Revista de Lenguas Modernas, N.º 36, 2022 / 01-22
ISSN electrónico: 2215-5643
ISSN impreso: 1659-1933
DOI: 10.15517/RLM.V0I36.48313

Assessment of Young English-Language Learners: Formative and Summative Strategies

La Evaluación de la competencia lingüística de inglés en estudiantes de primaria: Estrategias formativas y sumativas

Allen Quesada Pacheco, Ph. D.
Universidad de Costa Rica
Escuela de Lenguas Modernas
allen.quesada@ucr.ac.cr

Abstract

This article provides a comprehensive overview of formative and summative language assessment of young language learners, including the steps involved in developing a diagnostic test of English for primary school children. The importance of determining the level of English competence in young learners is emphasized, and various assessment and teaching are explored. The article covers both formative assessment, which allows for ongoing feedback and progress monitoring, as well as summative assessment. The four essential language skills - listening, speaking, reading, and writing - are examined in detail, and strategies for assessing each skill are discussed. The article also emphasizes the need to use assessment results for diagnostic purposes to improve teaching and learning outcomes. This article is intended for language educators and policymakers seeking to improve language instruction and evaluation in primary education settings.

Keywords: summative language assessment, diagnostic test, young language learners, formative assessments, Essential language skills

Recepción: 22-09-21 Aceptación: 03-09-23

Resumen

Este artículo proporciona una descripción completa de las evaluaciones formativas y pruebas de idiomas sumativas para estudiantes de primaria, incluyendo los procesos en el desarrollo de un examen diagnóstico en inglés para la población de escuela primaria. Se destaca la importancia de determinar el nivel de competencia en inglés de los jóvenes aprendices, y se exploran diversas técnicas de evaluación en el aula. El artículo cubre tanto la evaluación formativa, que permite retroalimentación continua y monitoreo del progreso en el aula, como la evaluación sumativa. Se examinan detalladamente las cuatro habilidades esenciales del lenguaje: escuchar, hablar, leer y escribir, y se discuten estrategias para evaluar cada habilidad. También se enfatiza la necesidad de utilizar los resultados de la evaluación con fines diagnósticos para mejorar los resultados de la enseñanza y el aprendizaje. Por último, el artículo está dirigido a educadores de idiomas y responsables de políticas educativas que buscan mejorar la enseñanza y evaluación de idiomas extranjeros en entornos de educación primaria.

Palabras claves: evaluaciones sumativas de lenguas extranjeras, prueba diagnóstica, estudiantes de primaria de idiomas, evaluaciones formativas, habilidades esenciales del lenguaje

Assessment… can contribute to the children's sense of pride in their achievement, and thus motivate them to make further progress.
Alan Maley

Introduction

Assessing the language proficiency of young learners is critical in the evaluation of language instruction and education policies. To determine the level of English competence, assessment techniques such as formative and summative assessments have been widely adopted. Formative assessment provides ongoing feedback and progress monitoring for students, while summative assessment is used to evaluate the learning and performance of students at the end of a specific period, such as a course, unit, semester, or academic year (Bulut & Ertem, 2018; Cohen, 2020). The aim of this article is to provide a comprehensive overview of both formative and summative assessment for young language learners, including computer-based language testing and test development processes. The article first discusses various formative assessment techniques for the four essential language skills: listening, speaking, reading, and writing. The article then explores summative assessment for young language learners. The use of computer-based language testing is also discussed, including its benefits, limitations, and impact on language learning. Finally, the article is aimed at language educators and educational policy makers who seek to improve the teaching and assessment of foreign languages in primary education settings.

Literature Review

Formative assessment techniques for young learners

Shaaban (2001) discussed several assessment techniques that can be used to measure the abilities, progress, and achievement of students effectively and practically in various educational settings. One such technique involves non-verbal responses, where young learners are expected to complete simple tasks based on basic instructions provided by their teachers. According to Shaaban (2001), this type of assessment does not create stress for students, as most of the tasks are an extension of their regular classroom activities and are seen as natural by the students. Hands-on activities such as creating diagrams, drawings, and charts are some examples of tasks that can be used for this type of assessment. Teachers commonly use the Total Physical Response approach (Asher, 1988; Tannenbaum, 1996; Shaaban, 2001) to implement this assessment technique with young learners.

Another technique is oral interviews. Young learners are shown images, and they have to engage in questions and answers to give responses accordingly. The teacher applies elicitation techniques to determine the proficiency level of young learners through questioning. Rich data can be obtained through this strategy (Pierce & O'Malley, 1992; Shaaban, 2001). Role-playing is another assessment technique that becomes very enjoyable for young learners because they can participate in believable situations of their daily life or they can simulate activities taught in content-based contexts. Role-plays and

simulations are forms of experiential learning. Learners take on different roles, assuming a profile of a character or personality, and interact and participate in diverse and complex learning settings (Russell & Shepherd, 2010). Altun (2015) "has maintained that it is beneficial to apply Role Play (RP) in EFL classes since they lead learners to develop communicative skills and improve their conversational abilities" (as cited in Soto et al., 2018, p. 51). "Furthermore, this technique enables students to link vocabulary, practical knowledge and topics being learned in class" (Alabsi, 2016, as cited in Soto et al., 2018, p. 51).

Apart from nonverbal communication, interviews, and role-plays, written narratives are outstanding assessment and teaching tools. A recommended activity that fosters the production of written narratives in young learners is creating storybooks; brief stories written by these EFL/ESL learners that can ignite their imagination and creativity. According to Wright (2002), utilizing storybooks is the most appropriate activity for young learner language teaching programs since stories are motivating and suitable for their cognitive level. Stories provide an authentic contextual framework that introduces children to vocabulary and language structures, and through stories, children develop literacy skills that aid them later in listening, reading, and writing. Shaaban (2001) has also noted that other genuine tasks such as writing letters to friends or TV program characters, having pen pals and writing to them, maintaining a personal diary, and writing everyday experiences can be highly effective assessment tools for narrative development.

