

Chapter 6

Standardized Tests for Assessing Young Learners' English Proficiency

*MARÍA NELLY GUTIÉRREZ ARVIZU

DOI: <https://doi.org/10.52501/cc.255.06>

Abstract

With the growth of English language teaching in elementary education around the world, a demand to assess language proficiency has emerged. Language testing companies have developed tests to determine achievement, certification, progress, and proficiency. These tests are designed for audience ranging from 7 to 12 years old. This chapter analyzes seven tests to assess English proficiency of young language learners around the world.

Resumen

Con el crecimiento de la enseñanza del idioma inglés en la educación primaria en todo el mundo, ha surgido una demanda para evaluar el dominio del idioma. Las empresas dedicadas a la evaluación de idiomas han desarrollado exámenes para determinar los logros, la certificación, el progreso y el dominio. Estas evaluaciones están diseñadas para público de entre 7 y 12 años. Este capítulo analiza siete exámenes para evaluar el dominio del inglés de jóvenes estudiantes de idiomas en todo el mundo.

Introduction

In recent years, English language programs have been implemented as part of the school curricula in many countries where English is not the

* PhD in Applied Linguistic. Associate Professor Foreign Languages, Universidad de Sonora.
ORCID: <https://orcid.org/0000-0001-7284-0939>

means of communication outside the classroom (Rich, 2014). In these English as a Foreign Language (EFL) contexts, schools may have also considered administering a test. But, what test would be most suitable for their students?, their curricula?, their needs? There are several options that testing companies have made available to schools. The purpose of this article is to provide information on English language tests that school administrators, and teachers may consider when deciding which test to administer. The information presented here is an overview of seven tests designed to measure young learner' proficiency in English.

First, let us turn our attention to the target audience for these language tests. Young language learners are students between the ages of 5 and 12 learning another language besides their native language (Hasselgreen, 2012). In general terms, the students in this age group are attending elementary education where learning a language is one school subject. They usually do not have a choice in the language they would like to study, nor the types of assessments they would have. Young learners have specific characteristics such as cognitive development, learning stage, motivation, and attention that require special attention. When developing curricula, materials, activities, and assessments, these characteristics must be considered to develop language assessment instruments to be age appropriate.

There are at least six considerations to be made when assessing young language learners. According to Rea-Dickins (2000), there is a consensus that the following features should be present: *a*) tasks should be appealing to the age group, interesting and captivating, preferably with elements of game and fun; *b*) many types of assessment should be used, with the pupil's, the parents' and the teacher's perspectives involved; *c*) both the tasks and the forms of feedback should be designed so that the student's strengths (what he or she can do) are highlighted; *d*) the test taker should, at least under some circumstances, be given support in carrying out the tasks; *e*) the teacher should be given access to and support in understanding basic criteria and methods for assessing language ability; and, *f*) the activities used in assessment should be good learning activities in themselves. In general, tests for children have visuals and colors, are age appropriate in terms of tasks and length, should be scenario based, should be a good learning experience, and should be motivating to learn (Choi, 2008; Shaaban, 2001).

To be more specific, large-scale testing has four characteristics. McKay (2006) pointed out that they should be targeted for schools, administrators, and parents and should have a rigorous development process. Also, tests and examiners should be available. The use of technology was also emphasized. The author also indicated that the use of these tests are for institutions, program administrators teachers, and parents to make changes to curricula, modify teaching (long run), certify their students, and track students' performance.

This chapter will present a test review of seven standardized tests to assess proficiency in EFL learners. First, the procedures to select the tests will be presented. This will be followed by information about the tests.

Procedures

Reviewing tests consists of systematically analyzing information on tests. The first step in a test review is to locate the tests according to specific criteria for allowing the selection of available tests for assessing young learners in EFL contexts. In this case the criteria were the following: 1. standardized tests, 2. in English, 3. targeted to young learners (6-12 years old), 4. in EFL contexts, 5. designed by testing companies, and 6. currently available for administration. An online search using key words such as "English," "tests," "young learners," "children," "EFL" was conducted. After that, websites for major testing companies and publishing companies were searched to look for tests that met the criteria. Finally, a selection of tests to be included in the test review was done. Seven tests were chosen because they fulfilled the criteria established. Table 1 shows information at a glance on the tests in the first column of the table. General information about the developer (i.e. publishing company), delivery mode (paper and pencil, computer, or online), skills assessed (listening, speaking, reading, writing), Common European Framework of Reference for Languages (CEFR) target level (Pre A1 to C2), and website is provided.

