How Do In-Service EFL Teachers Assess Student Language Learning? Analysis of English Assessment Instruments Used in Chilean Secondary Schools

ABSTRACT

The study analyses the assessment instruments designed by Chilean English as a Foreign Language (EFL) in-service teachers. The participants were 110 secondary school teachers from the central, northern, and southern areas of Chile. The data collection technique was document analysis of the assessment instruments designed in their teaching practices. The results suggest that the most typical instrument is a language test, while speaking/writing performance evaluations with rubrics and rating scales are used to a lesser degree, and all having writing, reading, grammar, and vocabulary as their primary assessment foci. It is suggested that universities and administrators of educational institutions promote opportunities for professional updating regarding language assessment. Moreover, pre-service teacher education needs to emphasize and strengthen this didactic dimension.

Keywords: language assessment, TEFL, in-service teachers, teacher professional development, pre-service teacher education

Kako učitelji ATJ ocenjujejo učenje jezika? Analiza instrumentov za ocenjevanje znanja angleščine, ki se uporabljajo v čilskih srednjih šolah

IZVLEČEK


Ključne besede: ocenjevanje jezika, poučevanje angleščine kot tujega jezika, učitelji na delovnem mestu, strokovno izpopolnjevanje učiteljev, začetno izobraževanje učiteljev
1 Introduction

Chile is a country located in South America, whose official language is Spanish. At a national level, learning English as a Foreign Language (EFL) is prioritized to boost the country’s competitiveness and accelerate its cultural and commercial integration with the rest of the globalized world (Government of Chile 2014). Moreover, the Chilean educational system considers EFL a mandatory school subject at primary education levels (from the fifth to the eighth grades, specifically) and during the four years of secondary education (Ministry of Education of Chile 2015). Based on the suggestions of the Chilean curriculum, students are expected to achieve a B1 level of English (Ministry of Education of Chile 2019). This means a lower-intermediate level according to the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001).

Nowadays, the practices of English language teaching and learning should promote the development of receptive skills (listening, reading) and productive skills (speaking, writing), following a communicative approach (Celce-Murcia 2014; Duff 2014; Kumaravadivelu 2006; Larsen-Freeman and Anderson 2011; Nunan 2015; Richards and Rogers 2014). The findings of studies specify that classroom language assessment practices can facilitate student achievements based on these linguistic competences (Namaziandost et al. 2020; Safdari and Fathi 2020). This is because English teachers collect the evidence of student learning through appropriate assessment instruments, which allows them to identify the impact of their instructional practices and, consequently, make the right decisions for evidence-based education.

Despite the above, studies show that many EFL teachers prioritize summative over formative assessment practices, focusing mainly on assessing linguistic knowledge rather than language skills (Minda and Chaka 2023; Tagle et al. 2022; Tsagari 2016; Wubshet and Menuta 2015; Yan, Zhang, and Dixon 2022). This behaviour might suggest that undergraduate professional training does not provide prospective teachers with sufficient language assessment literacy (Giraldo and Murcia 2019; Kong, Molnár, and Xu 2022). A study conducted by Tsagari and Vogt (2017) confirms this perspective, suggesting that EFL teachers’ lack of assessment training results in their dependence on using published assessment materials or assuming the practices of others without questioning them. The researchers conclude that an insufficient theoretical basis for assessment practices can impede the implementation of assessment innovations.

In Chile, English teacher education programmes usually include a one-semester course on evaluating students from a general pedagogical perspective (British Council 2015). This university subject does not address the specific techniques and instruments for assessing foreign language skills. Moreover, it does not develop a reflective attitude towards the assessment process. Based on these facts, it has been demonstrated that many prospective teachers learn how to assess student performance during their professional internships rather than their pre-service teacher training (Earl 2012).

In 2017, Chilean third-year secondary education students took a mandatory national English language examination that evaluated their reading and listening skills. The results indicated that 68% of Chilean EFL learners had a beginner level of the foreign language (A1,
according to the CEFR), while 32% demonstrated basic to intermediate proficiency (A2 and B2) (Agencia de Calidad de la Educación 2017). The data showed that Chilean students struggle to develop their English language skills further. These results may thus shed light on how assessment processes are carried out and which aspects of foreign language teaching and learning are prioritized in Chile. This indicates that English language teaching and assessment practices may not adequately address the development of students’ communicative skills at the national level.