Dialogue journals are helpful formative assessment techniques. A dialogue journal is a kind of notepad exchange where students and teachers write letters back and forth to each other over a period. This type of communication provides a meaningful bond or tie between teachers and students by inspiring young learners to exchange ideas, share topics of interest and beliefs, among others, providing levels of self-confidence. Shaaban (2001) has highlighted that dialogue journals are effective, enjoyable, and interactive in nature, regardless of the level of proficiency of young learners. Teachers can also use journals "to collect information on students' views, beliefs, attitudes, and motivation related to a class or program or to the process involved in learning various language skills" (Brown, 1998, p. 4).

Self-assessment is another formative assessment strategy. It allows young learners to reflect on their learning and express their feelings about their learning. Butler & Lee (2010) state that there are few empirical studies related to the use of self-assessment strategies among pre-school and elementary school children arguing that these young learners are not capable of self-evaluating their own performance. However, other studies have proven that children age 8 to 12 years old improve in self-assessing their performance, especially if they are assessing classroom tasks. McNamara and Deane (1995) have described that self-judgement helps young learners detect their strengths and weaknesses, (cited in Shaaban, 2001). Some of the activities that can be used for self-assessment are K-W-L charts (What I Know, What I Want to Know, What

I Learned) and learning logs. K-W-L charts activate the background knowledge of young learners regarding their interests, needs, likes, or dislikes; this graphic organizer empowers students to participate and help them track or monitor their own learning. Learning logs, on the other hand, maximize reflection among young learners. In the log, students record small chunks of information through writing, which makes them better thinkers. Students clarify questions which arise; they explain and share knowledge with peers.

Concerning peer evaluation, Shabaan (2001) has added that peer and group assessment is an excellent way to collaborate in academic performance, particularly in young learners. Spiller (2009) has claimed that peer assessment is a mutual process between students. The participation of students in commenting on the work of others increases their capacity for making intellectual choices and judgments, as well as the students receiving feedback from their peers helps them acquire a wide range of ideas about their work to promote and achieve development and improvement in their learning (p. 160, cited in Mohammed, 2017). In other words, peer assessment encourages collaborative learning and swapping ideas. It boosts confidence of students, encourages healthy discussions, and helps them develop their communication skills.

Without doubt, student portfolios are an excellent assessment technique, too. "A portfolio assessment is an individual collection of daily drawings, photographs, writings samples, audiotapes, video recordings, and other materials that provide visual and/or auditory documentation of a child's strengths" (Smith et al., 2003, p. 1). Portfolio assessment offers a variety of benefits, including: a) a means for recording children's ongoing development at different time periods in a school year by keeping, for example a file of a series of self-portraits; (b) rich information of each young learner to design curricula and instruction; (c) young learner´s involvement in their own work by reflecting in his/her strengths and weaknesses or by comparing their growth or progress from past experiences; d) a method of communication among the education community (teachers, parents, other students or peers, outside observers) that illustrate efforts, progress, achievement, doubts, among others (Cohen, 2020). It is important to highlight that portfolios do not compare children to other children. Instead, they illustrate the child's best work, building confidence and self-esteem.

Assessing the four skills in young language learners

Strategies for assessing listening

Learning how to listen can teach students how to communicate their ideas. The teaching and assessment of listening is of paramount importance in the overall evaluation of our learners' communicative ability. The following strategies allow teachers to assess listening skills to help students gain a better understanding of the skills themselves, how they are assessed appropriately, and what decisions should be made to improve them (Curtain, & Dahlberg, 2016). According to Curtain and Dahlberg (2016), there are four

basic types of classifications of listening skills: intensive, responsive, selective, and extensive. Assessments are designed to cater to these four categories accordingly.

Intensive listening:

Intensive listening assesses phonological and morphological elements of language. Example tasks include minimal phonemic and morphological pair recognition, past-tense markers as well as stressed and unstressed syllables. Students must listen carefully for components in a small string of language, such as phonemes, intonation, and discourse markers. Another activity to assess intensive listening is paraphrase recognition which is a type of listening task that assesses the student's ability to listen to a short piece of language and paraphrase it. Students listen to words, phrases, and sentences; then they are asked to choose the correct paraphrase from several choices (Buck, 2001; Ashcraft & Tran, 2010; Bulut & Ertem, 2018).

Responsive listening:

Responsive listening is another task used in listening assessments. This task is more authentic and more than likely used in an everyday setting inside or outside the classroom. It allows students to perform in a normal everyday English setting as well as teaches them functional tasks (i.e., asking for directions). Responsive listening can have specific questions or open-ended questions. Student's responses are measured by how accurately they answer questions. Students can speak and write their reply in open-ended responsive tasks (Buck, 2001; Ashcraft & Tran, 2010; Bulut & Ertem, 2018).

Selective listening:

The third listening strategy, selective listening, is when a student listens to a piece of information and must discriminate specific information. A listening cloze task is a popular assessment that requires the student to listen to a story, monologue, or conversation. Students see a transcript of the passage they are listening to and must fill in the missing information (deleted words or phrases). Students must filter out information that is irrelevant and retain the relevant information. Another example of selective listening is information transfer which is a technique that presents aural information that must be transferred to a visual representation such as a chart or a diagram. Picture-cued items are normally used for beginning ESL students. This assessment requires the student to actively listen, filter relevant information, and write the information where appropriate (Buck, 2001; Ashcraft & Tran, 2010; Bulut & Ertem, 2018).

Extensive listening:

Extensive listening tasks focus on macro-skills. These tasks are used for advanced English language learners. Extensive listening tasks include lectures, long conversations, and lengthy messages that require listeners to decipher information. Dictation is a widely researched genre used for assessing listening comprehension. Students taking this test listen to a passage of about 50 to 100 words in length three times. Students write down what they heard, which requires good listening as well as writing skills. Dictation provides a reasonable method of integrating

listening and writing skills implied in short passages. A more authentic example of extensive listening is a dialogue followed by multiple-choice comprehension items. The test-taker listens to a monologue or conversation and then is asked to answer a set of comprehension questions. In short, extensive listening includes listening to lengthy lectures or conversations to get a general idea of something. Listening for the main idea, details and making inferences are part of effective listening (Buck, 2001; Ashcraft & Tran, 2010; Bulut & Ertem, 2018).