Table 1. *General Information about English Tests for Young Language Learners*

Test	Developer	Delivery Mode	Skills Assessed	CEFR Target Level/ Results	Notes:
Center for Applied Linguistics English Proficiency Test for Students: Listening and Reading (CAL EPT)	Center for Applied Linguistics	Paper-pencil	Reading Listening	CENNI (0-10) PreA1 to B1	Designed and administer for Programa Nacional de Inglés in Mexico
A2 Key (KET for schools)	Cambridge English	Paper-pencil or computer based	Reading & Writing Listening Speaking Grammar & Vocabulary	A2 but results may be from A1 to B1	Formerly named Key English Test for Schools
Pre A1 Starters, A1 Movers, A2 Flyers (YLE Test)	Cambridge English	Paper-pencil or online	Listening Reading & Writing Speaking	Pre A1 to A2	Formerly named Young Learners of English (YLE) Test
Bronze, Silver, Gold Michigan Young Learners of English (MYLE)	Michigan Language Assessments	Paper-pencil using an answer sheet	Listening Reading & Writing Speaking	Pre A1 to A2	
Pearson Test of English Young Learners (PTEYL)	Pearson	Paper-pencil	Listening Reading Speaking Writing	Pre A1 to A2	
Oxford Young Learners Placement Test (OYLPT)	Oxford University Press	Online (adaptive)	Language Use: Vocabulary, Functions, & Grammar Listening	Pre A1 to B1	
Test of English as a Foreign Language Primary (TOEFL-P)	Educational Testing Service	Paper-pencil and computer	Reading Listening Speaking	Pre A1 to B1	

Note: The text will use the acronyms in parentheses.

Information about the Tests

After reviewing the tests, several salient features across tests can be observed. The analysis will focus on: 1) the skills used to measure proficiency in a language, 2) the framework used in tests, 3) the tasks examinees have to do, 4) the variety of English used in tests, 5) the reporting methods, 6) the materials offered by testing companies, 7) the reliability and validity evidence provided, 8) the delivery modes used, and 9) strengths and weaknesses.

The tests were designed to measure the construct of proficiency in English through assessing a combination of the language skills. All of the tests assessed the receptive skill of listening. The OYLTP did not assess reading as a separate skill but used written text to test language use through vocabulary, function, and grammar. The productive skill of writing was usually combined with reading as can be seen in YLE and MYLE. It seems that writing was limited to being able to write words or phrases. Only the KET for Schools and PTEYL had a writing task. In the former, the task was writing an email, while the latter was an integrated task with listening. For speaking, the OYLPT and CAL EPT did not assess it at all. For the rest of the tests, assessing speaking varied greatly. The KET for Schools had an interview with an examiner and two examinees at a time. A role-play between the students was one of the tasks. The YLE and MYLE had one-on-one conversation between the oral examiner and the examinee. There were tasks for finding differences, completing a story, finding the picture that did not belong, and a casual conversation about the examinee's life.

The PTEYL prompted speaking using a board game with the examinees which might make it less stressful in a testing situation. For the TOEFL-P, the speaking section consisted of descriptions, expressions, requests, questions, directions, and a narration. This was done using a computer in order for the rating to be scored by e-raters in a remote location using the feedback tool provided by the Educational Testing Service (ETS). All seven tests were aligned with the CEFR levels. This framework was developed to standardize curricula, materials, assessments of language in Europe (Council of Europe, 2001) but it has been adopted by other countries. Interestingly, one of the tests, the CAL EPT, was also aligned to the levels in

the Mexican certification of level in languages (CENNI) developed as a framework to assess and certify foreign language proficiency in Mexico, particularly in English (Secretaría de Educación Pública, n.d.). There are 20 levels in CENNI but the CAL EPT assessed up to the first 10 levels, which correspond to Pre-A1 to B1 levels in the CEFR.

The tasks that these tests present for the examinees attempt to resemble the tasks that school age children do in class. These are matching, completing sentences or paragraphs, coloring, describing pictures, completing stories, playing a board game, among others.

Hasselgreen (2005) highlighted the need to have a variety of tasks when assessing young learners. The tests that were examined offered a variety of tasks for the students but within the same format or context. Moreover, Wilson (2005) pointed out that tasks in tests for young learners should be a pleasant experience for the test takers. A common feature among the tests was the use of visuals as part of tasks throughout the tests. The KET for Schools, TOEFL-P, YLE, and MYLE had tasks that require examinees match pictures as part of the listening tasks. Table 2 has information of input, tasks, items, and time each section in the test has.

Regarding the variety of English used and accepted in the tests, British and American English were the dialectal varieties present. The KET for Schools, YLE, and OYLPT use British English in instructions and speech input but accept American English as responses to the tasks. MYLE focuses on American English accepting international varieties. The CAL EPT and TOEFL-P are tests that have been designed by American testing companies, and even though not specifically stated in the information available to the public, the focus might be American English but accepting other varieties in the responses from the test takers.