Based on the background mentioned above, this research aims to analyse the assessment instruments designed by Chilean in-service EFL teachers in the context of their instructional practices. Likewise, the present study seeks to answer the following specific research questions:

1) Which kind of language assessment instruments do Chilean in-service EFL teachers design?
2) What are the language dimensions included in the language assessment instruments designed by Chilean in-service EFL teachers?
3) What are the characteristics of the assessment instruments designed by Chilean in-service EFL teachers?

2 Assessment in English Language Teaching and Learning

Educational assessment involves the systematic documentation of evidence of student performance against learning objectives (Britton 2021; Solano-Flores 2016). This process includes data collection and interpretation by teachers. Within this, assessment is not an isolated didactic dimension because it operates within a system that links its procedures with different elements, such as the curriculum, academic content, and standards of linguistic competence, in addition to the pedagogical process (Gottlieb 2016).

Assessment can be diagnostic, formative, and summative (Marsh 2010; Ravela, Picaroni, and Loureiro 2019). The first type is conducted at the beginning of a didactic unit to identify students’ strengths and weaknesses to make decisions about lesson planning. Formative assessment is carried out during the pedagogical process to help students achieve their learning objectives. In contrast, summative assessment measures students’ learning at the end of a didactic unit, generally using a grade.

Teachers’ assessment practices can be associated with three approaches. First, it is possible to refer to the assessment of learning, which is summative in nature and is used to identify what students have learned as a result of a pedagogical intervention (Earl and Katz 2006). This type of assessment is usually implemented exclusively by the teacher to document learning, measuring and categorizing it to report this information to students, parents, and teachers (Chappuis and Stiggins 2020). In English language teaching and learning, these practices are linked to traditional assessment, encompassing instruments such as tests with multiple-choice, true-false, and gap-filling tasks (Harris 2013). Traditional assessment tends to focus on language accuracy, which means the correct use of linguistic structures or conceptual contents (Goh and Burns 2012). Research findings have posited that these evaluative practices seek, primarily, to prepare students to be successful test takers rather than to help them develop
communicative skills (Styron and Styron 2012; Wubshet and Menuta 2015). Moreover, other studies have reported that traditional assessment does not significantly impact the students' learning of English language skills (Forutan 2014; Goçtü 2012).

Another approach is assessment for learning, which is conducted for formative purposes. These assessment practices are employed to help students construct their learning throughout the instructional process (Butt 2010; Jones 2010). This involves providing continuous feedback and encouraging learners to self- and co-assess their performance. This type of assessment emphasizes communication or the production of clear ideas (fluency) over the correct use of linguistic structures (accuracy) (Türk 2022). Likewise, it understands learning as a social process so that interaction tasks and co-assessment/self-assessment procedures play a relevant role (Heritage 2022).

Alternatively, it is possible to refer to assessment as learning. This assessment approach corresponds to an active process of cognitive restructuring that occurs when learners interact with new ideas (Earl and Katz 2006). This type of assessment performance is related to metacognition, which implies knowledge of one's thinking processes.

2.1 English Assessment Instruments

Assessment instruments are used to collect information or evidence about students' learning (Chappuis and Stiggins 2020). Within the English language teaching and learning framework, instruments can be associated with traditional assessment, favouring the identification of linguistic knowledge, such as grammar, vocabulary, or pronunciation, and employing memorization (Forutan 2014; Shams and Tavakoli 2014). One of the most common instruments of this type of assessment is the written test, mainly employed in summative assessment practices as its purpose is to measure student language knowledge (Richard-Amato 2010).

It is also possible to refer to authentic assessment instruments, which focus on recognizing student learning concerning English language communicative competencies through formative assessment (Aliasin and Amianlu 2017; Brown and Abeywickrama 2018). These instruments relate to communicative performance assessment tasks performed in class, in which learners have an active role in listening, reading, speaking, and writing in English. These tasks are designed to help students achieve the learning objectives during the process, serving as evidence of achievement and allowing for feedback.