Strategies for assessing speaking

Speaking is often one of the areas of learning English that is not usually assessed, especially in a classroom setting. However, speaking is an important skill for students to develop, and there is a need to include activities in the English classroom that provide opportunities for students to speak in English. These could be telling a story, a role play, an interview, or a discussion. Assessing speaking activities can tell you about your students' progress in English, what they have learned, how confidently they can speak in English, or whether they are having problems speaking English. There are five types of strategies to assess speaking skills.

Intensive Speaking:

According to Brown and Abeywickrama (2010), intensive speaking involves producing a limited amount of language in a highly controlled context. One example of intensive speaking is "a read aloud task." In this activity, the teacher listens to a recording and evaluates students in a series of phonological factors and fluency. Some variations of this task are reading a scripted dialogue with someone else, reading sentences containing minimal pairs, reading information from a chart (Bachman & Palmer, 2010). In Reading-Aloud Tasks, the test-taker's oral production is controlled to assess prosodic stress and intonation among other oral skills (Bachman & Palmer, 2010).

Another way to assess intensive speaking is through a "sentence/dialogue completion task". Students are expected to read through the dialogue, so they can think about proper lines to fill in. The teacher produces one part orally and the students responds verbally. A third example is Pictured-Cued tasks which are one of the most popular ways to elicit oral language performance across proficiency levels. These tasks require a description from the test taker. Tasks are cued and the student will demonstrate their linguistic ability. Pictures can be very simple or more elaborate such as telling a story or event. This assessment can be customized and created to cater to teacher/student needs (Darmuki et al., 2017; Luoma, 2004).

Responsive Speaking:

According to Brown (2004), assessment of responsive tasks involves brief interactions with an interlocutor, differing from intensive tasks in the increased creativity given to the test-taker and from interactive tasks by the somewhat limited length of utterances. This aspect of assessment helps the teacher realize the student's ability to participate in discussions (Brown & Abeywickrama, 2010). Re-

sponsive speaking assessment is usually one on one (student and teacher) but may include other students. One way to assess responsive speaking is through question-and-answer tasks where students respond to questions that the test administrator asks; another way is through giving instructions and directions. Through this activity the test-taker is asked to give directions or instructions; another way is paraphrasing which requires the test-taker to paraphrase in two or three sentences what he heard or read. Questions should be impromptu and therefore will bring authentic and unrehearsed responses (Darmuki et al., 2017; Luoma, 2004).

Interactive Speaking:

Interactive speaking is a language performance that involves tasks requiring longer and sustained interaction with others, including interviews, role-plays, discussions, and games (Luoma, 2004). While responsive speaking involves shorter interactions, interactive speaking requires test-takers to engage in extended conversations with others, demonstrating their ability to sustain a conversation, negotiate meaning, and express opinions and ideas effectively (Darmuki et al., 2017).

One common method of assessing interactive speaking is through oral interviews, which involve face-to-face exchanges between test administrators and test-takers. The interview typically consists of four stages: a warm-up, a level check, probes, and a wind-down (Luoma, 2004). During these stages, the administrator may use prompts that are authentic and mimic real-world situations, such as those encountered in a restaurant or when giving directions. Another effective method of assessing interactive speaking is through role-plays. Role-plays are commonly used in communicative English classes as pedagogical activities (Luoma, 2004). Role-plays offer test-takers the opportunity to use language in a context that is difficult to elicit in other ways. The administrator can provide prompts that simulate real-world situations, allowing test-takers to demonstrate their ability to interact, negotiate, and problem-solve in English.

In addition to role-plays and oral interviews, discussions and conversations provide a level of authenticity and spontaneity in assessing interactive speaking (Darmuki et al., 2017). These tasks enable test-takers to engage in natural, free-flowing conversations, providing a more accurate assessment of their ability to use English for interactive communication.

Finally, games can also be used as an informal assessment task for interactive speaking, although they are not commonly used in testing contexts (Darmuki et al., 2017; Luoma, 2004). Test administrators should consider their assessment objectives and scoring criteria when selecting assessment tasks for interactive speaking. Indeed, interactive speaking is an important aspect of language proficiency that requires sustained interaction with others. Assessing interactive speaking can be accomplished through a variety of tasks, including oral interviews, role-plays, discussions, conversations, and games. Test administrators should select tasks that are appropriate for their assessment objectives and scoring criteria.

Extensive Speaking:

Extensive speaking is one of the most difficult aspects of speaking. This task involves complex, relatively lengthy types of discourse. For instance, oral presentations are used as an authentic life like assessment. It is common for individuals to present a brief report, a sales idea, or new product at some point in their life. Oral presentations allow students to use what they have learned in English by culminating everything in one presentation. A checklist or rubric is a common means of scoring and evaluation based on content and delivery. Another activity used for the assessment of extensive speaking is picture-cued storytelling, which requires students to describe a story based on a series of pictures that they have previously seen. The objectives of this task incorporate listening comprehension of original reading to production of oral discourse. Students are scored based on accuracy of vital information such as event order, fluency, and pronunciation (Darmuki et al., 2017; Luoma, 2004).

Imitative Speaking:

Imitative speaking tasks are based on repetition. Students just need to repeat a sentence they hear. This assessment focuses on the phonetic level of oral production (example, pronunciation), not meaning, and requires listening just for the prompt. This type of assessment helps teachers assess students' pronunciation skills. Examples include directed response tasks, reading aloud, sentence and dialogue completion and limited picture-cued tasks (Darmuki et al., 2017; Luoma, 2004). Similarly, Brown and Abeywickrama (2010) have explained that when students are involved in imitative speaking, communicative competence is not involved in the assessment criteria. The focus of this strategy is pronunciation which is a subskill of speaking.

