Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6, 9328-9344
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i6.3996

Learning to assess foreign language listening: Analyzing the assessment items designed by Chilean pre-service EFL teachers and the cognitive load they experience in this process

Tania Tagle[1], Paulo Etchegaray[2*], Claudio Díaz[3], Mabel Ortiz[4], Lucía Ramos[5], Marcela Quintana[6]
[1]Universidad Católica de Temuco, Chile; ttagle@uct.cl (T.T.).
[2]Universidad Autónoma de Chile; paulo.etchegaray@uautonoma.cl (P.E.).
[3]Universidad de Concepción, Chile; claudiodiaz@udec.cl (C.D.).
[4]Universidad Católica de la Santísima Concepción, Chile; mortiz@ucsc.cl (M.O.).
[5]Universidad Católica del Norte, Chile; luramos@ucn.cl (L.R.).
[6]Universidad Arturo Prat, Chile; maquinta@unap.cl (M.Q.).

Abstract: The general objective of this study is to analyze the types of items designed by Chilean pre-service EFL teachers to assess listening comprehension skills in English and the cognitive load they experience during the test design process. This study employs a mixed-method design, including 125 Chilean pre-service teachers of English in their third, fourth, and fifth year of professional preparation. As data collection techniques, the participants created a listening comprehension test for secondary education English language learners. Moreover, they completed the Cognitive Load Scale (Leppink et al., 2013). The findings indicate that all pre-service EFL teachers create close and open-ended questions to evaluate students' listening comprehension, focusing on identifying specific information. Fourth and fifth-year pre-service teachers design close and open-ended questions to identify general and specific information, encompassing higher levels of comprehension. On the other hand, regarding the cognitive load experienced by the participants in the test design, they exhibited moderate intrinsic cognitive load, demonstrating that the task was reasonably challenging. The extraneous load was generally low, suggesting minimal external distractions. Notably, the germane load was high across all groups, indicating that the task effectively promoted learning and competency development. Pre-service language teacher education programs should consider these findings to enhance listening assessment practices. Incorporating cognitive load theory can assist teacher candidates in building both pedagogical and disciplinary teaching competencies.

Keywords: *Cognitive load theory, Evaluation, Language assessment, Listening skills, Pre-service teacher education, Test design.*

1. Introduction

Following communicative practices, English language teaching and learning practices should focus on developing receptive (reading and listening) and productive (speaking and writing) language skills (Ashoori Tootkaboni, 2019). Implementing communicative approaches aids students in enhancing their abilities to understand and produce information in a foreign language (Porto, 2018). The evidence suggests that successful assessment practices could aid students in learning critical skills as teachers of English apply different assessment practices to collect data regarding students' learning progress (Fitriyah & Jannah, 2021; Fröjdendahl, 2018). This helps education professionals evaluate their teaching and make proper decisions.

* Correspondence: paulo.etchegaray@uautonoma.cl

Authors highlight various processes that English language teachers should develop when creating assessment tools (Bachman & Palmer, 2017; Fulcher, 2012; Giraldo, 2018; Taylor, 2009): relating assessment to relevant language learning and teaching approaches and theories; establishing the purpose of language assessment; constructing instrument specifications; designing instruments that are reliable, authentic, fair, ethical, practical, and interactive, among other processes.

Language teacher education aims to contribute to developing prospective EFL teachers' assessment competencies (Giraldo & Murcia, 2019; Güngör & Güngör, 2024). In this regard, language teaching programs should equip teacher candidates with assessment theoretical knowledge and practical skills, while boosting their confidence to design and implement multiple assessment practices (Oo et al., 2022). However, undergraduate professional preparation may not have a sufficient impact on the student teachers' assessment knowledge. According to some studies, pre-service teachers of English struggle with assessing students' foreign language skills learning and communicating assessment results (Prastikawati et al., 2021; Restrepo, 2020; Tagle et al., 2024). This suggests that teacher candidates may not successfully articulate teaching, learning, and assessment.

In Chile, English teacher training programs usually include a general semester course on assessing, which does not address specific techniques and instruments for assessing a foreign language, nor does it develop a reflective attitude towards the process (British Council, 2015). Based on this point, it has been shown that pre-service and in-service teachers state that they learn how to evaluate in professional practice and not in their university preparation (Earl, 2014; Fröjdendahl, 2018).

A study focused on the pedagogical and disciplinary knowledge of pre-service EFL teachers (Chilean Ministry of Education, 2023) highlights that evaluation and assessment skills need enhancement. Furthermore, data from the 2023 Teacher Evaluation (Chilean Ministry of Education, 2024) indicates a low number of educators who exhibited a successful performance in this competency area. Therefore, interventions in initial teacher training programs based on updated evaluation practices and the design of contextualized and well-constructed instruments are needed.

Different studies framed in human cognitive architecture have established that a low cognitive load is conducive to learning communicative skills in English (Lin et al., 2016; Nawal, 2018; Roussel et al., 2017). This relates to instructional and evaluation designs that present lower levels of interactivity in elements and resources, enhancing comprehension and supporting performance in receptive and productive skills in a foreign language. (Hanham et al., 2017). In this context, research has been conducted on the implications of cognitive load theory on how student teachers learn (see, for example, Odacı & Uzun, 2024); however, there is a deficit of studies focused on the cognitive load experienced by prospective English teachers when designing assessment instruments. Consequently, it is important to investigate this aspect within pre-service language teacher education since ineffective handling of cognitive load theory could hinder the development of assessment skills among teacher candidates.

Based on the previous point, the general objective of this research is to analyze the types of items designed by Chilean pre-service EFL teachers to assess listening comprehension skills in English and the cognitive load they experience during the test design process.

Likewise, the present study seeks to answer the following subsidiary research questions:

c) What types of items do Chilean pre-service EFL teachers design to assess listening comprehension skills in English?

b) What levels of cognitive load do Chilean pre-service EFL teachers experience when designing an instrument to assess listening comprehension skills in English?

2. Theoretical Framework

2.1. Listening Comprehension Skills in English

Listening comprehension is a communicative skill, like reading comprehension, oral production, and written production. It is essential in learning a foreign language since it consists of understanding information from oral texts (Siegel, 2014). This ability encompasses an active process of meaning construction, in which listeners decode from information related to minor linguistic components of oral

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

texts to more extensive ones (bottom-up form of information processing) and activate their prior knowledge or cognitive schemas related to the topics and information of oral texts (top-down form of information processing) (Bailey, 2020).

In addition, Rost (2013, 2016) mentions that listening comprehension is dynamic as it depends on various types of cognitive processing related to each other. From this author's perspective, listening is related to neurological processing, including hearing, awareness, and attention. Linguistic processing is the dimension of the skill that considers comprehension from a linguistic source. Semantic processing is associated with the construction of meaning in comprehension through the activation of appropriate memory structures. Another component of listening comprehension is pragmatic processing, which focuses on social and cultural context, assigning an active role to listeners in constructing meaning while comprehending.

Listening involves subskills, which look at specific reasons for listening to a spoken text or the skill's components that facilitate language users' comprehension (Zhao & Lee, 2022). Examples of listening subskills are listening to identify general information, which has to do with global comprehension of the oral presentation; listening to identify specific information, through which listeners recognize the key ideas of the text; listening to identify details, which involves understanding elements of a text, such as lexical components, grammar, sounds, among others; and listening to infer attitudes, through which language users identify the attitude expressed by speakers (Brown & Lee, 2015; Burns & Siegel, 2023).