In authentic assessment or assessment for learning, checklists are used with instruments or tasks to gather information and judge language learners' performance (Gottlieb 2021). Checklists follow a yes/no format when evaluating whether students have met specific criteria by observing them (Katz 2014).

Teachers may use scales to supplement their assessment instruments when documenting students' language performance, especially in the case of productive skills (Fulcher and Davison 2007; Knoch 2009; Kuiken and Vedder 2017; Murray and Christison 2020). On the one hand, rating scales are used to establish the degree or frequency of students' displayed language functions, behaviours, skills, or strategies. These include mastery levels or bands
based on criteria. On the other hand, rubrics are criterion-referenced instruments utilized to evaluate learners’ work based on uniform descriptors, which provide complete details on a specific performance. Rubrics can be holistic and analytic (Tedick and Lyster 2019). The former provides an overall assessment regarding student work, while the latter divides student performance into multiple dimensions or criteria.

Some authors suggest that rating scales and rubrics have distinct differences (see, for example, Plakans and Gebril 2015; Murray and Christison 2020). A rating scale is designed to determine a score, being more concise in describing different performance levels or indicators. In contrast, a rubric is mainly used to provide feedback to students, and therefore their criteria and descriptors must be more detailed and concrete.

2.2 Tasks Associated with the Assessment of Communicative Skills and Linguistic Content in English

In the context of English language learning, language assessment tasks collect information on students’ achievement of learning objectives (Katz 2014). Regarding listening and reading, open-ended comprehension activities allow students to produce their own responses considering the general or specific ideas of an oral or written text (Oakhill, Cain, and Elbro 2015). It has also been proposed that the development of these skills can be assessed utilizing close-ended comprehension tasks, which allow students to respond only by selecting previously defined options. These tasks contain, for example, multiple-choice, true-false, or gap-filling items, short-answer questions, and matching, among others (Alderson 2000; Buck 2001).

In speaking assessment, open-ended and structured tasks are used. Luoma (2004) indicates that the former allow students to put the skill into practice in terms of language communicative functions, such as describing, narrating, giving instructions, and explaining, among others. The latter are not based on the learners’ creation, but on the mastery of certain language components. Some examples of these types of activities include sentence drilling, reading aloud, oral sentence gap-filling exercises, and the formulation of short answer questions, among others.

To assess the learning objectives associated with writing in English, Brown and Abeywickrama (2018) highlight imitative tasks, which focus on the spelling correction of letters, words, or sentences. Likewise, the authors identify intensive or controlled tasks that involve the correct application of grammar or vocabulary when writing. Open-ended tasks, based on summarizing and integrating sentences within paragraphs, are also suggested. Finally, the authors refer to extensive written production tasks, in which learners must be competent in the ability to produce longer texts, such as essays, reports, and theses, among others.

3 Research Methodology

This study employs quantitative and qualitative research. The research design includes a collective case study, whose general objective is to analyse the assessment instruments designed by Chilean in-service EFL teachers. This research aim encompasses the types of assessment instruments, their respective assessment foci, and the tasks associated with these instruments.
3.1 Participants

The participants in this research were 110 EFL teachers working in secondary schools located in Chile. These institutions are located in the central, northern, and southern areas of the country. All study subjects signed a letter of informed consent, which specified that their participation was voluntary and that, in addition, the information they provided would be analysed anonymously, and their identity would be protected during and after the process.

3.2 Data Collection Technique

The document analysis technique was employed. According to Rapley (2018), this procedure involves examining materials produced by research participants to recognize their vision, actions, or what happens in their settings. In the present study, document analysis was used to process 824 assessment materials designed by the informants regarding their face-to-face instructional practices. Each of the teachers provided, voluntarily, between one and 10 documents.

First, quantitative document analysis was used by reviewing each of the instruments and completing a data matrix framed by the types of instruments, their assessment focus/foci (language dimension), and the tasks they are associated with. Then, the frequency of these components was calculated.

The ATLAS.ti software was used to interpret the instruments during the qualitative document analysis. The textual data were coded and subsequently organized into categories and subcategories, which focused on the characteristics of the documents. This was done by establishing relationships of meaning between the previously recognized codes. Conceptual networks were then created, centred on the findings.

