

Artículo científico de investigación

DOI: <http://doi.org/10.15517/revedu.v49i2.751>

Preschool English Foreign Language Vocabulary Test (PEVOT) for ESL Training in Costa Rica

*Instrumento de Evaluación de Vocabulario para la Enseñanza
del Inglés como Lengua Extranjera para Preescolares en
Costa Rica*

Verónica García-Castro
Universidad de Costa Rica
San José, Costa Rica
veronica.garciacastro@ucr.ac.cr (Correspondencia)
<https://orcid.org/0000-0003-2849-2855>

Delia Tamara Fuster Baraona
Universidad Nacional
Heredia, Costa Rica
delia.fuster.baraona@una.cr
<https://orcid.org/0000-0003-1326-4788>

Montserrat de Jesús Orozco Zúñiga
Investigadora independiente
San José, Costa Rica
orozcozm17@gmail.com
<https://orcid.org/0009-0004-5428-3883>

Silvia Benavides-Varela
Università degli
Studi di Padova, Padua, Italia
silvia.benavidesvarela@unipd.it
<https://orcid.org/0000-0003-4818-7372>

Kimberly Dayana Amador Romero
Investigadora independiente
San José, Costa Rica
kimberlyamador241@gmail.com
<https://orcid.org/0009-0008-9637-2605>

Recepción: 22 de febrero de 2025
Aceptado: 9 de mayo de 2025

¿Cómo citar este artículo?

García-Castro, V., Orozco Zúñiga, M. J., Amador Romero, K. D., Fuster Baraona D. T., & Benavides-Varela, S. (2025). Preschool English Foreign Language Vocabulary Test (PEVOT) for ESL Training in Costa Rica. *Revista Educación*, 49(2). <http://doi.org/10.15517/revedu.v49i2.751>

Esta obra se encuentra protegida por la licencia Creativa Atribución-NoComercial-CompartirIgual 4.0 Internacional



ABSTRACT

Vocabulary tests that match national curricula tend to be more suitable to the academic and social setting of students (Nikolov & Timpe-Laughlin, 2021). However, certain countries with limited educational budgets, such as Costa Rica (State of the Nation Report, 2021) may prioritize other areas, such as basic infrastructure or literacy, over developing and implementing standardized vocabulary assessments. This shortcoming was addressed in a descriptive study that used an innovative curriculum-based test to assess and explore English Foreign Language (EFL) vocabulary knowledge among Costa Rican school children. Participants included 128 (Mage = 5, SD = 6.8 months), students from rural and urban public schools. The methodology included a thorough analysis of the Ministry of Education (MEP) EFL preschool curriculum, followed by the design, piloting, and administration of a vocabulary test in English. The findings revealed that children of all ages are capable of recognizing EFL vocabulary included in the current EFL program at the Costa Rican Ministry of Education. However, urban public school students demonstrated a significantly higher academic performance than students from rural public schools. The study concludes that EFL testing is an effective method in assessing preschool student EFL vocabulary knowledge. The study also considers shortcomings and future recommendations.

KEYWORDS: Foreign Languages, Students Evaluation, Early Childhood Education, English as a Foreign Language, EFL, Bilingualism, Preschool Children.

RESUMEN

Las pruebas de vocabulario vinculadas a los currículos nacionales suelen ser más relevantes para los entornos académicos y sociales de los estudiantes (Nikolov y Timpe-Laughlin, 2021). Sin embargo, algunos países como Costa Rica, con presupuestos educativos limitados (Programa Estado de la Nación, 2021), pueden priorizar otras áreas, como la infraestructura básica o la alfabetización, por encima del desarrollo e implementación de evaluaciones de vocabulario estandarizadas. Este trabajo tuvo como objetivo abordar esta limitación presentando una prueba innovadora basada en el currículo para evaluar y explorar el conocimiento del vocabulario del idioma extranjero inglés (EFL) en niños de Costa Rica. Los participantes incluyeron 128 (edad media = 5, DE = 6,8) estudiantes de áreas rurales y urbanas que estudian en escuelas públicas. La metodología incluyó un análisis exhaustivo del currículo preescolar de EFL del Ministerio de Educación (MEP), seguido del diseño, la prueba piloto y la administración de la prueba de vocabulario en inglés. Los resultados mostraron que los niños, independientemente de su edad, pueden reconocer el vocabulario en inglés como lengua extranjera incluido en el actual programa de inglés como lengua extranjera del Ministerio de Educación. Sin embargo, el estudiantado de áreas urbanas demostró un desempeño significativamente diferente (puntajes más altos) en comparación con los estudiantes de áreas rurales. Se concluye que la prueba

