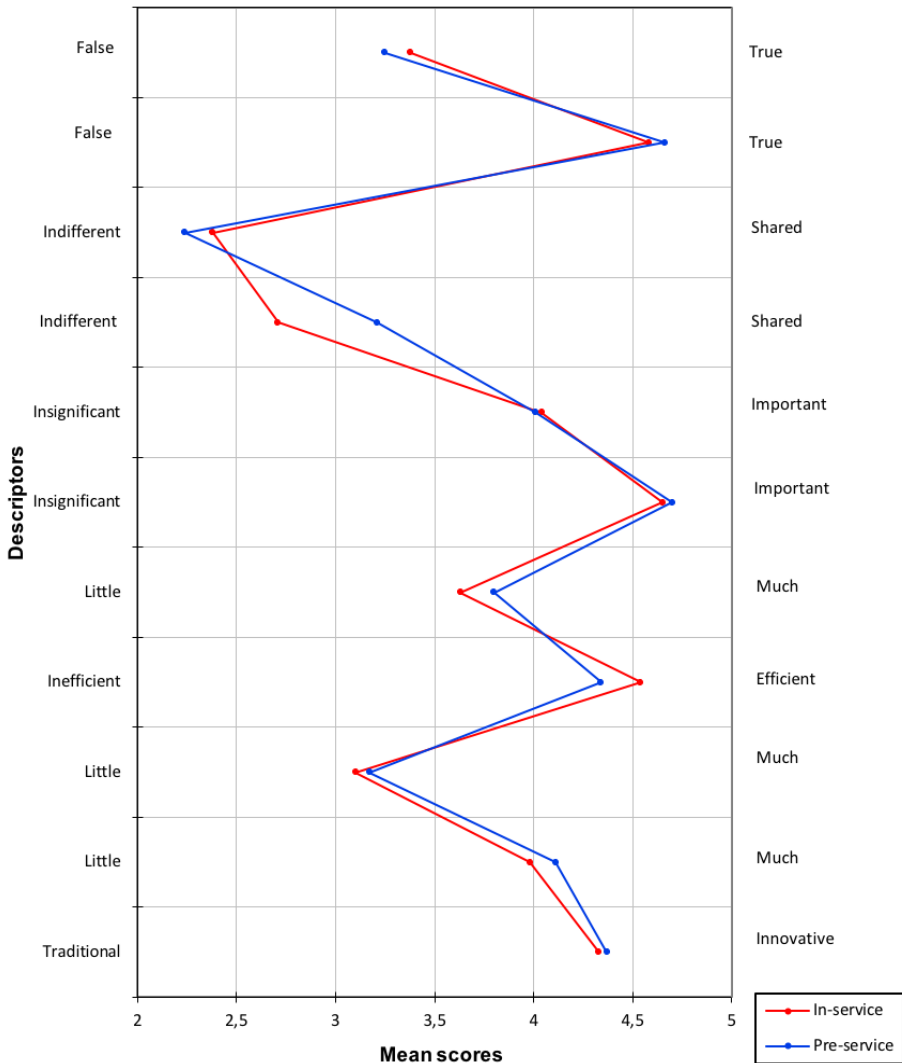


Figure 6. *In-service and pre-service teachers' mean scores and affective reactions for items 1-11 in the ADS towards English assessment*

Looking across all items, it can be observed that both groups tended to score in a positive direction on all items except items 3 and 4. This analysis reflects the positive affective disposition of in-service and pre-service teachers towards the assessment views represented on these items. Figure 7 below shows findings on items 12-22.