Considering comprehension development, Day and Park (2005) suggest six levels, which can also be associated with listening comprehension ability. The most basic level suggested by the authors is literal comprehension, which encompasses explicit textual information. Then, the reorganization level consists of making connections between ideas. According to Day and Park (2005), the third level of comprehension is inferential comprehension, in which language users create mental representations based on their prior knowledge to create meanings about the text. The fourth level of comprehension is prediction, which uses prior knowledge and understanding of the text to determine what will happen next in the text. The fifth level of comprehension is evaluation, whereby auditors critically judge the text. Finally, the sixth level of comprehension is personal response, in which individuals produce a text, indicating their feelings or other personal aspects, focusing on the textual information.

2.1.1. Listening Comprehension Strategies

In the context of listening comprehension skills, language users develop this ability by employing learning strategies. These are understood as actions that facilitate the learner's construction of knowledge, making this process self-directed and more transferable to new situations (Oxford, 1990). Different authors recognize three categories of listening comprehension strategies: cognitive, metacognitive, and socio-affective (Goh, 1998; O'Malley & Chamot, 1990).

First, cognitive strategies are associated with problem-solving in learning activities. Some examples of this type of strategy involve making inferences, elaboration, and note-taking, among others (Cross, 2009; Vandergrift, 2003).

Second, metacognitive strategies are procedures students use to manage their learning, performing success-based monitoring of this process (Goh & Aryadoust, 2015). In listening comprehension, some strategies are related to planning, monitoring linguistic performance, learner evaluation of the communicative situation, and self-assessment (Richards, 2008).

Third, social-affective strategies involve interacting with other human beings and controlling emotions (Thanh Nha & Hong Dung, 2020). Some of these strategies that individuals use when comprehending an oral text correspond to cooperation with others, requests for clarification, and self-motivation (Chamot, 2005; Cross, 2009).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

2.2. Assessment in English Language Teaching and Learning

Educational assessment is the systematic process of collecting and documenting evidence of student performance concerning established learning objectives (Britton, 2021; Solano-Flores, 2016). This process includes data collection, followed by teachers' interpretation of these data. In this line, 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, or standards on linguistic competence, in addition to the pedagogical process (Gottlieb, 2016).

Assessment can be diagnostic, formative, and summative (Marsh, 2010; Zeng & Huang, 2021). The first is carried out at the beginning of a didactic unit to identify students' strengths and weaknesses and make planning decisions. Formative evaluation is carried out during the pedagogical process to help students achieve the learning objectives. In contrast, summative evaluation involves measuring learning at the end of the didactic unit, usually using a grade.

Teachers can assess learning through three approaches. First, it is possible to refer to the assessment of learning, which is summative and is used to identify what students have learned as a result of a pedagogical intervention (Earl & 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 & Stiggins, 2020). In English language teaching and learning, these practices are linked to traditional assessment, in which instruments such as multiple-choice, true-false, and information completeness items are used, focusing on accurately using linguistic structures (Phongsirikul, 2018). Research findings have posited that these assessment practices primarily seek to prepare students to be successful test takers rather than to help them develop communicative skills (Wubshet & Menuta, 2015). Likewise, studies have reported that traditional assessment does not significantly impact learning English language skills (Forutan, 2014; Goçtü, 2012).

Another approach is assessment for learning, which is conducted for formative purposes. These evaluative practices are employed to help students construct their learning throughout the pedagogical 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 over linguistic structures. Likewise, it understands learning as a social process, so interaction tasks, co-assessment, and self-assessment procedures play a relevant role (Heritage, 2022).

On the other hand, it is possible to allude to assessment as learning. This evaluation approach corresponds to an active process of cognitive restructuring that occurs when learners interact with new ideas (Earl & Katz, 2006). This type of evaluative performance is related to metacognition, which implies knowledge of one's thinking processes.

2.3. Listening Comprehension Assessment Instruments

Listening comprehension in English is often evaluated using traditional assessment tools, such as standardized written tests. These tests typically include closed-response items (like multiple-choice questions, short answers, and sentence completion) and open-response items. Additionally, they may assess linguistic components such as grammar and vocabulary (Mihai, 2010). Different specialists have criticized these traditional assessment instruments because they emphasize the product and basic cognitive processes of comprehension and do not focus on language skill learning, evidencing a limited use of the English language (Bae & Lee, 2018; Chou, 2017).

On the other hand, to assess listening comprehension skills in English, authentic assessment instruments are suggested that consist of performance tasks based on the comprehension of different types of texts on real-world topics or situations, which involves the use of comprehension strategies and the development of higher-order cognitive processes (Ockey & Wagner, 2018; Sevilla & Chaves-Fernandez, 2019). These are employed during the learning process, privileging feedback for the learner. Some examples of authentic assessment instruments are listening comprehension tasks based on higher cognitive skills, including concept maps, portfolios, presentations, and dialogues, which can be complemented with self-assessment or co-assessment practices (Fan et al., 2020; Rojas, 2017).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

2.4. Cognitive Load Theory

Cognitive load theory primarily examines how individuals process and retain information while constructing knowledge, which is essential for effective instructional design (Gillmor et al., 2015; Leppink et al., 2013; Sweller, 2010; Van Merriënboer & Sweller, 2005). This process involves two types of memory: working memory, where information is initially organized, and long-term memory, where information is retained after learning has taken place. Cognitive load refers to the effort working memory requires to process incoming information and connect it to an individual's prior knowledge during the learning process (Paas et al., 2008; Sweller, 2018). From this viewpoint, learning occurs when working memory assigns meanings to information, organizing it into existing cognitive schemas. This information is then stored in long-term memory, which has an unlimited capacity.

It has been argued that long-term memory is limited in capacity and duration, so if the cognitive load is high, it collapses, and consequently, the construction of learning does not take place successfully (Ayres, 2020; Hong Yang & Farley, 2019). Considering this perspective, planning teaching and learning processes should aim to keep learners' cognitive load low to facilitate practical knowledge construction. This means reducing cognitive load by carefully selecting the content, activities, resources, and assessments (Roussel et al., 2017).

Cognitive load theory identifies three distinct types of cognitive load: intrinsic, extraneous, and germane cognitive load (Klepsch et al., 2017; Leppink et al., 2013; Martin, 2014). Intrinsic cognitive load refers to the complexity of the material or task and the learners' prior knowledge. This type of load is not something that teachers can change or adjust. Extraneous cognitive load is associated with ineffective instructional design, particularly regarding how information is presented to learners. Germane cognitive load occurs when cognitive resources are effectively utilized to construct, process, and automate cognitive schemas. This load is facilitated by instructional designs prioritizing relevant learning processes.

Teachers have control over extraneous and germane cognitive load when designing the teaching, learning, and assessment processes (Chen & Chang, 2009). In this framework, assessment of and for learning promotes learning construction as long as the materials and instruments used present low item interactivity (Hanham et al., 2017). Cognitive load is reduced if fewer components are processed simultaneously in working memory.

To reduce the cognitive load associated with assessment tools, it is recommended to emphasize essential information, organize test items attractively, and remove any unnecessary information or components (Gillmor et al., 2015). In the context of English language learning assessment, using visual or multimedia digital resources to present information is recommended as it promotes a balance between types of cognitive load (Costley et al., 2021; Liu. 2011).

3. Research Methodology

This study employs a mixed-method research design. The qualitative dimension of this research was used to describe the types of items Chilean pre-service EFL teachers design to assess listening comprehension skills in English in their third, fourth, and fifth year of professional preparation. Moreover, quantitative research was utilized to determine the cognitive load experienced by these research subjects in designing an instrument for evaluating listening comprehension skills in English.