4 Research Results

4.1 Assessment Instruments Designed by Chilean In-Service EFL Teachers

The quantitative document analysis revealed the frequency of the use of the 824 assessment instruments developed by the Chilean in-service EFL teachers in their instructional practices (see Table 1). Within this framework, the instrument they use the most is the test (74%). Then, in a smaller proportion of cases, these professionals design speaking/writing performance evaluations with rubrics (16%) and speaking/writing performance evaluations with rating scales (10%).

Table 1. Frequency of assessment instruments designed by in-service EFL teachers.

<table>
<thead>
<tr>
<th>Assessment instrument</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>610</td>
<td>74%</td>
</tr>
<tr>
<td>Speaking/writing performance evaluation with rubric</td>
<td>130</td>
<td>16%</td>
</tr>
<tr>
<td>Speaking/writing performance evaluation with rating scale</td>
<td>84</td>
<td>10%</td>
</tr>
<tr>
<td>Total</td>
<td>824</td>
<td>100%</td>
</tr>
</tbody>
</table>
Concerning the *speaking/writing performance evaluations*, these include scales, such as rubrics or rating scales. As stated in the Theoretical Framework, both help teachers evaluate, measure, and identify learners’ language productive performance as they integrate criteria (aspects to assess/evaluate) and indicators (criteria’s performance level). In Chile, rubrics provide students with more detailed and specific feedback, so they include descriptors (criteria’s performance level) that are useful to determine the quality of learner production. On the other hand, rating scales assign a score to students, presenting indicators (performance levels) based on criteria without describing them.

### 4.2 Language Dimensions Encompassed in the Assessment Instruments Designed by In-Service EFL Teachers

Considering the document analysis of the assessment instruments designed by the Chilean in-service EFL teachers, it is relevant to establish the frequency of particular language dimensions. This involved calculating how many language skills (listening, reading, speaking, and writing) and/or linguistic systems (grammar, vocabulary, or pronunciation) each assessment instrument covered as a macro category.

Assuming that assessment instruments may evaluate one or multiple skills or linguistic systems in second/foreign language assessment, a value of 1 was given for the presence of each. After analysing the 824 documents, all the values for each dimension/system were summed to calculate their frequency and percentage.

Firstly, the instruments focus on the assessment of language skills (see Table 2). Within this framework, the language ability they most seek to assess is *writing* in English (39%), followed by *reading* (31.3%). Consequently, the instruments assess the development of *listening* (18.2%) and *speaking* (11.5%) skills to a lesser degree.

<table>
<thead>
<tr>
<th>Language dimension</th>
<th>Dimension frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing</td>
<td>574</td>
<td>39%</td>
</tr>
<tr>
<td>Reading</td>
<td>462</td>
<td>31.3%</td>
</tr>
<tr>
<td>Listening</td>
<td>269</td>
<td>18.2%</td>
</tr>
<tr>
<td>Speaking</td>
<td>169</td>
<td>11.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,474</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Table 2. Frequency regarding the consideration of language skills in the assessment instruments.

Assuming that assessment instruments may evaluate one or multiple skills or linguistic systems in second/foreign language assessment, a value of 1 was given for the presence of each. After analysing the 824 documents, all the values for each dimension/system were summed to calculate their frequency and percentage.

Firstly, the instruments focus on the assessment of language skills (see Table 2). Within this framework, the language ability they most seek to assess is *writing* in English (39%), followed by *reading* (31.3%). Consequently, the instruments assess the development of *listening* (18.2%) and *speaking* (11.5%) skills to a lesser degree.

<table>
<thead>
<tr>
<th>Language dimension</th>
<th>Dimension frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>726</td>
<td>50%</td>
</tr>
<tr>
<td>Grammar</td>
<td>654</td>
<td>45%</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>76</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,456</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Table 3. Frequency regarding the consideration of linguistic contents in the assessment instruments.
Secondly, in terms of the linguistic contents in English that the assessment instruments present (see Table 3), it is possible to observe that these tools are mostly based on vocabulary (50%) and grammar (45%). Finally, the instruments also show that pronunciation in English is assessed to a lesser degree (5%).