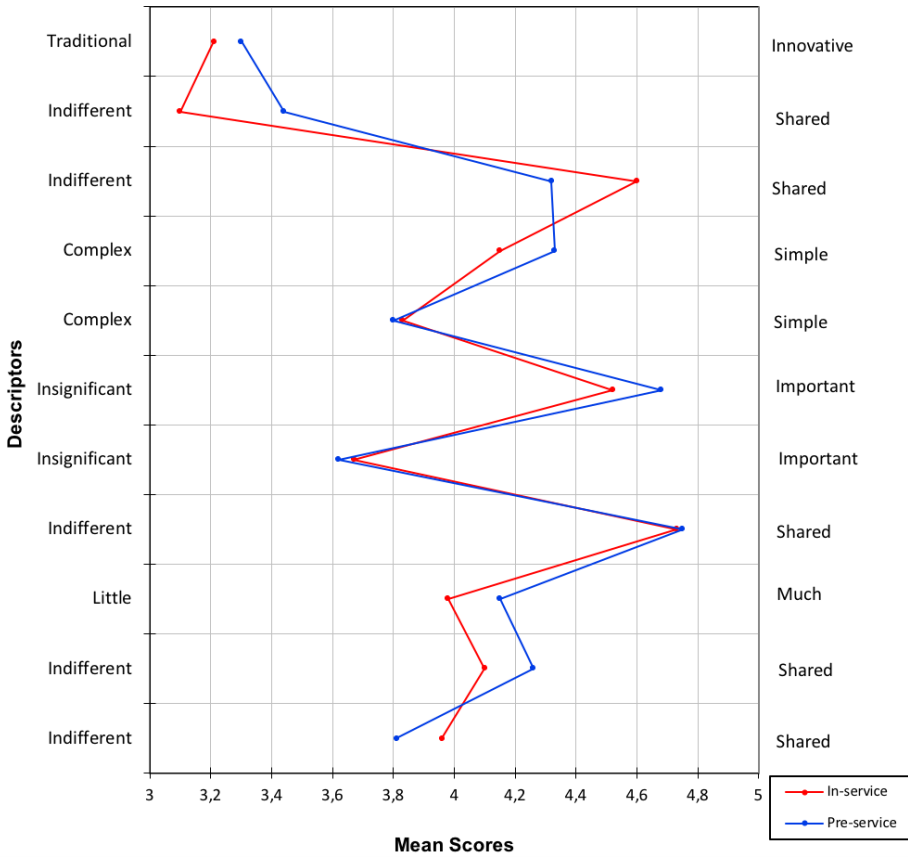


Figure 7. In-service and pre-service teachers' mean scores and affective reactions for items 12-22 in the ADS towards English assessment

In Figure 7 above it is shown that all mean scores are above 3.0. Even though items 19 and 20 are shown to be closer to the negative meaning adjectives, in general, the scores imply that the affective reaction towards English assessment is positive. This piece of data suggests that both groups tend to incline towards the positive meaning adjectives. Figure 8 below shows findings on items 23-33.

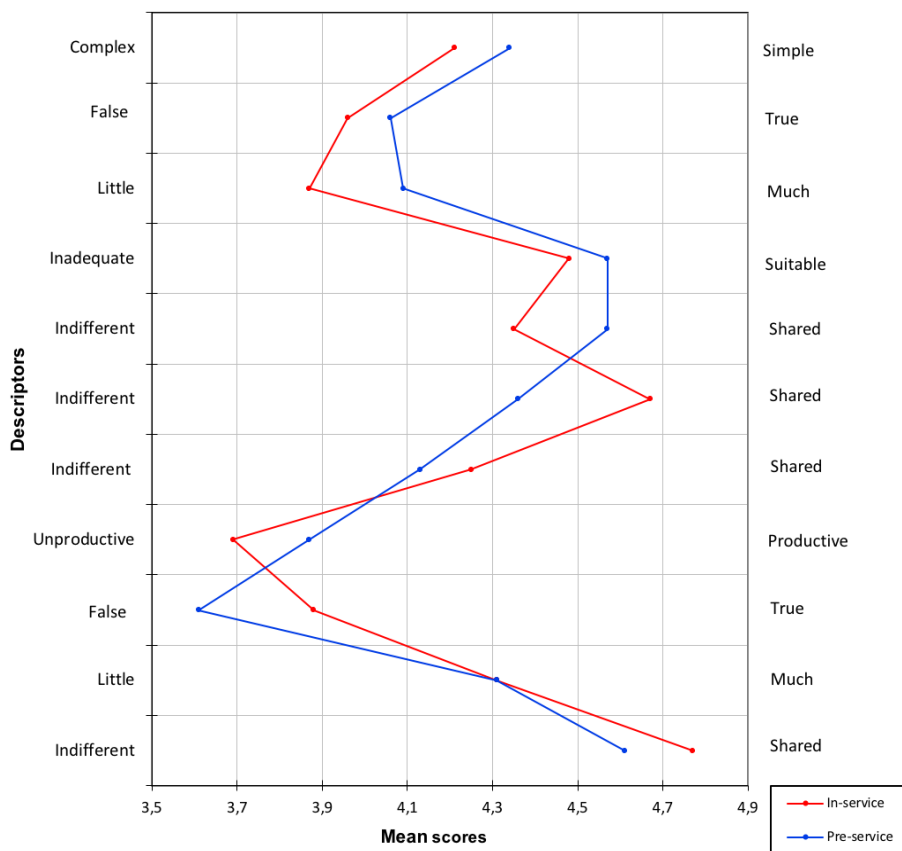


Figure 8. *In-service and pre-service teachers' mean scores and affective reactions for items 23-33 in the ADS towards English assessment.*