3.1. Participants

A total of 125 Chilean pre-service EFL teachers participated in this research. The participants were at different stages of their professional preparation: 46 were in their third year, 58 were in their fourth year, and 21 were in their fifth year. They represented four universities located in various regions of Chile, specifically in the Biobío Region, Antofagasta Region, Tarapacá Region, and Araucanía Region.

All research subjects signed a letter of informed consent, which indicated that their participation was voluntary and that the information they provided would be analyzed anonymously, ensuring their identity would be protected throughout the process.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

All pre-service teachers were enrolled in five-year language teacher education programs based on pedagogical and disciplinary standards established by the Chilean Ministry of Education (2021). Throughout their professional training, teacher candidates engage in various curricular activities to enhance their communicative competence in English. From their first to fifth year, they enroll in foreign language courses designed to provide opportunities for progressively achieving an advanced English language proficiency level. Moreover, these pre-service teachers also have university subjects on foreign language teaching didactics, practicum, linguistics, principles of pedagogy and education, research, and foreign language literature.

The subjects engage in progressive practicum activities from the second year of professional preparation. From the third year onwards, prospective teachers design and implement lesson plans in secondary schools that integrate the development of language skills in English (listening, reading, speaking, and writing). In these practicum experiences, they are encouraged to reflect on their professional performance and the beliefs that inform it.

3.2. Data Collection Techniques

Participants were tasked with creating a listening comprehension test for secondary education English language learners, including close-ended and open-ended questions, with a time limit of 45 minutes for test design. They had to specify the underlying listening strategy, include clear student instructions, and assign a score to each item. Afterward, the teacher candidates submitted their production.

After the research subjects designed their listening comprehension test, they completed the Cognitive Load Scale (Leppink et al., 2013). This scale included three subscales: intrinsic cognitive load (3 items), extraneous cognitive load (3 items), and germane cognitive load (4 items). This instrument had a Cronbach's alpha between .79 and .82 for Latin American contexts. This scale evaluates the difficulty and complexity of designing an English assessment instrument on a scale ranging from 1 (completely disagree) to 10 (completely agree). Operationally, cognitive load is understood as the effort that the individual declares to make when processing incoming information and relating it to his or her previous knowledge when designing an assessment instrument.

The researchers analyzed the teacher candidates' listening test design using documentary analysis. This procedure involves examining previously written documents to understand the authors' perspectives, their actions, or what is occurring to them (Rapley, 2018).

The documentary analysis of the participants' tests contemplated the use of ATLAS.ti software. Data coding was performed, and they were subsequently organized into categories and subcategories focused on the characteristics of the items. This was done by establishing relationships of meaning between the previously recognized codes. Consequently, conceptual networks were created centered on the findings.

On the other hand, the data obtained from the Cognitive Load Scale (Leppink et al., 2013) were analyzed using descriptive statistics. This included measures of central tendency, such as the mean, minimum, and maximum values, as well as measures of dispersion, like standard deviation. Likewise, the Kruskal-Wallis test was utilized to examine differences among groups in the third, fourth, and fifth years of initial teacher training across three dimensions of cognitive load (intrinsic, extraneous, and germane). The quantitative analysis was conducted using SPSS software.

4. Results

4.1. Items to Assess Listening Comprehension Designed by Third-Year Chilean Pre-Service EFL Teachers

In the evaluation of *listening comprehension items created by third-year Chilean pre-service EFL teachers* (see Figure 1), it is noted that this group designed *multiple-choice items*, *true-false items*, and *open-ended questions*.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

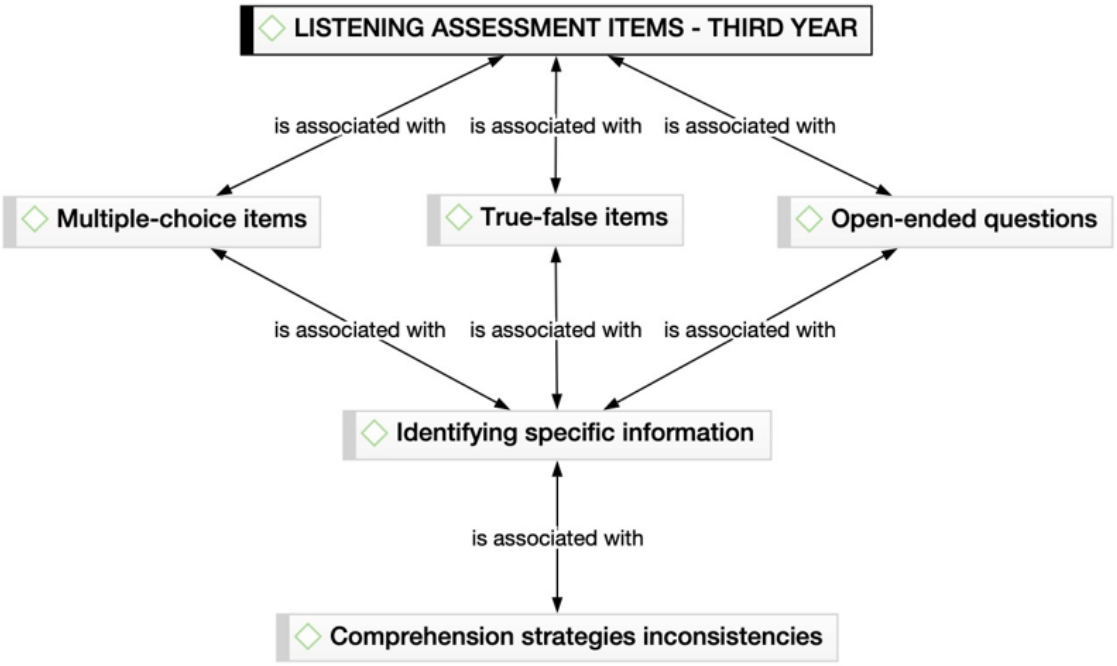

Figure 1.
Conceptual network regarding items to assess listening comprehension designed by third-year Chilean pre-service EFL teachers.

Third-year student teachers create assessment items to evaluate listening skills in English. These assessments primarily consist of closed-ended activities, such as *multiple-choice* or *true-false questions*. The materials developed by the students indicate that their focus is on identifying specific information from an oral text as a comprehension strategy. An example of this can be seen in the following closed-response item: "What time does Susan need to pick up her brother? a) 5:30 pm, b) 12:30 pm, c) 4:30 pm". Likewise, another prospective teacher designs a true-false equivalent item: "According to the conversation, write T if the statement is true and F if the statement is false... Susan has to go to the dentist. Then, she has to clean the house".

On the other hand, the third-year pre-service EFL teachers elaborate on *open-ended questions*, targeting a literal level of listening skills. Therefore, the questions created by these participants usually have only one correct answer according to the information in the text. Evidently, this type of item type focuses on identifying specific information as a comprehension strategy. The following is a segment of the aforementioned task: "Answer the following questions about the text. Why does Ted have to call Susan before he arrives at her house?".

According to the content analysis, it was possible to determine *inconsistencies* considering the *comprehension strategies* stipulated by the participants. This is because those comprehension actions they allude to do not correspond precisely to the type of item they design.