4.3 Characteristics of the Assessment Instruments Designed by In-Service EFL Teachers and the Frequency of Their Assessment Tasks

Test

The qualitative document analysis identified the tests as assessment instruments designed by the in-service EFL teachers, whose characteristics are illustrated through a conceptual network (see Figure 1). The types of test tasks included in these instruments relate to the following learning objectives: to develop comprehension skills, to develop comprehension production skills, and to apply linguistic contents.

![Figure 1. Tests as assessment instruments designed by in-service EFL teachers.](image)

On the one hand, the tests designed by in-service EFL teachers include test tasks that seek to assess the learning objective to develop comprehension skills. The tests include reading activities (739 out of 1585, 46.6%) and listening activities (314 out of 1585, 19.8%), which are used similarly. They consist of open-ended and close-ended tasks, through which students must recognize and interpret general and specific information from oral or written texts that provide exposure to the foreign language. Regarding the document analysis, the open-ended tasks (187, 11.8% of the total) are mainly questions learners must answer with short or paragraph responses in English based on a text. The following test excerpt is an example of this type of reading test task: “Answer the following questions: What is the most serious problem presented in the text? What is the most effective plan of action? Why?” (Participant 14). Another participant formulates a similar listening test task: “What is the purpose of the text? Why?” (Participant 32).

On the other hand, within the framework of the learning objective to develop comprehension skills, the analysed tests contain items only related to close-ended tasks (866, 54.6% of the
In these the students must respond by producing or selecting a predetermined answer considering general and specific information from the texts they have read or listened to. Within this, the English tests show multiple-choice test tasks (283, 17.9% of the total), in which learners must answer, opting for only one of the choices presented to them. Of all the activities, 11.8% involve reading (187) and 6.1% involve listening (96). The following test excerpt is an example of this: “Listen again and answer by marking the correct alternative: What activity will Ted do? a) He will have a party, b) He will meet with his friends, c) He will study for his exams” (Participant 2). Another informant designed a similar item in one of his reading tests: “Regarding the written text, answer the questions, circling the correct alternative” (Participant 7).

Considering the listening or reading close-ended tasks included in English tests, another activity corresponds to true-false tasks (110, 6.9% of the total). In these, students must read statements presented to them and specify whether they integrate information that appears in the texts or not. Of all the cases, 5.8% involve reading (92), while only 1.1% involve listening (18). The following test excerpt refers to this type of reading test task: “According to the text you read, write T if the information is true or F if it is false: 1) The scene takes place in the living room; 2) Bob is visiting his family” (Participant 10). Another participant designed a similar item to assess listening: “While you are listening to the text, decide whether the following sentences are true or false. Justify the false ones” (Participant 35).

Based on this document analysis, banked gap-fill tasks are also close-ended exercises within tests that assess comprehension skills, worth 18.9% of the total test tasks (299). In these, students had to fill in blanks in paragraphs or sentences with specific information from oral and/or written input texts in English. Of all these exercises, 12.7% involve reading (201), while 6.2% involve listening (98). This can be seen in the following segment of a reading test task: “Complete these sentences with information from the text you are reading: a) Teenagers are different from those of the past. Their ___ habits, way of life, and ___ are an example” (Participant 12). Another test includes the following task: “Listen and fill in the blanks: 1) Cinthia’s hair is ___ and ___. She goes to the hairdresser every ___” (Participant 41).

Additionally, the English tests designed by the study subjects contain test tasks related to the learning objective to develop production skills, which is mostly associated with the writing skills in the foreign language (201 test tasks, worth 12.7% of the total). This involves the production of written texts, including from paragraphs to texts of greater complexity and length within a test. This is shown in the following test segment: “Write, in a paragraph, about your opinion based on the text you read. Justify your answer” (Participant 72). Another test task presents a similar item: “According to the text you listened to, write a short essay about pollution in the world. This should have an introduction, body, and conclusion” (Participant 98). It should be noted that this type of test task is graded by using rating scales or rubrics, which are part of the main evaluation instruments designed by the in-service EFL teachers.