de inglés como lengua extranjera es eficaz para evaluar el conocimiento del vocabulario en inglés como lengua extranjera de los estudiantes preescolares. También se discuten las limitaciones y las recomendaciones futuras.

PALABRAS CLAVE: Lengua extranjera, Evaluación de estudiantes, Educación de la primera infancia, Niñez en edad preescolar, Inglés como Lengua Extranjera, Bilingüismo, Niñez preescolar.

INTRODUCTION

Interest in adding foreign language learning into early childhood education and day care has been on the rise in many countries (Bergström et al., 2016; Ferjan-Ramírez et al., 2020). Learning foreign languages offers cognitive, social, economic, and cultural benefits that range from individuals to all of society, preparing people to thrive in a multicultural and interconnected world. In Costa Rica, a person who speaks English is likely to receive a 20.6% salary increase, work 0.8% fewer hours per week, have a 2.5% higher probability of receiving paid vacation, and have a 2.7% higher chance of paid sick leave compared to non-English speakers (Abarca & Ramírez, 2017).

English-language fluency, therefore, provides social mobility and, undoubtedly, cognitive and social benefits. However, investment in public English teaching is not widespread throughout the country. Only 23.2% of the population in urban public preschools have the option of taking English as a Foreign Language (EFL). The figure dwindles to only 6.8% of students in rural areas (State of the Nation Report, 2021). EFL refers to classroom-based instruction and learning of the English language in a country where English is not the primary language of communication (Richards & Schmidt, 2013). Furthermore, a dearth of curriculum-based tests to help assess the degree to which classes reflect the learning goals of the Ministry of Education, ensure standardized language proficiency across regions, and support teachers in lesson planning and student progress. Although Costa Rica has an age-appropriate vocabulary test for Spanish as a first language (L1) simply translating it would not be appropriate for assessing EFL proficiency, given that L1 tests the native language or the first language acquired by the student (Richards & Schmidt, 2013).

In very young learners, the typical process of language acquisition proceeds effortlessly (Tomasello, 2010), yet the quality of L1 linguistic input contributes to syntactic and semantic learning (Hoff, 2006; Tomasello, 2003) as well as vocabulary growth (Huttenlocher et al., 2010 & Edwards, 2018). L1 language development is also shaped by the number of conversational turns between the child and the primary caregiver (Romeo et al., 2018) and repeated exposure to linguistic input (Tomasello, 2010). Considering all these factors, L1 vocabulary tests for preschool children assess the extent of the vocabulary by presenting complex tasks such as recognition of synonyms, adjectives, and categories (Murillo-Rojas, 2012), among others.

Employing an L1 vocabulary test designed for native-speaking children with children learning EFL may pose challenges due to differences in the quality and frequency of the L2 input when compared to richer L1 input (Goriot et al., 2021), with L2 referring to the process of learning English after acquiring a first language in a context where English is necessary for daily communication (Richards & Schmidt, 2013). For instance, children typically have less exposure to EFL than L1, and the quality of EFL input may also lack the linguistic and cultural richness of L1 scenarios. Additionally, EFL children may use the foreign language mainly in classroom settings, whereas native-speaking children use it daily. Another challenge is that in L1 vocabulary tests, children could potentially deduce the meaning of unknown terms based on L1 linguistic context (Cain et al., 2004; Nakatsuhara et al., 2020), which early L2 learners may lack. Consequently, L2 vocabulary tests tend to be simpler and focus on word meaning recognition rather than categorizing words, reducing their cognitive load (Schmitt, 2008).