4.2. Items to Assess the Listening Comprehension Skill Designed by Fourth-Year Pre-Service EFL Teachers

The category of *items to assess the listening comprehension skill designed by fourth-year English teaching students* is presented through a conceptual network (see Figure 2). It illustrates the types of items formulated by the participants, which are associated with *multiple-choice items*, *true-false items*, and *open-ended questions*.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

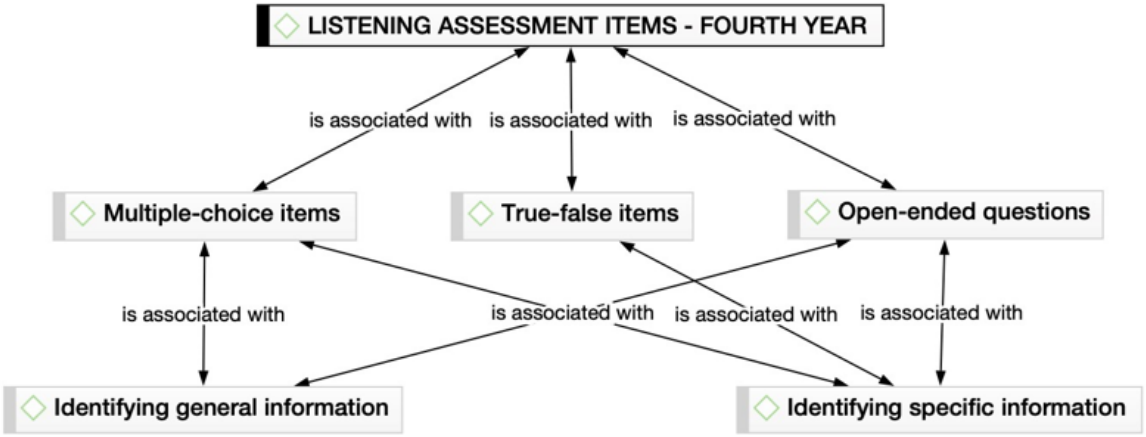

Figure 2.
Conceptual network regarding items to assess the listening comprehension skill designed by fourth-year pre-service EFL teachers.

Fourth-year pre-service EFL teachers design *types of items to assess listening comprehension in English*, which correspond to closed activities, such as *multiple-choice items*. According to the documentary analysis, these focus on *identifying general information* as a comprehension strategy, which implies recognizing or making inferences about implicit ideas in the oral text. This is illustrated in the following excerpt from a multiple-choice item elaborated by a participant: "What is the main topic of the conversation? a) Dating ideas, b) Asking for directions, c) Speaking about work, d) Daily routines".

On the other hand, the items based on closed activities designed by fourth-year future teachers, which include *multiple-choice* and *true-false items*, also focused on the strategy of *identifying specific information* in the same way that third-year students do. An example of this is the segment of a multiple-choice item elaborated by an informant: "Why can't Susan go out with Ted? a) Because she has to go to the doctor, b) Because she has to pick up her friend Julie from soccer, c) Because she is going to be busy all day". Additionally, the excerpt of a true-false item used for the same purpose is presented: "Listen to the audio and state if the following sentences are true (T) or false (F). 1) Susan has to finish her science assignment".

This group of pre-service EFL teachers also formulate questions to evaluate listening comprehension. In this context, it is possible to recognize *open-ended questions* that encourage learners to produce answers based on the spoken text. As in the case of third-year prospective teachers, these tasks target the level of literal comprehension; however, this group also focuses on higher levels of this skill. From the review of these items, it is established that they are based on the strategy of *identifying specific information*. Evidence of this is shown below, where a question that includes a more basic level of comprehension is presented, with only one possible answer: "What were the reasons why Susan could not meet Ted?". In the same way, an example of an open question is presented that allows candidates to produce an answer, integrating their opinion: "Answer the following questions related to the recording in about three to four lines. 1) Why do you think Susan feels like a slave sometimes? Give examples of situations that could have made her feel like that".

4.3. Items to assess the listening comprehension skill designed by fifth-year pre-service EFL teachers

The category of *items to assess listening comprehension skills designed by fifth-year pre-service EFL teachers* is presented through a conceptual network (see Figure 3). It illustrates the types of items proposed by the participants, which are associated with *multiple-choice items*, *true-false items*, and *open-ended questions*.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

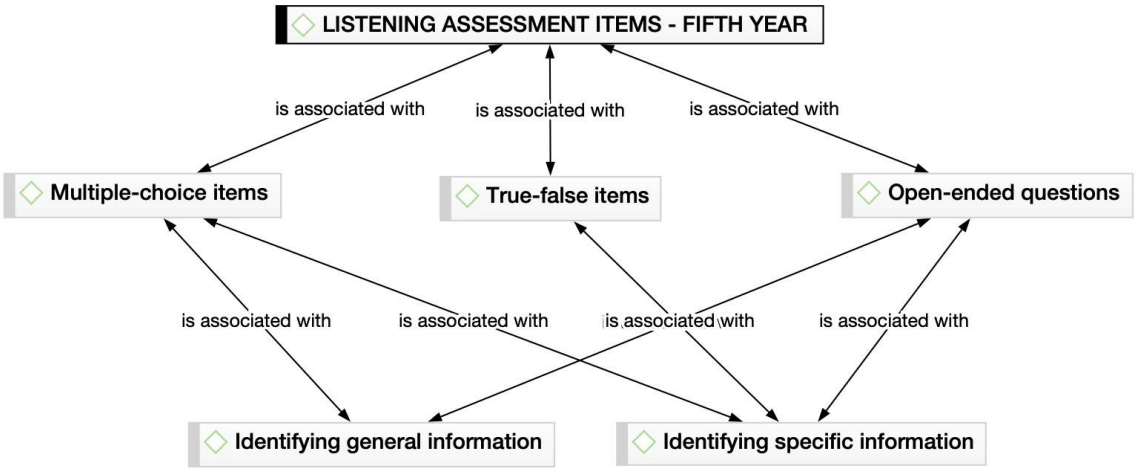

Figure 3.
Conceptual network regarding items to assess the listening comprehension skill designed by fifth-year pre-service EFL teachers.

In their fifth year of professional training, prospective English teachers design item types to assess listening comprehension ability, including closed-ended activities, such as *multiple-choice items*. These present several predetermined answers, including a correct answer that the listener must choose based on the oral text. One of their focuses is the strategy of *identifying general information*, which involves recognizing implicit ideas from the text and making inferences. This is reflected in an item excerpt that points to the type of text: "What type of text is it? a) An interview, b) A monologue, c) A dialogue, d) A podcast". Likewise, another student teacher includes an equivalent item: "What is the purpose of Ted's messages in the conversation? a) to inform, b) to invite, c) to persuade, d) to describe".

Similarly, these participants designed another type of closed activity to assess listening associated with *true-false items*. In these, listeners read statements and determine whether or not they correspond to what the oral text states. According to the documentary analysis of these items, they focus on the strategy of *identifying specific information*. This practice is similar to what third and fourth-year trainee teachers do, as shown in the following item segment: "Write T if the sentence includes true information or F if it is false. 1) Susan says she feels like a slave with all the things she has to do, 2) Ted wants to invite Susan to play football".

Additionally, other items proposed by the fifth-year prospective teachers to assess listening in English correspond to *open-ended questions*. Firstly, this group of student teachers formulate more basic questions since they request the listener's production but consider only one correct answer based on the oral text. On the one hand, these are associated with the strategy of *identifying general information* because they require making inferences, as illustrated below: "Answer the questions according to the dialogue. 1) What is Susan's attitude towards Ted's invitations?" In addition, it is possible to determine that the open-ended questions of a more basic level designed by the future teachers in the fifth-year student teachers focus on the strategy of *identifying specific information* since they point to explicit ideas from the oral text. This practice is evidenced in the following excerpt: "Answer the following questions. 1) What does Susan mean when she says she is "going to be running around all day"?