Figure 8 illustrates mean scores and affective reactions for items 23 to 33. All mean scores are above 3.5. This indicates that even though in-service and pre-service teachers seem inclined to choose negative meaning adjectives in some items, it does not mean that the affective reaction towards English assessment is negative. Furthermore, an important number of responses are shown to lean towards the positive meaning adjectives.

Overall, the t-tests applied to both in-service and pre-service teachers' results of the ADS showed that $t = -1.06, p > 0.05$, meaning that there are no statistically significant differences found between both groups. This restates what was described before; the affective reaction towards English assessment for in-service and pre-service teachers is similar and positive.

5. DISCUSSION

5.1. About assessment literacy

The results obtained by the Assessment Literacy Inventory for Chilean in-service and pre-service teachers are worrying as both groups answered less than 50% of the inventory correctly, suggesting that their knowledge of language assessment is deficient.

The findings show that in-service teachers achieved 42% of the inventory while pre-service teachers achieved 40% of it. The highest overall performances for pre-service and in-service teachers were on Standard 1, Standard 4, and Standard 7; whereas the lowest mean scores correspond to Standard 3, Standard 5, and Standard 6. These results are similar to those found in a study conducted by Perry (2013), looking at the level of assessment literacy of Montana high school teachers. Perry's (2013) study also used the classroom assessment literacy inventory and found that, on average, teachers answered 63% of the questions correctly, with participants scoring the highest on Standard 4 and the lowest on Standard 3, Standard 6, and Standard 7. While participants in Perry's (2013) study and those in this current study scored the highest on the same standards, the two groups differed in terms of overall performance. Perry's analysis showed that Montana teachers' knowledge towards assessment is higher than that of the Chilean in-service and pre-service teachers participating in this study. Additionally, Montana teachers tended to score the lowest on Standard 7, whereas Chilean in-service and pre-service teachers tended to score the highest on Standard 7.

Mertler (2003) surveyed in-service and pre-service teachers to measure their assessment literacy using the Classroom Assessment Literacy Inventory (*CALI*). The results showed that in-service teachers achieved 62% of the inventory correctly and their highest and lowest mean scores were on Standard 3 and Standard 5, respectively. Moreover, pre-service teachers' highest mean performance was on Standard 1 and the lowest was on Standard 5. In comparison to Mertler's (2003) results, the present study shows that the mean scores obtained by the Chilean teachers tended to be quite low for Standard 3. Overall, in-service teachers' in this study performed the best on Standard 1 and demonstrated the lowest performance on Standard 5, while pre-service teachers' highest mean score was on Standard 7 and their lowest mean score was on Standard 5. These comparisons suggest that Montana teachers demonstrated higher literacy on administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods than in-service and pre-service teachers participating in this study. Overall, Mertler's (2003) results show that in-service and pre-service teachers have high literacy towards assessment, whereas Chilean in-service and pre-service teachers show low literacy for language assessment.

A different study done by Muhammad & Bardakçı (2019) to investigate the assessment literacy of Iraqi English language teachers through the Classroom Assessment Literacy Inventory showed that Iraqi EFL teachers' assessment literacy level was much less than satisfactory, as teachers answered 42% of the inventory correctly. This result was the same as the one obtained by Chilean in-service teachers and similar to the Chilean pre-service teachers' results. Looking at the performances

by standards, further similarities between these two studies can be identified. Iraqi teachers' highest mean score was for Standard 4 while Chilean in-service and pre-service teachers' in this study also tended to perform well on Standard 4. This means that in both studies, in-service and pre-service teachers were skilled at using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement. However, Iraqi teachers tended to score the lowest on Standard 7, while Chilean teachers tended to score the highest on this standard. This suggests that Chilean in-service and pre-service teachers were more skillful than Iraqi teachers at recognizing unethical, illegal, and, otherwise, inappropriate assessment methods and uses of assessment information.