In terms of delivery mode, three different types were observed: paper-pencil, computer adaptive, and computerized. The most common was the paper-pencil administration. The OYLPT is the only computer adaptive tests. The YLE has a computer version but it is not adaptive. The purpose of the OYLPT is to place students in a level based on the answers they give through the computer.

When reporting scores to children and parents, due to the nature of these tests, the certificates have symbols such as shields (YLE), medals

Table 2. Skills Assessed per Test

Test	Listening			Reading			Writing			Speaking		
	Input	Tasks Items	Time	Input	Tasks Items	Time	Input	Task	Time	Input	Tasks	Time
CAL EPT for Students YLE	Not specified	Not specified	45 min	Not specified	Not specified	45 min						
	Conversations	5 25	25 min	Labeled pictures Picture Story Sentences Passage	6 40	30 min	Included in Reading testlets		In Reading time	Find the differences Story with pictures Odd one out Personal questions	4	5
MYLE	Multiple dialogues Conversation Dialogues	5 25	25 min	Labeled pictures Picture Story Sentences Passage	6 40	30 min	Included in Reading testlets		In Reading time	Find the differences Story with pictures Odd one out Personal questions	4	6
	Short conversations Long conversations Radio host	5 25	30 min	Ads Sentences Conversations Article Definitions Cloze Invitation	8 55	70 min	Email	1	In Reading time	Personal conversation Roleplay among candidates	2	10
OYLPT	Monologues Dialogue	3 12		(Not Reading) Language Use: Vocabulary Function Grammar cloze	3 18							
PTEYL	Continuous dialogues on the same topic	2 15	60 min total	Questions in dialogue Sentence to Picture Picture Passage	3 20	60 min total	Answer questions	1	60 min total	Board game Short talk	2	20
TOEFL Primary	Sentences Conversations Lecture	Not specified 36	30 min	Picture to choose sentence Definitions Email Passage	Not specified 36	30 min				Description Expression Requests Questions Directions Narration	6	20

Note: Center for Applied Linguistics English Proficiency Test (CAL EPT); Key English Test for Schools (KET); Young Learners of English (YLE) Test; Michigan Young Learners English (MYLE); Pearson Test of English Young Learners (PETEYL); Oxford Young Learners Placement Test (OYLPT); Test of English as a Foreign Language Primary (TOEFL-P)

(MYLE), badges, or stars (TOEFL-P) and even a metaphor of launching a rocket (OYLPT). This makes the reports more appealing and easier to understand for the target population, children. However, this type of report might also be vague complicating the process of making decisions such as curricular changes or teaching training for institutions. For the institutions, the testing companies provide different types of reports that are numerical and have more information that can be used to monitor progress, or achievement. The CAL EPT reports provide two results; one is on the CEFR and the other on the CENNI. The reason for this might be to provide information to teachers, school administrators, and curricula developers in Mexico about the proficiency and progress of students in public schools.

Most of the testing companies offered preparation materials for the students and teachers to use in class. The materials could be vocabulary lists (KET for Schools, YLE), sample items (TOEFL-P), previous tests (YLE, PTEYL), even lesson plans (OYLPT).

Regarding psychometric information on these tests of English for children, the information on reliability indices and standard error of measurement, was not available to the public in most of the cases. YLE and KET for Schools are produced by Cambridge English Language Assessments and information on reliability was provided. The coefficients for Cronbach's alpha and the Standard Error of Measurement provided indicate that they are reliable assessments. CAL EPT pilot test data was analyzed using the Rasch model. In that case, medium to high coefficients of person reliability were found. This reliability shows how well test takers' proficiency levels are differentiated (Myford, 2006). Higher values indicate higher differentiation. MYLE reported the percentages of distribution of test taker scores (in medals) per section in each of the tests for the 2014 administrations. This information did not provide evidence for reliability. An interesting aspect to mention is that Cambridge English Language Assessment and Michigan Language Assessment have similar tests since both organizations have worked closely. Weyant and Chisholm (2014) described the how test fairness is approached by both institutions. CaMLA and Cambridge English Language Assessment adhere to six principles of fairness: demonstrate respect for people, demonstrate respect for personal convictions and beliefs, demonstrate sensitivity to population differences and world knowledge,

avoid undue negativity, avoid unduly controversial or upsetting topics, and avoid construct-irrelevant knowledge.

Each test had different strengths and weaknesses. These features may depend on the use that the test had. In other words, it depends on what the test is needed for, the institution's resources (computers, spaces, finances), and practical implications, that the characteristics may make a specific test more suitable than the rest. The last two columns of table 3 show the strengths and weaknesses of each test.