Despite the increasing need to evaluate EFL proficiency among very young learners (Nikolov & Timpe-Laughlin, 2021) and the importance of appropriate contextual and cognitive demanding vocabulary assessments for these L2 learners, there are currently no EFL tests available for very young learners in Costa Rican. This study aims to develop and administer an EFL vocabulary assessment to examine EFL acquisition among very young learners in Costa Rica, with the goal of improving EFL instruction and learning in the country. A properly-designed, country-specific EFL assessment allows educators to promptly identify children requiring additional language assistance as well as facilitate the design and creation of customized activities to support them and ensure seamless multilingual development prior to the start of primary school. The study specifically seeks to design and develop an EFL vocabulary assessment for children that aligns with the Costa Rican EFL preschool curriculum, evaluate preschoolers EFL vocabulary proficiency using the newly created test, and investigate the impact of sociodemographic factors, including age and geographical region, on test performance.

The study addresses the following research questions:

1. Is it possible to design an EFL vocabulary test for children based on the Costa Rican EFL preschool curriculum? If so,
2. What is the receptive EFL vocabulary knowledge of preschoolers as assessed by the newly created test?
3. Does test performance vary according to sociodemographic variables such as age and geographic location?

Given that there is less coverage of English language instruction in public preschools in rural than in urban areas (State of the Nation Report, 2021), it is anticipated that preschoolers from urban areas settings will perform better than those in rural areas

The study provides an extensive literature review of EFL instruction and acquisition in very young learners, together with vocabulary assessment among preschool-aged children in Costa Rica (Section 2). Section 3 details the methodology encompassing the participants and setting, test design, content validation, testing procedures, and data analysis. Section 4 delineates the findings, content and test validity, vocabulary assessment of the full sample, and the influence of sociodemographic characteristics. Section 5 analyzes the findings while Section 6 outlines the conclusion, study limitations and recommendations for future research, followed by a reference list and an appendix.

Literature review

EFL teaching and learning in very young learners in Costa Rica

EFL instruction and acquisition in early childhood occur after exposure to the second language (L2) after the age of three (Lakshmanan, 2013). While a specific critical period for L2 acquisition remains undefined, it is generally accepted in order to learn foreign languages more effectively, the younger, the better (Mihaljević-Djigunović et al., 2008; Ministerio de Educación Pública [MEP], 2007). EFL vocabulary acquisition can be classified into two categories - receptive and productive learning. The former indicates that learners can comprehend a word in its spoken or written form, while the latter refers to their ability to use it in verbal or written communication in the foreign language (Schmitt, 2010). Receptive vocabulary knowledge refers the ability of a learner to comprehend a word when they encounter it (input), whereas productive vocabulary knowledge pertains to the capacity of a learner to produce and articulate the word as.

The national preschool EFL curriculum in Costa Rica considers both EFL receptive and productive learning through teaching methods such as the communicative approach, total physical response, community language learning, the natural approach, and whole language (Ministerio de Educación Pública [MEP], 2007). Exposure to a foreign language at an early age improves a child's cognitive abilities and cultural awareness. The Ministry of Education of Costa Rica (MEP) asserts that the English curriculum enhances understanding of Costa Rican culture and overall oral communication among children. Although the curriculum promotes listening and speaking skills, the primary objective of EFL instruction is oral communication.

Despite an EFL curriculum with pertinent L2 themes and their corresponding assessment, the quality of preschool instruction has been scrutinized for several reasons, including inadequate classroom interaction and insufficient reading time (Barquero et al., 2019). Preschooler pre-reading skills development appears to be slow with notable disparities between urban and rural areas (Barquero et al., 2019). Findings also reveal significant gaps in reading and learning skills among students from schools outside the Greater Metropolitan Area (GAM) compared to schools within the GAM. A study on this issue was conducted in schools from Costa Rican provinces of Heredia, Puntarenas,

and San José, where vocabulary assessments, phonological awareness evaluations and word-letter recognition exams were administered. Despite the positive results, experts argue that these outcomes may be associated with the development and growth of the child rather than the school attributes or effectiveness of activities, such as shared reading (Barquero et al., 2019).