Perry's (2013) and Mertler's (2003) studies show that teachers' knowledge towards language assessment was higher in the US-context than for in-service and pre-service teachers in Chile. On the other hand, the study conducted by Muhammad & Bardakççi (2019) suggests a comparative level of knowledge with Iraqi teachers' knowledge towards language assessment, which tended to be as low as Chilean in-service and pre-service teachers' literacy towards language assessment.

The strongest and weakest aspects of teachers' assessment literacy show little difference across the literature reviewed on language assessment. The results may be due to the differences in assessment courses in teacher education programs, curriculum differences, and education policies.

5.2. About the affective disposition towards language assessment

According to the present study, the affective reaction of in-service and pre-service English teachers was similar and positive between both groups. The general results of in-service and pre-service teachers indicate that they perceive English assessment as a guided instance that self-motivates students to create something meaningful.

This study finds that in-service and pre-service teachers focus on the continuity of the process at the moment of assessing their learners' language competence rather than seeing the final product as of higher importance. Teachers signaled feedback as a paramount aspect of the assessment process. Mahfoodh & Pandian (2011) conducted a study to investigate English as a Foreign Language (EFL) students' affective reactions and perceptions of their teachers' written feedback. The findings of Mahfoodh & Pandian's (2011) study show that EFL students consider their teachers' feedback useful, suggesting that teachers have some level of authoritative power. Students expected teachers to focus on all aspects of their written texts but place more emphasis on grammatical corrections. The current study showed that pre and in-service teachers' positive reactions towards item 8 (*Feedback should be individual and it should continually inform students' about their assessment performance*) aligns with Mahfoodh & Pandian's findings (2011) in the sense that teachers perceive themselves as the primary guide for their students at the moment of assessing.

Rahayu & Rahayu (2019) carried out a study based on several journal articles about writing assessment and teachers' affective dimension towards it. The authors concluded that, when administering assessments, teachers need to care about their students' affective dimension and their own, so students can value what they learn, and

teachers can pay attention to their students' learning. This conclusion relates to item 32 of the *ADS*, which claims that English assessment is an information source for teachers. Chilean in-service and pre-service teachers had a positive perception towards this item, which suggests that these teachers perceived assessment as a source of information for the improvement of the learning-teaching process for both themselves and their students. Stabler-Havener (2014) stated that a sympathetic interlocutor affects test-takers' performances on examinations. She concluded that when a test-taker cannot answer a question because of its level of difficulty, a sympathetic test administrator (in contrast to an unsympathetic one) asks another, often easier question, in order to elicit a speech sample that can be rated, as opposed to not assigning a score to that test question. In the current study, Chilean in-service and pre-service teachers shared a similar perception on item 9, *English assessment should not focus on students giving correct answers*. In other words, teachers in this study would find a way to help a student reach the expected answer if that student was unable to answer a question correctly during an assessment. As explained before, that might be through the election of a new simpler query.

Moreover, Chilean teachers from both groups tended to demonstrate a positive affective perception on item 31, *English assessment is not an independent and different process from the teaching-learning process*. This finding aligns with what was stated in the previous paragraphs, regarding the fact that teachers tended to perceive assessment as a complement of the teaching and learning process.

6. CONCLUSIONS

The first objective of this study was to compare in-service and pre-service teachers' literacy towards language assessment. As mentioned earlier, it was expected that in-service would demonstrate a higher level of literacy than pre-service teachers as they have been assessing a foreign language longer. Nonetheless, findings revealed that in-service and pre-service teachers' knowledge towards language assessment was weak, with neither group answering more than forty-two percent of the assessment literacy scale correctly. This finding is critical since it suggests that in-service teachers are working in the educational system assessing language without possessing the necessary assessment knowledge. Similarly, pre-service teachers are starting their professional careers in the educational system with an insufficient level of language assessment literacy.

The second objective of this study was to compare in-service and pre-service teachers' affective disposition towards language assessment. The results of this instrument were expected to be similar between both groups since pre-service teachers may develop certain attitudes and perceptions towards assessment through their practicum experiences rather than during their university training. This prediction was proven correct, as in-service and pre-service teachers shared a positive reaction towards language assessment. Both groups agree on the fact that assessment should be complementary to the teaching and learning process and that the focus should not

be on students giving a correct answer but instead on their ability to communicate something meaningful in the target language.