Other English test tasks relate to the learning objective to apply linguistic contents. Through these, students must demonstrate their conceptual knowledge regarding the correct use of grammar and vocabulary. One of these test tasks focuses on banked gap-filling (136 test tasks, worth 8.6% of the total), in which learners must fill in the gaps of the provided sentences or
paragraphs, considering the relevant language structures or words. Of the total tasks, 4.9% are related to grammar (78), while only 3.7% are related to vocabulary (58). The following is an example: “Complete with don’t or doesn’t. Bill __ plays tennis every Saturday” (Participant 11). Another text excerpt presents the instructions for a task related to vocabulary: “Write the following words in the empty spaces of the sentences: apply – temporary job – CV – interesting – full-time job – boring” (Participant 14).

Sentence transformation (6.4%, 101) is another test task associated with the learning objective to apply linguistic contents. In this context, students must paraphrase, demonstrating their knowledge of English grammar. The following test excerpt illustrates this form of assessment: “Considering the learning contents that relate to reported speech, create the second sentence, expressing a meaning similar to the first one and use the specified word” (Participant 20). Another test presents a similar item: “Transform these sentences, using the content of passive voice correctly: 1) Paul and Peter watched the soccer game” (Participant 54).

Word order (1.9%, 30) is also one of the types of test tasks present in English tests that evaluate the learning objective to apply linguistic contents. In these, students are presented with sentences whose components appear out of order. Therefore, they must rewrite them, considering their correct structure in the English language, as shown in the following example: “Order the following sentences: 1) I / the piano / played” (Participant 6). The following excerpt also integrates an equivalent test task: “Order the statements... 5) exciting / is / hobby / kayaking / an” (Participant 8).

There are also matching tasks based on lexical items in the tests (4.1%, 65), framed within the learning objective to apply linguistic contents. They essentially emphasize English vocabulary knowledge and include activities where students must match two equivalent test tasks: “Match the following words with their corresponding English definition” (Participant 14). A similar test also integrates one of these test tasks: “By writing an arrow, match the words in column A with their synonyms in column B” (Participant 28).

Table 4 refers to the frequency of tasks included in tests designed by the in-service EFL teachers. In this context, the quantitative document analysis helps to reveal that the most common tasks focus on assessing the development of reading (739, 46.6% of the tasks), listening (314, 19.8% of the tasks), grammar (209, 13.2% of the tasks), writing (201, 12.7%), and vocabulary (123, 7.8% of the tasks). Hence, it is possible to establish that in-service EFL teachers measure receptive skills, writing, and the application of linguistic contents to a greater extent by using tests.

The test task frequency was calculated by counting the assignments that were part of the tests provided by the participants. These instruments included from one to 10 test tasks with items within them.

**Speaking and writing performance evaluations utilizing rubrics and rating scales**

To assess the learning objective to develop production skills, the in-service EFL teachers also consider speaking and writing performance evaluations that are utilized with rubrics and rating
scales. However, these are used to a lesser degree. Their characteristics are illustrated through a conceptual network (see Figure 2), and these instruments include assessment criteria related to language fluency and accuracy.

The analysed rating scales and rubrics contain assessment criteria, whose fulfilment is assessed through indicators that provide information about each level of performance (criteria) achieved by students. These are equivalent to scores. As stated in the Theoretical Framework, rubrics have more detailed descriptors (criteria’s performance level), while rating scales present only the level of performance (indicator) for each assessment criterion. The analysed rubrics are all analytic, meaning they divide student performance into multiple dimensions or criteria.

Both scales are designed to assess the development of production skills in English to support speaking and writing performance evaluations. They incorporate, in the first place, criteria linked to fluency. Within this, students must demonstrate that their ideas are understandable and follow a logical order when writing or speaking in the foreign language. This is expressed in the following segment of a rating scale’s criterion: “All interactions, in the role-playing performance, follow a clear sequence” (Participant 16). A similar element is evident in a rubric descriptor associated with an assessment criterion: “The student produces the language without interruptions. His/her contributions are relevant, and there is a clear organization of ideas (level 5)” (Participant 20).