Table 3. *Analysis of Tests*

<i>Test</i>	<i>Purpose</i>	<i>Delivery Mode</i>	<i>Time</i>	<i>Skills Assessed</i>	<i>Reliability</i>	<i>Strengths</i>	<i>Weaknesses</i>
CAL EPT	Proficiency (Program success)	Paper and Pencil	90 min	Listening Reading	IRT person index around .80	– Tailored for a specific context	– 2 skills – Lengthy
KET for Schools	Certification at A2 CEFR	Paper and Pencil or Computer	110 min	Reading & Writing Listening Speaking	Total alpha higher than .90	– Designed with school topics	– Lengthy
YLE	Proficiency and Achievement	Paper and pencil or Online	60 min	Listening Reading & Writing Speaking	alpha per section higher than .80	– 4 skills – Trained oral examiners	– Oral examiners needed on site
MYLE	Proficiency and Certification form Pre-A to A2 CEFR	Paper and pencil with answer sheet	60 min	Listening Reading & Writing Speaking	Not available	– 4 skills	– Oral examiners needed on site
OYLPT	Placement	Online	30-40 min	Language use (Gr and Vo) Listening	Not available	– Fast scores (Computer adaptive)	– Not by skills
PTEYL	Proficiency	Paper and Pencil	60 min	Listening Reading Speaking Writing	Not available	– Scenario based	– Limited times for test administration
TOEFL-P	Placement and Achievement	Paper and Pencil with computer for speaking	80 min	Listening Reading Speaking	Not available	– Academic oriented	– Multi-mode delivery – Tasks may not be appropriate for target population

Conclusion

This chapter presented an overview of the tests that may be useful for different purposes. Each test has a unique set of features that make it a good fit for the purposes that schools and programs may have. It is up to the user to decide which test would be the most beneficial for their goals. In the case of young language learners, it is important to consider the tasks, testing time, as well as cognitive maturity to select the test that will be used.

In summary, the analysis of standardized tests for assessing young learners' English proficiency underscores the diverse array of options available to educators and institutions. Each of the seven tests reviewed offers distinct features tailored to various needs and contexts, reflecting the growing emphasis on accurately measuring language development in young learners.

The effectiveness of a language assessment tool hinges on several factors, including the alignment with developmental stages of children, the range of language skills evaluated, and the relevance of the tasks to everyday language use. This review highlights that while all the tests incorporate essential elements such as listening comprehension and reading, their approaches to assessing writing and speaking vary significantly. This variation points to the necessity for schools and programs to carefully consider the specific requirements and contexts of their students when selecting an assessment tool.

Furthermore, the diverse delivery modes, ranging from paper-pencil tests to computer-based formats, demonstrate the integration of technology in modern language assessment, offering flexible options for administrators. The choice of assessment should also consider factors such as the test's alignment with established proficiency frameworks like the CEFR, and the availability of preparatory materials which can enhance both teaching and learning experiences.

Ultimately, the decision on which test to implement should be guided by the test's suitability to the educational objectives, the cognitive and developmental needs of the learners, and practical considerations such as resources and administrative capacity. By aligning assessment tools with

these factors, educators can better support young learners in their language development journey, ensuring that assessments are both effective and meaningful in evaluating English proficiency.

References

- Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39-62.
- Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Press Syndicate of University of Cambridge.
- Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), 337-354.
- Hasselgreen, A. (2012). Assessing young learners. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing*. Routledge. <https://doi.org/10.4324/9780203181287>
- McKay, P. (2006). *Assessing young language learners*. Cambridge University Press.
- Myford, C. (2006). *Analyzing rating data using Linacre's FACETS computer program: A set of training materials to learn to run the program and interpret output*. [Unpublished manuscript]. University of Illinois, Chicago.
- Rea-Dickins, P. (2000). Assessment in early years language learning contexts. *Language Testing*, 17(2), 115-122.
- Rich, S. (2014). *International perspectives on teaching English to young learners*. Palgrave Macmillan.
- Secretaría de Educación Pública. n.d. CENNI: Tipos de Cenni [Website]. Available at https://dgair.sep.gob.mx/tipos_cenni (accessed on 20 November, 2024).
- Shaaban, K. (2001). Assessment of young learners. *English Teaching Forum*, 39(4), 16-27.
- Weyant, K., & Chisholm, A. (2014). Safeguarding fairness principles through the test development process: A tale of two organizations. *Research Notes*, 55(1), 3-6.
- Wilson, J. (2005). Ethical issues in the testing of young learners. *Research Notes*, 22(4), 6-7.