Strategies for assessing reading

Reading assessment helps teachers understand the strengths and needs of each language learner. Although all reading assessments should share this purpose, the way individual assessments provide information and how teachers use the particular assessment information are varied. For this reason, it is important to be aware of the different types of readings students will be involved in to be able to assess them. Assessing this skill addresses perceptive, selective, interactive, and extensive reading.

Perceptive reading

Perceptive reading tasks involve attending to the components of larger pieces of discourse, such as letters, words, punctuation, and other graphemic symbols, implying a bottom-up processing approach. In the beginning stages of reading a second language, fundamental tasks include recognizing alphabetic symbols, capitalized and lowercase letters, punctuation, words, and grapheme-phoneme correspondences (Rumelhart, 1977; Schank & Abelson, 1977).

Reading aloud is one of the assessments used to measure a student's literacy, as it is a reading comprehension and oral production task. For instance, students can read letters, words, and/

or sentences separately and sequentially. The teacher can select a story appropriate for the student's proficiency level, and multiple-choice responses can be used to evaluate ESL literacy skills, such as minimal pair distinction tasks and grapheme recognition tasks (Afflerbach, 2012; Grabe & Stoller, 2011).

Selective Reading:

Selective reading focuses on the lexical and grammatical aspects of language. A common activity to test vocabulary and reading knowledge is multiple-choice items, which are easily administered and scored, and serve the purpose of vocabulary and/or grammar check. An assessment that requires both reading and writing performance is gap-filling tasks. A simple gap-filling task is to create sentence completion items where test-takers read part of a sentence and complete it by writing a phrase. The administrator will have to use their judgment on what comprises a correct response. Correct responses will reflect reading comprehension of the first part of the sentence. A combination of bottom-up and top-down processing may be both used to assess lexical and grammatical aspects of students' reading ability. Assessment activities can include items such as multiple-choice (form-focused criteria), matching tasks, editing tasks, picture-cued tasks and gap-filling tasks (Afflerbach, 2012).

Interactive Reading:

Interactive reading uses personal experience and prior knowledge of young learners to engage these readers more fully when reading a given text. This reading approach enables children to be consistently challenged, but also encourages them to use what they already know.

One way of offering interactive reading instruction is through guided reading. By arranging students into small reading groups, the teacher can encourage readers to engage with the text actively. This approach involves a series of strategies and techniques that help readers to connect with the material, ask questions, make predictions, and reflect on what they are about to read. By using this group strategy, children or young are involved in peer negotiation for meaning that will assist them in the comprehension of the text. This model highlights four important aspects: actively engaging with the text, checking to understand, using context to focus, and constructing meaning. This way, the interactive reading model emphasizes the importance of readers' knowledge, elaboration, monitoring, and situational context in comprehending what they read (Rivera, 2022).

One assessment for interactive reading is cloze tasks, which is one of the most popular types of reading assessments. They can be created relatively easy and customized for any student. Other assessment tasks include cloze tasks, multiple choice tasks for reading comprehension, short-answer questions, editing tasks, scanning, ordering tasks, non-verbal tasks for information transfer such as charts, maps, graphs, and diagrams (Afflerbach, 2012).

Extensive Reading:

Extensive reading involves longer complex texts such as journal articles, essays, technical reports, professional

articles, short stories, and books. Global understanding is the goal for assessment. Top-down processing is assumed for most extensive tasks. Skimming tasks are used to get the main ideas. Note-taking and outlining are both used frequently for higher-ordered learning. A common method of assessing extensive reading is asking students to write a summary of a text. Students receive an appropriate but complex text to read and then summarize. Students are to accurately determine the main idea and supporting details. Their summaries should be written in their own words and organized accordingly. Other tasks like short-answers, editing, scanning, ordering, and information transfer tasks can also be used to assess extensive reading (Afflerbach, 2012).

Strategies for assessing writing

Assessment of student writing is a process. Assessment of student writing and performance in the class should occur at many different stages throughout the course and could come in many different forms. One of the major purposes of writing assessment is to provide feedback to students. This feedback is crucial to writing development. The following strategies refer to ways in which writing can be assessed.

Imitative Writing

Basic tasks such as writing the alphabet, individual words and very short sentences are some types of imitative writing tasks. Imitative writing requires students to demonstrate skills in the fundamental tasks of writing letters, spelling words, and placing punctuation marks correctly and constructing very brief sentences (Brown, 2004). A good example of imitative writing would be a spelling test. Writing out numbers is also a great way to test students' imitative writing ability. Tests like these may lack authenticity but at the imitative stage, form is the primary focus. Beginning-level language learners need basic training in and assessment of imitative writing. Indeed, imitative writing includes the rudiments of forming letters, words and simple sentences, and these levels of learners can benefit from these easy tasks to more complex ones such as detecting phoneme-grapheme correspondences (Ketabi & Somaye, 2015; Mckay, 2008).

Intensive Writing

According to Brown (2004), form-focused, guided writing, and grammar writing are other terms used for intensive writing. Intensive writing is characterized by requiring students to demonstrate their ability to produce appropriate vocabulary within a given context, use of collocations and idioms, and correct grammatical features up to the length of a sentence. While some argue that tasks involving grammatical transformation, such as combining two sentences into one using a relative pronoun, lack meaningful value, others suggest that it is a good method for students to memorize and apply grammar rules. An alternative approach to controlling the responses students create while still enabling them to work with the grammatical and syntactical aspects of language is to use picture-cued story sequences.

At this level, writing is considered display writing, and ordering tasks at

the sentence level are particularly appealing to students who enjoy word games and puzzles. For instance, students are asked to put a scrambled set of words into a coherent sentence or to reorder them correctly. This task requires not only writing skills, but also logical reasoning and background knowledge in order to properly order the sentences, and it taps into the rules of grammatical word ordering (Ketabi & Somaye, 2015; McKay, 2008).