On the other hand, this group of participants proposes other *open-ended questions* that aim at a higher level of comprehension, where listeners must produce their answers regarding the oral text but also refer to their opinions. These items point, on the one hand, to the strategy of *identifying general information* because they involve making inferences about implicit ideas. The following is an example of this assessment task: "Answer two questions about the text in no more than two paragraphs. 1) What do you think about the dating ideas that Ted used?". On the other hand, the open-ended questions

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

designed by the participants underlie the strategy of identifying specific information, focusing on explicit ideas of the text, as evidenced in the following excerpt: "Answer the following open-ended questions according to the oral text. 1) Why do you think Susan is too busy?".

4.4. Cognitive Load Experienced by Chilean Pre-Service EFL Teachers when Designing an Instrument to Assess Listening Comprehension Skills in English

Table 1 presents the results of the cognitive load experienced by participants while designing an instrument to assess listening comprehension skills in English. A descriptive analysis of cognitive load dimensions and their respective items, including intrinsic, extraneous, and germane loads, is provided. This quantitative descriptive analysis utilizes measures of central tendency, including the Mean, Minimum (Min), and Maximum (Max) values, as well as the Standard Deviation (SD) as measures of dispersion.

Table 1.
Results by dimension and items for all participants.

Dimension / item	Mean	SD	Min	Max
Dimension: Intrinsic load	3.43	2.19	0	8
1. The topic of designing a listening test was very complex.	3.31	2.54	0	8
2. The test design task covered listening comprehension tasks and strategies that I perceived as very complex.	3.72	2.40	0	9
3. The test design task covered concepts and definitions on listening comprehension testing that I perceived as very complex.	3.29	2.47	0	10
Dimension: Extraneous load	1.96	2.09	0	10
4. The written instructions and explanations used during this listening test design task were very unclear.	2.01	2.36	0	10
5. The written instructions and explanations used during this listening test design task made it difficult to understand.	2.03	2.33	0	10
6. The written instructions and explanations during this listening test design task were, in terms of learning, very ineffective.	1.86	2.18	0	10
Dimension: Germane load	5.35	2.83	0	10
7. This test design task really enhanced my understanding of how to create a test for English listening.	5.21	2.98	0	10
8. This test design task really enhanced my understanding of both the strategies and tasks for testing English listening.	5.08	2.83	0	10
9. This test design task really enhanced my knowledge of how to write instructions for a listening test.	5.36	3.11	0	10
10. This test design task really enhanced my knowledge and understanding of the difference among multiple choice, true or false and full answer question items.	5.74	3.44	0	10

First, *intrinsic cognitive load* refers to the difficulty inherent in the task, regardless of the design or instructions. It is the intrinsic complexity of the learning material. On average, participants perceived a moderate intrinsic cognitive load (3.43 out of 10). This fact suggests that the material's content or the task was of intermediate difficulty for most pre-service teachers, requiring moderate mental effort to process. A high standard deviation (2.19) is evident, indicating that individual differences in prior

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

knowledge, cognitive skills, and learning strategies may have influenced the perception of difficulty. Some participants may have had more prior knowledge about the topic, facilitating comprehension, while others lacked this knowledge and experienced greater difficulties.

Extraneous cognitive load refers to the challenges that arise from the design of a learning task or material. This dimension includes factors such as the complexity of the instructions, the way information is presented, and the clarity of the objectives. Participants' responses to these elements resulted in a mean score of 1.96 (out of 10). The findings reveal that, on average, participants perceived a low extraneous cognitive load. This fact suggests that the design of the task or material did not impose notable additional difficulties. However, the relatively high standard deviation of 2.09 indicates considerable response variability. While most participants found the task straightforward, some perceived it as much more complex than others.

Germane cognitive load refers to the part of the task that promotes deep and meaningful learning. The results indicate a high perception of germane cognitive load (mean of 5.35 out of 10). Despite perceived difficulties in the other cognitive load dimensions, participants reported that the task helped them improve their understanding of key concepts, develop new skills, and apply their knowledge effectively. However, the large standard deviation (2.83) indicates that some subjects might have found the task enhanced their understanding and knowledge of test design, while others did not.

Table 2 exhibits the comparative results according to the dimension of cognitive load and level of professional training. The Kruskal-Wallis test was conducted to analyze differences among groups in the third, fourth, and fifth years of initial teacher training regarding three dimensions of cognitive load: *intrinsic*, *extraneous*, and *germane*.

Table 2.
Comparative results according to dimension and year of professional preparation.

Dimension	Year	N	Mean	SD	*p*
Intrinsic load	Third year	46	3.34	2.06	0.146
	Fourth year	58	3.80	2.30	
	Fifth year	21	2.67	2.08	
Extraneous load	Third year	46	2.28	2.18	0.092
	Fourth year	58	1.99	2.11	
	Fifth year	21	1.19	1.68	
Germane load	Third year	46	5.01	2.66	0.477
	Fourth year	58	5.68	2.85	
	Fifth year	21	5.21	3.19	

Analysis of the *intrinsic cognitive load* ($p = 0.146$) experienced by participants when creating an instrument to assess English listening comprehension found no statistically significant differences across various levels of professional training. This fact suggests that, regardless of the year of university preparation, pre-service EFL teachers perceive a similar complexity inherent in the tasks. The comparison of means by group establishes that fourth-year prospective teachers perceive a higher level of intrinsic cognitive load (mean = 3.80) compared to third (mean = 3.34) and fifth-year candidates (mean = 2.67), but it is moderate. In this context, fifth-year candidates experience a lower cognitive load when facing this assessment task.

As for the *extraneous cognitive load* ($p = 0.092$) experienced by prospective EFL teachers when designing an instrument to assess listening comprehension in English, although it does not reach a conventional significance level ($p < 0.05$), there is a tendency for fifth-year students to perceive a lower extraneous cognitive load (mean = 1.19) compared to third (mean = 2.28) and fourth-year (mean = 1.99) prospective teachers. This situation could indicate that, as pre-service teachers undergo more advanced professional training and practicum experiences, they become more efficient in managing external factors that hinder learning when designing assessment instruments.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

There were no statistically significant differences in *germane cognitive load* (p = 0.477) among different levels of professional training when student teachers designed an instrument for assessing listening comprehension in English. This reality indicates that prospective EFL teachers, regardless of their professional preparation levels, perceive a similar degree of meaningful learning and task relevance when designing language assessment instruments. As for the comparison of means, fourth-year candidates perceive the germane cognitive load as slightly higher (mean = 5.68) compared to third-year (mean = 5.01) and fifth-year candidates (mean = 5.21), which evidences that this group has a slightly more positive perception of the learning process related to assessing listening comprehension in English.

5. Discussion of Results

Based on the research findings, pre-service EFL teachers in their third, fourth, and fifth year of professional preparation create close and open-ended questions to evaluate students' listening comprehension skills, focusing on identifying specific information from the spoken texts. This evaluation task can help confirm how good learners are at developing listening comprehension strategies to identify specific information or details (Brown & Lee, 2015). From the point of view of Day and Park (2005), these evaluation practices may involve literal comprehension, which is the most basic level and encompasses explicit textual information.

In their fourth and fifth year of professional preparation, pre-service EFL teachers create close and open-ended questions to evaluate students' listening comprehension skills, focusing on identifying general and specific information from spoken texts. In light of the evaluation practices mentioned, these candidates incorporate the subskill of listening to identify general information that contributes to the overall understanding of oral texts (Zhao & Lee, 2022). According to Day and Park (2005), this type of comprehension is linked to higher levels of understanding, including inferential comprehension. In this context, language users form mental representations based on prior knowledge to derive meaning from the text. From this viewpoint, the evaluation practices of prospective teachers in their fourth and fifth years are linked to the fourth level of comprehension, prediction. Learners draw upon their prior knowledge and understanding of the text at this level to anticipate what will happen next (Day & Park, 2005; Flowerdew & Miller, 2014).