On the other hand, the analysed rating scales and rubrics also have assessment criteria framed in language accuracy. They state that students must demonstrate the correct use of the

### Table 4. Frequency of test tasks.

<table>
<thead>
<tr>
<th>Test task</th>
<th>Task frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading banked gap-fill</td>
<td>201</td>
<td>12.7%</td>
</tr>
<tr>
<td>Reading multiple-choice</td>
<td>187</td>
<td>11.8%</td>
</tr>
<tr>
<td>Answering reading questions</td>
<td>149</td>
<td>9.4%</td>
</tr>
<tr>
<td>Reading matching</td>
<td>110</td>
<td>6.9%</td>
</tr>
<tr>
<td>Sentence writing</td>
<td>103</td>
<td>6.5%</td>
</tr>
<tr>
<td>Grammar transformation</td>
<td>101</td>
<td>6.4%</td>
</tr>
<tr>
<td>Listening banked gap-fill</td>
<td>98</td>
<td>6.2%</td>
</tr>
<tr>
<td>Text writing</td>
<td>97</td>
<td>6.1%</td>
</tr>
<tr>
<td>Listening multiple-choice</td>
<td>96</td>
<td>6.1%</td>
</tr>
<tr>
<td>Reading true-false</td>
<td>92</td>
<td>5.8%</td>
</tr>
<tr>
<td>Grammar banked gap-fill</td>
<td>78</td>
<td>4.9%</td>
</tr>
<tr>
<td>Vocabulary matching</td>
<td>65</td>
<td>4.1%</td>
</tr>
<tr>
<td>Listening matching</td>
<td>64</td>
<td>4.0%</td>
</tr>
<tr>
<td>Vocabulary banked gap-fill</td>
<td>58</td>
<td>3.7%</td>
</tr>
<tr>
<td>Answering listening questions</td>
<td>38</td>
<td>2.4%</td>
</tr>
<tr>
<td>Word order</td>
<td>30</td>
<td>1.9%</td>
</tr>
<tr>
<td>Listening true-false</td>
<td>18</td>
<td>1.1%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,585</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>
conceptual knowledge of language form or lexicon when writing or speaking in English. For example, a segment of a rating scale refers to language accuracy: “The student, in his/her written report, writes correctly using the simple present tense in English” (Participant 32). Similarly, the following rubric descriptor exhibits the assessment of grammar knowledge: “Excellent performance – 3 points: The student demonstrates full command of simple and complex grammatical forms when speaking” (Participant 36).

The assessment criteria of the rating scales and rubrics used to assess the language accuracy of speaking and writing skills also allude to the vocabulary knowledge of English. This is illustrated below in a rating scale segment: “The student, in his/her oral presentation, uses words studied in the third learning unit” (Participant 29). This is further evidenced in the assessment criteria of another rubric: “5 points – No errors considering vocabulary. The student presents a high command in this dimension” (Participant 40).

Considering the assessment criteria linked to linguistic accuracy, these are also associated with knowledge of pronunciation. Within this, the assessment seeks to recognize whether the student correctly articulates the sounds of the language concerning vocabulary when developing the speaking skills. The following example presents a rating scale’s assessment criterion: “In the oral presentation, all words are pronounced correctly” (Participant 37). This aspect is also mentioned in a criterion descriptor present in a rubric used to assess speaking: “Not achieved (0 points) – The student presents many pronunciation problems that interfere with the comprehension of the dialogue” (Participant 9).
Table 5 presents the frequency of assessment tasks related to the speaking and writing performance evaluations designed by in-service EFL teachers. These activities are employed along with rubrics and rating scales. The quantitative document analysis made it possible to identify that these activities mainly focus on producing written texts (60.1%) and oral presentations (29%). The data show that the study subjects tend to assess writing skills more often with these instruments.

The task frequency was calculated by counting the speaking and writing performance instruments provided by the participants. These documents included the instructions or procedures and the rating scale or rubric.

Table 5. Frequency of assessment tasks related to speaking and writing performance evaluations utilized with rubrics and rating scales.

<table>
<thead>
<tr>
<th>Assessment task</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production of written texts</td>
<td>172</td>
<td>60.1%</td>
</tr>
<tr>
<td>Oral presentations</td>
<td>83</td>
<td>29%</td>
</tr>
<tr>
<td>Open-ended speaking tasks</td>
<td>21</td>
<td>7.3%</td>
</tr>
<tr>
<td>Role-play</td>
<td>10</td>
<td>3.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>286</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

5 Discussion of Results

The participants demonstrated a strong tendency to design their tests as an English language assessment instrument. This preference is related to traditional assessment, which may minimize the development of competences or performances that are put into practice in daily life (Al-Nouh, Taqi, and Abdul-Kareem 2014). The reason for this evaluative practice could be due to the versatility of this instrument, since in practice it is possible to adjust it to the characteristics of various educational scenarios (Giraldo 2018; Scully 2017). As a result, summative evaluations predominate, which could mean there are fewer opportunities to provide continuous student support or feedback (Butt 2010).