Responsive Writing

In this type of writing, young EFL/ESL students learn about sentence-level grammar and are more concerned about discourse. Form is important at discourse level, and meaning and context are emphasized. Brief descriptions, short reports, summaries, and interpretation of charts and graphs are examples of responsive writing tasks. A guided question and answer activity is a lower-order task. Paraphrasing is also a good example of responsive writing because it gets students to use their own words while offering variety in their expression. In addition, guided writing activities provide a list of criteria for students to use while they construct their first paragraphs in the second language. Guided writing texts of students may be as long as two to three paragraphs (Ketabi & Somaye, 2015; Mckay, 2008).

Extensive Writing

Extensive writing requires students to achieve a purpose, organize and develop ideas logically, use details to support or illustrate ideas, demonstrate syntactic and lexical variety, and engage in the process of multiple drafts to achieve a final product up to the length of an essay, term paper, major research project report, or thesis (Brown 2004).

Having students write essays as a response to a book, lecture, or video is one way to assess students' extensive writing abilities. Students' responses will reflect the message (meaning) of the original text through supporting details, expressing their opinions, conforming to the expected length of paper and taking a stance that either defends or supports their opinion effectively. By using this type of written assessment, students build communication skills and improve reading comprehension. Journaling allows students to experiment with a variety of writing skills and genres. Teachers give appropriate topics that agree with proficiency level and establish guidelines for what the response to literature should entail (Ketabi & Somaye, 2015; Mckay, 2008).

Formative assessment and summative assessment for young language learners

Many educational institutions all over the world are beginning to use standardized testing for young learners (Mostafa, 2019). Special care should be considered when developing the standardized instrument based on the premise those tasks must be appropriate to young learners' level of cognitive development (Garton & Copland, 2019). Stevens and DeBord (2001) have stated that:

> an assessment system should include a variety of instruments for various categories or purposes.

> Clarifying the main purpose of the assessment, determining what should be measured, establishing procedures for data collection, and selecting data sources (child work, standardized tests, teacher report, parent report) are all components in an assessment process. (p. 2)

According to Katz (1997) and Kagan et al. (1998), assessment of individual young learners is currently used to determine progression on meaningful developmental achievements, to place or promote, to detect special needs, learning, and teaching problems, to assist with curriculum and instruction decisions, to help a child assess his or her own progress, to boost learning, to evaluate programs, to monitor trends, and for high-stakes accountability (cited in Garton, S. & Copland, F., 2019).

Kagan et al. (1998) have highlighted that there are common principles that can guide assessment policies and practices of young children. First, assessments should benefit children by improving the quality of educational programs or in providing direct services to children. Added to this, the purpose for assessments should be specific and should provide fairness, reliability, and validity (Wolf & Butler, 2017). Reliability and validity in young children's assessments must increase with children's age, and the data collection method selected should also be age appropriate (National Education Goals Panel, 1998). Assessment involves collecting evidence and making judgements or forming opinions about learners' knowledge skills and abilities (Garton, S., & Copland, F., 2019). It often also involves keeping an informal or formal record of those judgements. It is a key professional responsibility of all teachers to become effective at assessment.

Tsagari et al. (2018) have mentioned that teachers have two main purposes for assessing learners in their classes. One purpose is to improve learning by checking that learners are progressing (assessment for learning). They do this so that they can decide whether to give additional help, try a different explanation or use different materials when learners find things difficult, or whether to provide more challenging activities when learners are ready for these. The other purpose is to judge how successful learners have been in mastering the content of a course to report this to parents, school management or educational authorities (assessment of learning). This usually involves deciding on grades or scores. The former of these purposes is called formative assessment or assessment for learning. The latter is called summative assessment or assessment of learning.

In short, formative assessment and summative assessment serve distinct roles in the education system. Formative assessment is a powerful way and an ongoing process aimed at supporting and enhancing student learning. It involves providing timely feedback to students and teachers throughout the learning journey. Teachers use formative assessments to gauge students' understanding, identify areas of improvement, and adjust instructional strategies accordingly. These assessments can take various forms, such as quizzes, discussions, and observations, and are usually non-graded, creating a low-pressure environment that encourages active learning (Green et al., 2022).

On the other hand, summative assessment occurs at the end of a specific learning period and is designed to evaluate overall learning outcomes and achievement. Unlike formative assessments, summative assessments are more formal and standardized, often involving major exams or projects. The focus is on determining how well students have mastered the material and met the learning objectives. While they provide a summary of students' achievements, the feedback in summative assessments is limited to final grades or scores, which may have significant implications for students' academic progress and future opportunities (Green et al., 2022). In other words, formative assessment is for learning, while summative assessment is of learning. Finally, assessment, if done correctly, can provide a common ground between educators and parents or families to use in collaborating on a strategy to support the children (Tsagari et al., 2018).

Computer-based language assessment for young language learners

According with Chapelle and Douglas (2006), with the advancement of technology as a powerful mechanism to transform education, the use of computer technology, in the field of language assessment and testing, has been widely contemplated since the advent of CALL (Computer Assisted Language Learning) and CALT (Computer Assisted Language Testing).

CALT is an integrated procedure in which language performance is elicited and assessed with the help of a computer (Noijons, 1994). CALT encompasses computer-adaptive testing (CAT), the use of multimedia in language test tasks, and automatic response analysis (Chapelle & Douglas, 2006). Chapelle (2010) has distinguished three main motives for using technology in language testing: efficiency, equivalence, and innovation. Efficiency is attained through computer adaptive testing and analysis-based assessment, employing automated writing evaluation (AWE) or automated speech evaluation (ASE) systems. Equivalence refers to research aimed at ensuring computerized tests are comparable to the gold standard in language testing, which is paper and pencil tests. Innovation implies leveraging technology to genuinely transform language testing (cited in Sulaiman & Khan, 2019).