Even though Chilean in-service and pre-service teachers showed insufficient language assessment literacy, they still regarded language assessment positively and focused on the students' learning process rather than the grading of a final product. For this reason, participants perceived updating assessment practices as important and they tried to incorporate different assessment approaches into their teaching. Furthermore, they believed that the training provided by Chilean teacher preparation programs in language assessment is insufficient and should be improved.

Lastly, additional qualitative research in this area is highly encouraged in order to gather more detailed data on how teachers perceive language assessment. We were not able to explore every aspect of the topic in this study and, therefore, believe more research would provide interesting insights into the experiences of in-service and pre-service teachers. In future studies, we would also like to include a variable measuring teachers' years of experience as we consider that experience may have an important impact on teachers' knowledge and attitudes towards language assessment. Finally, we believe conducting a study at a national level could provide important insight as the research data would be more generalizable and could be applied to a wider population of Chilean in-service and pre-service teachers.

7. REFERENCES

- AMERICAN FEDERATION OF TEACHERS, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION & NATIONAL EDUCATION ASSOCIATION. 1990. *The Standards for Teacher Competence in the Educational Assessment of Students*. Retrieved November 9th, 2019 from <https://buros.org/standards-teacher-competence-educational-assessment-students>.
- COFFEY, J., DOUGLAS, R. & STEARNS, C. 2008. *Assessing Science Learning: Perspectives from Research and Practice*. National Science Teachers Association.
- FORGAS, J. P. 2001. *Feeling and thinking: The role of affect in social cognition*. Cambridge University Press.
- FULCHER, G. 2012. Assessment literacy for the language classroom. *Language Assessment Quarterly* 9(2): 113–132.
- INBAR-LOURIE, O. 2008. Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing* 25(3): 385–402. doi:10.1177/0265532208090158.
- JEONG, H. 2013. Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing* 30(3): 345–362. <https://doi.org/10.1177/0265532213480334>.
- KAHVECI, M. & ORGILL, M. (Eds.). 2015. *Affective dimensions in Chemistry education*. Dordrecht: Springer.
- MAHFOODH, O. & PANDIAN, A. 2011. A qualitative case study of EFL students' affective reactions to and perceptions of their teachers' written feedback. *English Language Teaching* 4(3): 14–25.
- MERTLER, C. A. 2003. Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference? *Paper presented at the annual meeting of the Mid-Western Educational Research Association*, Columbus, OH.

- MERTLER, C. A. & CAMPBELL, C. 2005. *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- MINISTERIO DE EDUCACIÓN. 2014. *Estándares Orientadores para Carreras de Pedagogía en Inglés*. Retrieved from: https://www.cpeip.cl/wp-content/uploads/2019/03/Estándares_Inglés.pdf
- MUHAMMAD, F. & BARDAKÇI, M. 2019. Iraqi EFL teachers' assessment literacy: perceptions and practices. *Arab World English Journal* 10(2): 420-430. <https://dx.doi.org/10.24093/awej/vol10no2.33>
- PERRY, M. 2013. *Teacher and principal assessment literacy*. Doctoral thesis. University of Montana. Retrieved from: <https://scholarworks.umt.edu/etd/1391>
- PLAKE, B. & IMPARA, J. 1993. Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher* 6(1): 21-27. Retrieved from: <https://eric.ed.gov/?id=EJ460188>
- PURPURA, J. E. 2016. Second and foreign language assessment. *The Modern Language Journal* 100: 190-208. doi:10.1111/modl.12308
- RAHAYU, E. & RAHAYU, E. 2019. Teacher's cognitive and affective versus teachers' writing assessment. *Premise: Journal of English Education and Applied Linguistics* 8(1): 102-116.
- ROSENBERG, B. & NAVARRO, M. 2018. Semantic differential scaling. In B. Frey (Ed.), *SAGE encyclopedia of educational research, measurement, and evaluation*: 1504-1507. Thousand Oaks: SAGE.
- STABLER-HAVENER, M. L. 2014. Learning-oriented assessment: The affective dimension. *Working Papers in TESOL & Applied Linguistics* 14(2): 53-55.
- STIGGINS, R. J. 1991. Assessment literacy. *The Phi Delta Kappan* 72(7): 534-539. Retrieved from: <https://eric.ed.gov/?id=EJ422814>
- TAYLOR, L. 2009. Developing assessment literacy. *Annual Review of Applied Linguistics* 29: 21-36.
- YAMTIM, V. & WONGWANICH, S. 2013. A Study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences* 116: 2998-3004. Doi.org/10.1016/j.sbspro.2014.01.696.