The research findings indicate that third-year pre-service teachers tend to design basic items for assessing listening comprehension skills. This trend occurs because they primarily focus on extracting specific information from spoken texts. Furthermore, there are inconsistencies in the comprehension strategies they claim to use for each assessment item. These issues may arise because the candidates are still in the process of learning the didactics of teaching a foreign language and need more opportunities to design and implement listening evaluation practices gradually during their practicum experiences (Tagle et al., 2024). The findings are consistent with a study by Tsagari and Vogt (2017), which asserts that teachers of English must possess sufficient theoretical knowledge to create effective assessment practices. Lacking this knowledge, they may rely on published assessment materials or replicate the assessment methods used by other educators without judging them.

To address the challenges that language assessment encompasses, it is recommended that pre-service language teacher education programs offer candidates opportunities to develop and practice various evaluation performances related to different comprehension items in their didactics and practicum courses (Farmasari et al., 2023; Giraldo & Murcia, 2019). Moreover, teacher candidates need to incorporate higher levels of comprehension when assessing listening skills. For example, they should create listening evaluations, where students critically analyze the spoken message or create a text expressing their feelings or personal experiences related to the content, focusing on information from oral texts. Within this framework, pre-service language teacher education should also promote guidelines for successful evaluation practices, which involve designing assessments that have easier scoring methods while adhering to appropriate time constraints (practicality); effectively measuring what the assessment is supposed to evaluate (validity); consistently and accurately measuring student's

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

language skill learning (reliability); designing language assessment tasks that are realistic (authenticity), and being conscious of the influence of assessment practices on students' learning (washback) (Brown & Abeywickrama, 2018; Plakans & Gebril, 2015).

Furthermore, prospective EFL teachers should have opportunities to reflect on the evaluation practices they design. Their reflection should focus on the theoretical principles of communicative approaches to language teaching and learning to enhance their listening evaluation practices in English. In such a context, reflection is necessary to move away from traditional evaluation paradigms, which fail to aid in language learning (Lőrincz, 2023; Rahman et al., 2018). Teacher education programs can encourage teacher candidates to reflect on their assessment practices during post-observation conversations, which can foster a dialogic environment that facilitates reflection and evaluation, enabling trainee teachers to feel empowered within the teaching profession (Turan & Yiğitoğlu Aptoula, 2023). Additionally, Richards and Farrell (2011) suggest some activities that facilitate teacher reflection, which include audio- or video-recording lessons, lesson written accounts, teaching journals, case reports, teaching portfolios, critical incidents, action research, or student-teacher support groups.

Based on the previously mentioned ideas, the goal of pre-service language teacher education in listening assessment is for teacher candidates to evaluate this skill using an interactive comprehension approach (Rost, 2013, 2016). This perspective highlights the importance of linguistic knowledge (bottom-up processing) and background knowledge (top-down processing) in the comprehension process, suggesting that both dimensions should be considered together in teaching and assessment. Moreover, from a constructivist perspective, listening assessment should incorporate comprehension strategies to help listeners become autonomous learners (Chamot, 2005; Oxford, 1990).

On the other hand, regarding the cognitive load experienced by the pre-service teachers when they designed items to evaluate listening comprehension skills in English, the participants perceived a moderate intrinsic cognitive load, suggesting that the material's content or the task was of moderate difficulty for most pre-service teachers, requiring moderate mental effort to process. Fourth-year students perceive intrinsic cognitive load as higher compared to third- and fifth-year students. However, the differences are not statistically significant. This suggests that, although there are variations, students generally perceive a similar level of task complexity, regardless of educational level.

Moreover, on average, participants perceived a low extraneous cognitive load when designing listening evaluation items. This fact suggests that the design of the task or material did not impose notable additional difficulties (Leppink et al., 2013). In this context, there is a tendency for fifth year candidates to perceive a lower extraneous cognitive load compared to third- and fourth-year prospective teachers. This fact could indicate that more advanced students develop more efficient strategies to manage external factors that hinder learning (Gillmor et al., 2015).

Despite perceived difficulties in the other cognitive load dimensions, all participants evidenced a high level of germane cognitive load. This happened as the pre-service teachers reported that the task helped them improve their understanding of key concepts, develop new skills, and apply their knowledge effectively. A high germane cognitive load suggests that, despite these difficulties, the task achieved its primary objective: to promote learning (Vandewaetere & Clarebout, 2013).

The comparison of cognitive load experienced by third, fourth, and fifth pre-service EFL teachers showed no statistically significant differences. The overall results determine stability in the perception of intrinsic and germane cognitive load when designing an instrument to assess listening comprehension skills in English. According to the interpretation of these results, teacher candidates, regardless of their level of professional preparation, perceive similarly the complexity inherent to the tasks and the level of meaningful learning they construct (Klepsch et al., 2017).

The results of this study also exhibit a tendency for extraneous cognitive load to decrease over time as participants progress through higher levels of professional training. In this regard, fifth-year students could be developing more effective strategies to manage external factors that hinder learning while designing an instrument to assess listening comprehension, reflected in a lower perception of extraneous cognitive load (Jordan et al., 2019).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

It is important to note that a high cognitive load is not always negative. In fact, a certain level of cognitive challenge is necessary to promote long-term learning. However, it is critical to balance the intrinsic and extraneous cognitive load and ensure that the germane cognitive load is high enough to promote meaningful learning (Leppink et al., 2013). Implementing cognitive load theory principles is crucial in preparing future foreign language teachers, especially as they learn to assess students in their practicum experiences. Assessment is a key competence of educational practice, and pre-service teachers feel this is a complicated area of study (see, for example, Fröjdendahl, 2018). Therefore, effective management of intrinsic, extraneous, and germane cognitive load theory within pre-service teacher programs can scaffold these future teachers in understanding pedagogical and disciplinary concepts, which would augment their professional knowledge and performance (Timothy et al., 2023).

6. Conclusions

This study analyzed pre-service EFL teachers' listening comprehension evaluation items and their cognitive load during the test design. Research findings indicate that pre-service teachers improve their ability to create assessment items, demonstrating a better focus on specific and general information than in their earlier years of professional training. Third-year teachers primarily concentrate on basic, specific information, while those in higher years include more complex levels of comprehension, such as inferential and predictive skills.

Participants exhibited moderate intrinsic cognitive load, demonstrating the task was reasonably challenging. The extraneous load was generally low, suggesting minimal external distractions. Notably, the germane load was high across all groups, indicating that the task effectively promoted learning and skill development.

These findings emphasize the necessity for pre-service teacher education programs to improve the development of listening evaluation skills. To achieve this, future teachers should have opportunities to design various evaluation tasks and instruments, including those that assess higher-order thinking skills such as critical analysis and personal response. Furthermore, pre-service EFL teacher education must encourage prospective teachers to critically examine their evaluation practices through reflection and identify areas for improvement.

Pre-service teacher education programs must address cognitive load effectively. This study suggests that teacher preparation should balance challenging tasks and the appropriate support needed to minimize extraneous cognitive load while maximizing germane load. This approach is particularly important for courses vital to the professional development of pre-service teachers, such as English language proficiency, didactics, and practicum. For this reason, it is recommended that teacher educators receive training on cognitive load theory to help future teachers develop profound learning.