One important advancement concerning test item development is what is called Technology-enhanced items (TEIs). TEIs are assessment items (questions) that utilize technology other than multiple choice items to improve the interaction with the item beyond what is possible with paper. Technology enhanced items can improve examinee engagement, assess complex concepts with higher fidelity, improve precision/reliability, and enhance face validity. The goal is to improve assessment, by increasing things like reliability/precision, validity, and fidelity.

Technology-enhanced items (TEIs), if they are well-designed, can provide advantages over conventional multiple choice and constructed response items (Boyle & Hutchinson, 2009; Jodoin, 2003; Kane, 2006; Parshall et al., 2010; Tarrant et al., 2006). TEIs can broaden construct measurement; present more authentic contexts for the demonstration of skills and knowledge, reduce the effects of random guessing;

reduce construct irrelevance; increase measurement opportunities; facilitate time- and cost-efficient scoring of constructed responses; and improve test-taker motivation through greater engagement (Bryant, 2017). That is why developing a test for young learners with TEIs is a must to engage students and help them feel at ease. Table 1 presents a description of each of the technology-enhanced item types that can be used.

Table 1
Technology-enhanced item (TEI) types

TEI types	Description
Drop-down	Examinee chooses the correct answer from a drop-down list of options.
Drag-and-drop	Examinee selects and drags a label, an image or text to a predetermined drop-zone in the response area (an image, area of text, or label area).
Drag the words	Examinee drags and drops words from the bank to the corresponding blank fields.
Matching	Given a word(s), sentence, number(s), and/or object from the right the column, the examinee clicks to match the appropriate corresponding word(s), sentence, number(s), or object from the left column
Ordering/Sequencing	Examinee orders elements by dragging them into the correct order, for example, chronologically or smallest to larges
Hot spot	Examinee clicks areas on an image. Examinee selects a single answer or multiple answers.

Technology-enhanced assessment is a valuable teaching and learning tool that provides primary school students with an engaging way to interact with class material. Furthermore, it encourages the development of digital literacy skills by requiring students to drag and drop answers, highlight relevant data, complete sentences in a drop-down menu, and other digital activities. By providing this type of interactive experience during tests, it prepares students for life after school as they will be adept at navigating technology tools in their future academic endeavors.

Test development for young language learners

There are important steps and procedures that should be followed for the design and development of high-quality standardized tests (See table 2 on a test development process).

Defining objectives and the purpose of the test is crucial. It is important to have clarity on what to measure (construct) by identifying the skills or knowledge to be included. Once a decision is made, test developers should ask some fundamental questions: Who will take the test and for what purpose? What skills and/or areas of knowledge should be tested? How should test takers be able to use their knowledge? What kinds of questions should be included? How many of each kind? How long should the test be? And how difficult should the test be? (Flutcher & Davidson, 2007, Flutcher & Davidson, 2017; ALTE, 2011; Cambridge ESOL, 2011). There is also a need to know the cognitive and emotional characteristics of young learners in order to develop the testing conditions (Wolf & Butler, 2017; Garton & Copland, 2019).

The second step is to organize and create the item development committee. This group is in charge of the item specifications and item development which may include defining test objectives and specifications, helping ensure test questions are unbiased, determining test format (e.g., multiple-choice, drag and drop, sequencing, matching, select from a list, fill in the blanks, short answer, etc.), reviewing test questions, picture vocabulary, picture naming, sound and letter word identification, or test items, and writing test questions (Flutcher & Davidson, 2007, Flutcher & Davidson, 2017; ALTE, 2011; Cambridge ESOL, 2011).

The third one deals with writing and reviewing questions. That is, each test question is reviewed and revised to determine clarity as well as to make sure that items have only one correct answer among the options provided on the test based on the rules and specifications of the test. Scoring guides for open-ended responses, such as short written answers, and oral responses go through similar reviews (Flutcher & Davidson, 2007, Flutcher & Davidson, 2017; ALTE, 2011; Cambridge ESOL, 2011).

Steps two and three are closely related to task characteristics. Bachman (1990, cited in Befhrouz & Nahvi, 2013) provides a framework of task characteristics which include a set of features that describe five aspects of tasks: setting, test rubrics, input, expected response, and relationship between input and response. Setting refers to the physical condition under which testing takes place. Test rubric includes those features that show how the test takers should proceed during the test to accomplish the tasks. The characteristics of rubric include the organization (structure) of the test, instructions, the duration of the test as a whole and of the individual parts, and how the language that is used is evaluated and scored. Input consists of the material contained in a given test task which the test takers are going to process in some way and to which they are expected to respond. The type of input may be either an item or a prompt. The purpose of an item is to elicit either a selected or a limited response. An example of a test item is the familiar multiple-choice question or picture identification. The purpose of a prompt is to elicit an extended production response. Regarding the expected response, response characteristics can be described as closed ended, limited, and open ended. Depending on the input, test takers will be engaged in any of these responses.

The next step is to validate the test; in other words, the items and questions

are pretested with a sample group or pilot group, like the population to be tested. The results enable test developers to determine the difficulty, ambiguity, revision elimination or replacement of each question. The final step in a sound construction of a test consists of assembling the test. Each reviewer answers all questions independently and submits a list of correct answers to the test developers. Any discrepancies are resolved before the test is published. Even after the test has been administered, statisticians and test developers must make sure that test questions are working as intended. Before final scoring takes place, each question must undergo preliminary statistical analysis (Flutcher & Davidson, 2007, Flutcher & Davidson 2017; ALTE, 2011; Cambridge ESOL, 2011).

In a nutshell, young language learner's assessment should aim at making a language test a motivating and enjoyable experience for students. Such tasks and questions should ensure that the act of test experience is stress-free and engaging according to the age group.