Additionally, in-service EFL teachers mostly consider writing and reading skills as their assessment foci, and listening and speaking skills are less emphasized. These assessment practices are also in line with what occurs in the area of linguistic contents, since those associated with pronunciation are almost entirely neglected. Consequently, learning the oral dimension of the foreign language and its assessment could be underrated, which then may have the effect that students do not develop the four linguistic skills. Therefore, experts in second language teaching and learning suggest that assessment in this area should consider the component of authenticity (see, for example, Brown and Abeywickrama 2018; Coombe 2012). This refers to designing assessment practices and instruments that incorporate language as naturally as possible, based on tasks involving communicative actions from daily life.

The document analysis presented in this study establishes that EFL teachers design production performance evaluations using rubrics and rating scales to a lesser degree, and show a tendency to evaluate the learning of writing. These practices may lack a link between
comprehension and production language skills, as in contexts where communication is developed authentically and functionally (Aliasin and Amianlu 2017).

Based on the previous points, the preference for summative/traditional assessment instruments that also prioritize certain language skills over others, segregating some or treating them separately, can be risky for the development of learners’ English language skills, because the foreign language would not be taught within the framework of communicative approaches. Instead, different specialists suggest that language learning, and its corresponding assessment, should emphasize the use of both comprehension and production skills, simulating meaningful situations where the language is used as in everyday life (Murray and Christison 2020; Richard-Amato 2010). Several authors also state that effective language instruction requires opportunities for both language exposure and production (see, for example, Gottlieb 2016; Nunan 2015).

The findings of this research highlight some needs related to the didactic knowledge of in-service teachers of English. For this reason, it is suggested that higher education institutions and school administrators generate instances of professional development improvement and disciplinary updating. Moreover, teachers should reflect on the impact of their assessment practices on their students’ learning of English (Giraldo 2019). In this sense, action research could be used so that teachers assume the role of researchers to contribute practical improvements to the community through a self-reflective, critical, and inquiry-oriented process (see, for example, Burns 2010). The principles of this proposal perceive teachers as reflective professionals who evaluate their practices and professional growth, constantly questioning themselves about student learning.

In addition, English teacher education institutions need to strengthen and monitor the development of the assessment competency of pre-service teachers in the undergraduate subjects of didactics and disciplinary practice (Tagle et al. 2022). Therefore, potential English teachers should be familiar with a variety of language assessment instruments, both traditional and authentic, that focus on comprehension and productive language skills from a communicative perspective. Likewise, in their professional preparation pre-service EFL teachers should be encouraged to design and implement assessment practices by questioning and reflecting on their impact on student English language learning.

6 Conclusions

The findings of the present study indicate that the in-service teachers of English mainly design their tests as an evaluation instrument, considering production performance evaluations using rubrics and rating scales to a lesser extent. Their assessment practices focus on the writing and reading skills, and the linguistic contents of vocabulary and grammar, which suggests that the oral dimension of the language is considered a secondary aspect. This logic may hinder the learning of the four linguistic skills in an integrated manner, since they would be taught in a piecemeal fashion, not following the principles of communicative approaches.

As per the above, the results show a strong tendency to use traditional and summative evaluation, marginalizing authentic and formative assessment. Consequently, these
practices may contribute little to constructing and verifying student learning related to their communicative competences in a foreign language.

It is suggested that universities and administrators of educational institutions encourage in-service teachers of English to increase and update their didactic knowledge focused on assessment competencies. Similarly, teacher education programmes should strengthen courses focused on this didactic dimension, favouring the design and implementation of assessment instruments in practice, complementing this with the competency of pedagogical reflection. Within this, longitudinal studies should be conducted to monitor how the assessment practices of teaching professionals change from the time they are undergraduate students until they are integrated into educational establishments as in-service teachers. This would allow us to keep track of the impact of pre-service teacher training on the pedagogical performance of candidates.

References