Funding:

This manuscript was produced in the context of the research grant Fondecyt 1220307 entitled "Estudio sobre el diseño de instrumentos de evaluación del idioma inglés: procesos y carga cognitiva, respuesta afectiva y desempeños de candidatos a profesores".

References

[1] Ashoori Tootkaboni, A. (2019). Teachers' beliefs and practices towards communicative language teaching in the expanding circle. *Revista Signos*, *52*(100), 265-289. http://dx.doi.org/10.4067/S0718-09342019000200265

[2] Ayres, P. (2020). Something old, something new from cognitive load theory. *Computers in Human Behavior*, *113*, 1-5. https://doi.org/10.1016/j.chb.2020.106503

[3] Bachman, L., & Palmer, A. (2017). *Language assessment in practice*. Oxford University Press.

[4] Bae, M., & Lee, B. (2018). Effects of text length and question type on test-takers' performance on fill-in-the-blank items in Korean CSAT. *English Teaching*, *73*(4), 149-174. https://doi.org/10.15858/engtea.73.4.201812.149

[5] Bailey, K. M. (2020). *Teaching listening and speaking in second and foreign language contexts*. Bloomsbury Academic.

[6] British Council (2015). *Annual report 2014-15*. https://www.britishcouncil.org/sites/default/files/annual-report-2014-2015.pdf

[7] Britton, M. (2021). *Assessment for learning in primary language learning and teaching*. Multilingual Matters.

[8] Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy* (4th ed.). Pearson.

[9] Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed). Pearson.

[10] Burns, A., & Siegel, J. (2023). Teaching listening: Dichotomies, choices and practices. In E. Hinkel (Ed.), *Handbook of practical second language teaching and learning* (pp. 225-237). Routledge.

[11] Butt, G. (2010). *Making assessment matter*. Continuum.

[12] Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, *25*, 112-130.

[13] Chappuis, J., & Stiggins, R. (2020). *Classroom assessment for student learning: Doing it right - using it well* (3rd ed.). Pearson Education.

[14] Chen, I. J., & Chang, C. C. (2009). Teoría de carga cognitiva: Un estudio empírico sobre la ansiedad y el rendimiento en tareas de aprendizaje de idiomas. *Electronic Journal of Research in Educational Psychology*, *7*(18), 729-746. https://doi.org/10.25115/ejrep.v7i18.1369

[15] Chilean Ministry of Education (2021). *Estándares de la profesión docente: Carreras de pedagogía en inglés educación básica/media*. MINEDUC.

[16] Chilean Ministry of Education (2023). *Informe resultados nacionales: Evaluación nacional diagnóstica de la formación inicial docente*. MINEDUC.

[17] Chilean Ministry of Education (2024). *Informe Nacional Portafolio 2023*. MINEDUC.

[18] Chou, M. H. (2017). Impacts of the Test of English Listening Comprehension (TELC) on teachers and teaching in Taiwan. *Asian-Pacific Journal of Second and Foreign Language Education*, *2*(5), 1-14. https://doi.org/10.1186/s40862-017-0028-9

[19] Costley, J., Fanguy, M., Lange, C., & Baldwin, M. (2021). The effects of video lecture viewing strategies on cognitive load. *Journal of Computing in Higher Education*, *33*, 19-38. https://doi.org/10.1007/s12528-020-09254-y

[20] Cross, J. (2009). Effects of listening strategy instruction on new videotext comprehension. *Language Teaching Research*, *13*(2), 151-176.

[21] Day, R. R., & Park, J. S. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, *17*(1), 60-73.

[22] Earl, L. M. (2014). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd ed.). Corwin.

[23] Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Manitoba Education, Citizenship, and Youth, School Programs Division.

[24] Fan, J., Frost, K., & Baoquan, L. (2020). Teachers' involvement in high-stakes language assessment reforms: The case of Test for English Majors (TEM) in China. *Studies in Educational Evaluation*, *66*, 1-10. https://doi.org/10.1016/j.stueduc.2020.100898

[25] Farmasari, S., Wardana, L. A., Baharudin, B., Herayana, D., & Suryaningsih, H. (2023). Pre-service EFL teachers' language assessment literacy satisfaction and assessment preparedness. *IJoLE: International Journal of Language Education*, *7*(4), 646-660. https://doi.org/10.26858/ijole.v7i4.47445

[26] Fitriyah, I., & Jannah, M. (2021). Online assessment effect in EFL classroom: An investigation on students and teachers' perceptions. *Indonesian Journal of English Language Teaching and Applied Linguistics*, *5*(2), 265-284.

[27] Flowerdew, J., & Miller, L. (2014). Dimensions of academic listening. In M. Celce-Murcia, D. M. Brinton & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (pp. 90-103). Cengage Learning

[28] Forutan, A. (2014). Traditional versus alternative writing assessment. *International Journal of Foreign Language Teaching & Research*, *2*(7), 10-22.

[29] Fröjdendahl, B. (2018). Pre- and in-service teachers' assessment literacy: A qualitative approach. *Literacy Information and Computer Education Journal (LICEJ)*, *9*(2), 2886- 2894.

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

[31] Gillmor, S. C., Poggio, J., & Embretson, S. (2015). Effects of reducing the cognitive load of mathematics test items on student performance. Numeracy, *8*(1), 1-18. http://dx.doi.org/10.5038/1936-4660.8.1.4

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

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

[34] Goçtü, R. (2012). Comparison of traditional and portfolio assessment efficiency in English language teaching in high schools. *Journal of Education*, *1*(2), 43-52.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

[35] Goh, C. (1998). How ESL learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research*, *2*(2), 124 -147.

[36] Goh, C. C. M., & Aryadoust, V. (2015). Examining the notion of listening subskill divisibility and its implications for second language listening. *International Journal of Listening*, *29*(3), 109-133. https://doi.org/10.1080/10904018.2014.936119

[37] Gottlieb, M. (2016). *Assessing English language learners: Bridges to educational equity* (2nd ed.). Corwin.

[38] Güngör, M. A., & Güngör, M. N. (2024). Relocating assessment in pre-service teacher education: an emerging model from activity theory lens. *European Journal of Teacher Education*, *47*(2), 348-368. https://doi.org/10.1080/02619768.2024.2339522

[39] Hanham, J., Leahy, W., & Sweller, J. (2017). Cognitive load theory, element interactivity, and the testing and reverse testing effects. *Applied Cognitive Psychology*, *31*(3), 265-280. https://doi.org/10.1002/acp.3324

[40] Heritage, M. (2022). *Formative assessment: Making it happen in the classroom* (2nd ed.). Corwin.

[41] Hong Yang, H., & Farley, A. (2019). Quantifying the impact of language on the performance of international accounting students: A cognitive load theory perspective. *English for Specific Purposes*, *55*, 12-24. https://doi.org/10.1016/j.esp.2019.03.003

[42] Jones, J. (2010). The role of assessment for learning in the management of primary to secondary transition: Implications for language teachers. *Language Learning Journal*, *35*(2), 175-191. https://doi.org/10.1080/09571730902928052

[43] Jordan, J., Wagner, J., Manthey, D. E., Wolff, M., Santen, S., & Cico, S. J. (2019). Optimizing lectures from a cognitive load perspective. *AEM Education and Training*, *4*(3), 306-312. https://doi.org/10.1002/aet2.10389

[44] Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01997

[45] Leppink, J., Paas, F., Van der Vleuten C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*, 1058-1072. https://doi.org/10.3758/s13428-013-0334-1

[46] Lin, J. J. H., Lee, Y. H., Wang, D. Y., & Lin, S. J. (2016). Reading subtitles and taking enotes while learning scientific materials in a multimedia environment: Cognitive load perspectives on EFL students. *Educational Technology & Society*, *19*(4), 47-58.