Table 2
Test development process

	Groups involved	**Input gathering/ considerations**	**Outcomes**
Planning	Students, teachers, schools, administrators, English experts, educational institutions,	Surveys, interviews, workshops, training sessions	test requirements
Design	test management team, item writers, testing experts, IT researchers, psychometricians	Test construct, test usefulness, technical characteristics, procedural/logistics matters, standards (CEFR)	table of specification
Trials/Piloting	pretest and research analysis, reviews, adaptation item, elimination	test data	improved test construction/design
Stakeholders	test takers, teachers students, schools administrators, English experts, educational institutions, employers, Ministry of Education	Mock tests, test website, result explanations	Evidence of achievement Monitoring Advancement. Diagnostics. Teacher support

In a nutshell, young language learner's assessment should aim at making a language test a motivating and enjoyable experience for students. Such tasks and questions should ensure that the act of test experience is stress-free and engaging according to the age group.

Conclusion

As many teachers and students are working in blended or learning environments for the first time due to COVID-19, the role of formative and summative assessments has become more important than ever. By understanding exactly what their students know before and during instruction, English language teachers have much more power to improve student mastery of the foreign language than if they find out after a lesson or unit is complete. Formative assessment provides timely and actionable feedback to inform instructional decisions and improve student learning outcomes. Surely, summative assessment for young learners is also a tool that can be used to gather and provide educators, parents, and families with critical information about a young learner's language development and growth in the second language.

For this reason, the strategies provided above for assessing the four language skills offer feedback on the proficiency of young language learners. This information is a valuable source of input for teachers, who can monitor students' progress and record their strengths and weaknesses. There is a risk, however, that the types of tasks and tests may not be the best in terms of motivating and stimulating young learners. That is, they might be cognitively beyond young learners or that those tasks could be boring for young language learners affecting their enjoyment of learning English. However, it is a reality that the use of tests and assessments as instruments of education policy and practice is growing, mainly to identify learning differences among students or to inform pedagogical and instructional planning.

Therefore, language assessment has an important role to play in revealing young learners' development of language growth and mastery, and ways of interacting with and understanding the new culture so that English teachers can choose a pedagogical approach and curricular materials that will support young language learners' further language mastery and learning. There is a need to continue researching on young language learners (YLLs) assessment especially because of the large numbers of primary school children being tested worldwide and the stake holders making decisions as a result of the evidence provided with these types of assessment.

References

Afflerbach, P. (2012). *Understanding and using reading assessment, K-12*. Routledge.

Alabsi, T. A. (2016). The effectiveness of role-play strategy in teaching vocabulary. *Theory and Practice in Language Studies*, 6, 227-234. doi: http://dx.doi.org/10.17507/tpls.0602.02

ALTE (2011). *Manual for language testing development and examining: For use with the CEFR*. Council

of Europe. Retrieved from https://rm.coe.int/manual-for-language-test-development-and-examining-for-use-with-the-ce/1680667a2b

Altun, M. (2015). Using role-play activities to develop speaking skills: A case study in the language classroom. Proceedings of the 6th International Visible Conference on Educational Studies and Applied Linguistics, Iraq, 354-363.

Ashcraft, N., & Tran, A. (Eds.) (2010). *Teaching listening: Voices from the field*. Alexandra, VA: TESOL Press

Asher, J. (1988). *Learning another language through actions: The complete teacher's guidebook*. Los Gatos, CA: Sky Oaks Productions.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press. Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Behfrouz, B. & Nahvi, E. (2013). The Effect of Task Characteristics on IELTS ReadingPerformance. *Open Journal of Modern Linguistics, 3* (1), pp. 30-39 Boyle, A., & Hutchinson, D. (2009). Sophisticated tasks in e-assessment: What are they and what are their benefits? *Assessment & Evaluation in Higher Education, 34* (3), 305- 319.

Brown, H. D. (2003) *Language assessment. Principles and classroom practices*. Pearson Longman, San Francisco, California.

Brown, H. D. (2004). *Language assessment: principles and classroom practices*. New York: Pearson Education.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. Pearson Longman.

Brown, J. D. (1998). *New ways of classroom assessment*. Alexandria, VA: TESOL.

Bryant, W. (2017). Developing a Strategy for Using Technology-Enhanced Items in Large Scale Standardized Tests. *Practical Assessment, Research & Evaluation, 22*(1). Retrieved from http://pareonline.net/getvn.asp?v=22&n=1

Buck, G. (2001). *Assessing listening*. New York: Cambridge University Press.

Bulut, B. & Ertem, I. (2018). *A Think-Aloud study: Listening comprehension strategies used by primary school students*. Faculty of Education, Adnan Menderes University, Aydın, Turkey. Retrieved from: https://files.eric.ed.gov/fulltext/EJ1175612.pdf

Butler, Y. G & Lee, J. (2010). On-task versus off-task among Korean elementary school students studying English. *The Modern Language Journal, 90* (4), 506-518.

Cambridge ESOL (2011) *Using the CEFR: Principles of Good Practice*. Cambridge: Cambridge ESOL. Retrieved from https://www.cambridgeenglish.org/images/126011-using-cefr-principles-of-good-practice.pdf

Chapelle, C. A. (2010). Technology in language testing [video]. Retrieved July 14, 2023 from http://language-testing.info/video/main.html

Chapelle, C. & Douglas, D. (2006) *Assessing language through computer technology*. Cambridge: Cambridge University Press. Center for Applied Linguistics. Retrieved from https://www.academia.

edu/38423343/COMPUTER_ASSISTED_LANGUAGE_TESTING-CALT_ISSUES_AND_CHALLENGES

Cohen, L. (2020). *The power of portfolios*. Retrieved from https://www.scholastic.com/teachers/articles/teaching-content/power-portfolios/

Curtain, H. A., & Dahlberg, C. A. (2016). *Languages and children, making the match: new languages for young learners*, Grades K-8. Pearson Education.

Darmuki, I., Andayani, S., Nurkamto, J., & Saddhono, K. (2017). Developing an Interactive Speaking Test: The Challenges and Solutions. *English Language Teaching, 10*(7), 133-141. https://doi.org/10.5539/elt.v10n7p133

Encalada, M. R. (2018). Role-plays as an assessment tool in English as a foreign language (EFL) class. In S. Soto, E. Intriago Palacios, & J. Villafuerte (Eds.), *Beyond Paper- and Paper Tests: Good Assessment practices for EFL Classes* (pp. 49-73). Machala, Ecuador: Editorial UTMACH.

Flutcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York, NY: Routledge

Flutcher, G. & Davidson, F. (2017). *The Routledge handbook of language testing*. New York, NY: Routledge

Garton, S., & Copland, F. (Eds.). (2019). *The Routledge Handbook of Teaching English to Young Learners*. New York, NY: Routledge

Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading*. Pearson Education Limited. Green, C. García-Millán, C. Lucendo-Noriega, A. (2022). *Spotlight: Formative assessment: Improving learning for every child*. https://jacobsfoundation.org/wp-content/uploads/2022/06/hundred_formative_assessment_digital.pdf

Jodoin, M.G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement, 40* (1), 1-15.

Kagan, S. L. (1999). A5 Redefining 21st-Century Early Care and Education. *Young Children. 54*(6). p. 2-3.

Kagan, S. L., Shepard, L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington D.C.: U.S. Government Printing Office

Kane, M. (2006). Content-related validity evidence in test development. In S.M. Downing & T.M. Haladyna (Eds.) *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Katz, L. G. (1997). *A Development Approach to Assessment of Young Children*. ERIC Digest.

Ketabi, S., & Somaye, S. (2015). Different methods of assessing writing among EFL teachers in Iran. *International Journal of Research Studies in Language Learning, 4*(1), 27-38. Retrieved from https://www.researchgate.net/publication/278329954_Different_methods_of_assessing_writing_among_EFL_teachers_in_Iran_Different_methods_of_assessing_writing_among_EFL_teachers_in_Iran

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

McKay, P. (2008). *Assessing Young Language Learners*. UK: Cambridge University Press.

McNamara, M. J. & D. Deane. (1995). Self-assessment activities: Towards autonomy in language learning. *TESOL Journal, 5* (1). 17-21.

Mohammed, J. (2017). The effect of peer assessment on the evaluation process of students. *International Education Studies, 10* (6), 159-173. https://www.researchgate.net/publication/317257376_The_Effect_of_Peer_Assessment_on_the_Evaluation_Process_of_Students

Mostafa, R. (2019). Standardized testing for young learners: An overview. *Education Sciences, 9* (2), 144. doi: 10.3390/educsci9020144.

National Education Goals Panel (1998). *Principles and recommendations for early childhood assessments*. https://govinfo.library.unt.edu/negp/reports/prinrec.pdf

Noijons, J. (1994). Testing computer assisted language tests: Towards a checklist for CALT.*CLICO Journal, 12* (1), 37-58

Parshall, C.G., Harmes, J.C., Davey, T., & Pashley, P. (2010). Innovative items for computerized testing. In W.J. van der Linden & C.A.W. Glas (Eds.*) Computerized adaptive testing: theory and practice* (2nd. ed.). Norwell, MA: Kluwer Academic Publishers.

Pierce, L. V. & J. M. O'Malley. (1992). *Performance and portfolio assessment for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.

Rivera, A. (2022). *Interactive reading as an approach to enhance reading comprehension through English texts in seventh graders*. https://repositorio.ucaldas.edu.co/bitstream/handle/ucaldas/18899/AngelaMaria_RiveraSalazar_2023.pdf?sequence=1&isAllowed=y

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI* (pp. 573–603). Academic Press.

Russell, C. and Shepherd, J. (2010). Online role-play environments for higher education. *British Journal of Educational Technology 41*(6), 992–1002.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates

Shabaan, K. (2001). Assessment of young learners. *FORUM, 39* (4), 16-22. https://www.researchgate.net/publication/234709045_Assessment_of_Young_Learners

Smith, J., Brewer, D., & Heffner, T. (2003). Using portfolio assessments with young children who are at risk for school failure: *Preventing school failure, 48*, 123-129. doi: 10.1080/1045988X.2003.10871078

Soto, S., Palacios, E. & Villafuerte, J. (2018). *Beyond paper-and-pencil tests: Good assessment practices for EFL students*. Ecuador: Universidad Técnica de Machala. http://repositorio.utmachala.edu.ec/bitstream/48000/14443/1/Cap.2%20Role-plays%20as

Spiller, D. (2009). Assessment Matters: Self-Assessment and Peer Assessment. Teaching Development, The University of Waikato. Retrieved from http://www.waikato.ac.nz/tdu/pdf/booklets/8_SelfPeerAssessment.pdf

Stevens, G. & DeBord, K. (2001). Issues of assessment in testing children under age eight. *The Forum for Family and Consumer Issues*, 6 (2). pp 1-7.

Sulaiman, Z. & Khan, M. (2019). Computer assisted language testing (CALT): Issues and challenges. *International Journal of Higher Education and Research, 9* (1), 1-11. https://www.researchgate.net/publication/331311015_COMPUTER_ASSISTED_LANGUAGE_TESTING_CALT_ISSUES_AND_CHALLENGES

Tannenbaum, J. (1996). *Practical ideas on alternative assessment for ESL students*. ERIC Digest. ED395500, Washington, DC: ERIC Clearinghouse on Languages and Linguistics.

Tarrant, M., Knierim, A., Hayes, S.K., & Ware (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education Today, 26* (8), 354-363.

Tsgari, D., Vogt, K., Froehlich, V., Csépes, I., Fekete, A., Green, A., Hamp-Lyons, L., Sifakis, N., & Kordia, S. (2018). Handbook of Assessment for Language Teachers, Erasmus +: TALE Project. https://taleproject.eu/pluginfile.php/2129/mod_page/content/12/TALE%20Handbook%20-%20colour.pdf

Wolf, K. M., & Butler, Y. G. (2017). *English Language Proficiency Assessments for Young Learners* (1st ed.). Routledge.

Wright, A. (2002), *Storytelling with children*. Oxford University Press 978-0132855211