[47] Liu, K. (2011). Reducing cognitive load in multimedia-based college English teaching. *Theory and Practice in Language Studies*, *1*(3), 306-308. https://doi.org/10.4304/tpls.1.3.306-308

[48] Lőrincz, M. (2023). EFL student teachers' beliefs about language teaching approaches and instructional practices. *Electronic Journal of Foreign Language Teaching*, *20*(2), 167-192.

[49] Marsh, C. (2010). *Becoming a teacher: Knowledge, skills and issues* (5th ed.). Pearson.

[50] Martin, S. (2014). Measuring cognitive load and cognition: Metrics for technology-enhanced learning. *Educational Research and Evaluation*, *20*(7-8), 592-621. https://doi.org/10.1080/13803611.2014.997140

[51] Mihai, F. M. (2010). *Assessing English language learners in the content area: A research-into-practice guide for educators*. The University of Michigan Press.

[52] Nawal, A. F. (2018). Cognitive load theory in the context of second language academic writing. *Higher Education Pedagogies*, *3*(1), 385-402. https://doi.org/10.1080/23752696.2018.1513812

[53] O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press.

[54] Ockey, G. J., & Wagner, E. (2018). Introduction. In G. J. Ockey & E. Wagner (Eds.), Assessing L2 listening: Moving towards authenticity (pp. 2-10). John Benjamin Publishing Company.

[55] Odacı, M. M., & Uzun, E. (2024). Measuring cognitive loads and attitudes of pre-service teachers in computer based testing environment. *Journal of Educational Technology & Online Learning*, *7*(1), 84-101.

[56] Oo, C. Z., Alonzo, D., & Asih, R. A. (2022). Acquisition of teacher assessment literacy by pre-service teachers: A review of practices and program designs. *Issues in Educational Research*, *32*(1), 352-373.

[57] Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Heinle Cengage Learning.

[58] Paas, F., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning: Theory, methods and applications. In D. H. Robinson & G. Schraw (Eds.), *Recent innovations in educational technology that facilitate student learning* (pp. 11-35). Information Age Publishing.

[59] Phongsirikul, M. (2018). Traditional and alternative assessments in ELT: Students' and teachers' perceptions. *REFLections*, *25*(1), 61-84.

[60] Plakans, L., & Gebril, A. (2015). *Assessment myths: Applying second language research to classroom teaching*. University of Michigan.

[61] Porto, M. (2018). Intercultural citizenship in foreign language education: An opportunity to broaden CLIL's theoretical outlook and pedagogy. *International Journal of Bilingual Education and Bilingualism*, *24*(7), 927-947. https://doi.org/10.1080/13670050.2018.1526886

[62] Prastikawati, E. F., Mujiyanto, J., Saleh, M., & Fitriati, S. W. (2021). Language assessment course: Impact on pre-service EFL teachers' language assessment literacy. *ISET: International Conference on Science, Education and Technology*, *7*(1), 160-167.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996

[63] Rahman, M. M., Singh, M. K. M., & Pandian, A. (2018). Exploring ESL teacher beliefs and classroom practices of CLT: A case study. *International Journal of Instruction*, *11*(1), 295-310. https://doi.org/10.12973/iji.2018.11121a

[64] Rapley, T. (2018). *Doing conversation, discourse and document analysis*. SAGE.

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

[66] Richards, J. C. (2008). *Teaching listening and speaking: From theory to practice*. Cambridge University Press.

[67] Richards, J. C., & Farrell, T. S. C. (2011). *Practice teaching: A reflective approach*. Cambridge University Press.

[68] Rojas, J. (2017). Making sense of alternative assessment in a qualitative evaluation system. *Profile: Issues in Teachers' Professional Development*, *19*(2), 73-85. https://doi.org/10.15446/profile.v19n2.57178

[69] Rost, M. (2013). *Listening in language learning*. Routledge.

[70] Rost, M. (2016). *Teaching and researching: Listening* (3rd ed.). Routledge.

[71] Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction*, *52*, 69-79. https://doi.org/10.1016/j.learninstruc.2017.04.007

[72] Sevilla Morales, H., & Chaves-Fernandez, L. (2019). Authentic assessment in the listening comprehension classroom: Benefits and implications. *GIST - Education and Learning Research Journal*, *19*, 6-30. https://doi.org/10.26817/16925777.704

[73] Siegel, J. (2014). Exploring L2 listening instruction: Examinations of practice. *ELT Journal*, *68*(1), 22-30. https://doi.org/10.1093/elt/cct058

[74] Solano-Flores, G. (2016). *Assessing English language learners: Theory and practice*. Routledge.

[75] Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, *22*, 123-138. https://doi.org/10.1007/s10648-010-9128-5

[76] Sweller, J. (2018). Measuring cognitive load. *Perspectives on Medical Education*, *7*, 1-2. https://doi.org/10.1007/s40037-017-0395-4

[77] Tagle, T., Etchegaray, P., Díaz, C., Ortiz, M., Quintana, M., & Ramos, L. (2024). How do in-service EFL teachers assess student language learning? Analysis of English assessment instruments used in Chilean secondary schools. *ELOPE: English Language Overseas Perspectives and Enquiries*, *21*(1), 111-127. https://doi.org/10.4312/elope.21.1.111-127

[78] Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, *29*, 21-36. https://doi.org/10.1017/S0267190509090035

[79] Thanh Nha, C., & Hong Dung, N. T. (2020). Listening learning strategies employed by English majors at the Saigon International University. *International Journal on Studies in English Language and Literature (IJSELL)*, *8*(7), 16-27.

[80] Timothy, V., Fischer, F., Watzka, B., Girwidz, R., & Stadler, M. (2023). Applying cognitive load theory in teacher education: an experimental validation of the scale by Leppink et al. *Psychological Test Adaptation and Development*, *4*, 246-256. https://doi.org/10.1027/2698-1866/a000052

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

[82] Turan, P., & Yiğitoğlu Aptoula, N. (2023). Between teacher candidates' reflection and teacher educators' evaluation: Fluctuations in epistemic (a)symmetry in feedback conversations. *Modern Language Journal*, *107*, 1011–1034. https://doi.org/10.1111/modl.12886

[83] Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147-177. https://doi.org/10.1007/s10648-005-3951-0

[84] Vandergrift, L. (2003). Orchestrating strategy use: Towards a model of the skilled L2 listener. *Language Learning*, *53*(3), 463-496.

[85] Vandewaetere, M., & Clarebout, G. (2013). Cognitive load of learner control: Extraneous or germane load? *Education Research International*, 902809. https://doi.org/10.1155/2013/902809

[86] Wubshet, H., & Menuta, F. (2015). Investigating the practice of alternative assessment in English classrooms: The case of selected grade nine English teachers' assessment practices. *International Journal of Scientific Research in Education*, *8*(4), 159-171.

[87] Zeng, J., & Huang, L. (2021). Understanding formative assessment practice in the EFL exam-oriented context: An application of the theory of planned behavior. *Frontiers in Psychology. 12*. https://doi.org/10.3389/fpsyg.2021.774159

[88] Zhao, J., & Lee, C. I. (2022). Teaching EFLLs listening subskills with a speaking-listening model in a computer-mediated communication setting. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.836013

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9328-9344, 2024
DOI: 10.55214/25768484.v8i6.3996