Vocabulary testing in very young learners

Early vocabulary tests are important because they play a critical role in tracking language development, guiding instructional choices, and promoting cognitive and social growth among children. However, there are few tests available to examine L1 or L2 vocabulary recognition among young learners. For one, the Peabody Picture Vocabulary Test (PPVT-III) aims to measure verbal ability, receptive vocabulary knowledge, and comprehension of spoken English as a first language (Dunn & Dunn, 2007; Eigsti, 2017); that is, listening and understanding spoken single-vocabulary words. The test “does not require the child to speak, read, or write. Although most examinees respond by touching or pointing to pictures, the test can be administered to individuals with significant motor impairment” (Stein & Lukasik, 2009, p. 795). The set consists of 12 different items, all of which the examinee must correctly identify, except for one or none of the test items. One by one, the examiner must say a stimulus word aloud that the examinee then identifies by pointing or saying the number of the test plate.

Vocabulary tests designed for L1 speakers frequently prove to be unsuitable for L2 learners, since they fail to consider the distinct learning requirements, backgrounds, and developmental stages of L2 learners. Consequently, significant efforts have been made to create vocabulary assessments tailored to specific cultural backgrounds or countries. One such assessment is the English Picture Vocabulary Test (EPVT). The objective of EPVT is to assess the capacity of very young learners to comprehend and articulate EFL vocabulary knowledge within specific subjects (Güngör & Önder, 2023). The EPVT was administered to a sample of Turkish children aged five and six years, all students of sixteen different preschool institutions representing low and middle-high socioeconomic backgrounds. All of the study participants received English language instruction at their institutions, at a high, moderate, or low intensity level. Güngör and Önder (2023) assert that this assessment considers the characteristics and their learning needs of young learners. The EPVT can be used to evaluate language development in children, vocabulary acquisition levels and the influence of different curricula and instructional methods on their language structure learning (Güngör & Önder, 2023).

The Cambridge Young Learners English (YLE) is a test battery that evaluates the English language proficiency of young learners. It evaluates four macro skills of the English language (i.e., reading, writing, listening, and speaking), emphasizing oral and listening skills, since they are the skills that are mostly developed in children. Writing activities consist of words and phrases, since children at this learning stage have yet to acquire the organizational and cognitive skills necessary for producing writ-

ten content (Taylor & Saville, 2002). The assessments were formulated using textbooks and classroom resources to identify the main topics often present in preschool syllabi and aiming to monitor challenges related to L2 language development skills in children from beginner to low-intermediate levels (Taylor & Saville, 2002). The YLE tests are structured as task-based assessments and aim to integrate brief, game-like activities such as coloring drawings, in contrast to the aforementioned Peabody Picture Vocabulary Test (Dunn & Dunn, 2007) and the English Picture Vocabulary Test (Güngör & Önder, 2023).

To the best of our knowledge, the TEVOPREESC is the only early-level vocabulary test currently available in Costa Rica. The TEVOPREESC is a recognition test designed to assess the vocabulary knowledge of preschoolers aged three to six whose L1 is Spanish. The TEVOPREESC includes linear black and white illustrations (Murillo-Rojas, 2012). The test is administered one-on-one and includes six main sections: naming, classification of categories, comprehension, knowledge of colors, word definitions, and lexical richness.

To summarize, a plethora of language exams are available for assessing children worldwide. Each, however, is customized to a specific context and assessment objective. The TEVOPREESC is the only language instrument available in Costa Rica to evaluate preschooler L1 vocabulary. It is inappropriate to translate it to evaluate EFL, since it is similar to other evaluations that are intended to evaluate L1. Therefore, it is imperative to develop an EFL assessment instrument for preschoolers that is tailored to the Costa Rican context to assess their receptive vocabulary. In addition to providing valuable results for L2 language classrooms and teaching practices, a specialized EFL test for children in Costa Rica would also enhance national education, particularly to attain greater professional and cultural development in a globalized world.

METHODOLOGY

Research questions and aims

The aim of the study was to design and administer an EFL vocabulary assessment for preschool-aged children consistent with the Costa Rican EFL preschool curriculum. It also sought to evaluate the receptive EFL vocabulary knowledge of preschoolers, which refers to words that the child may comprehend, and to investigate the impact of sociodemographic factors on test performance, such as age and geographical region. The study sets out to answer the following research questions:

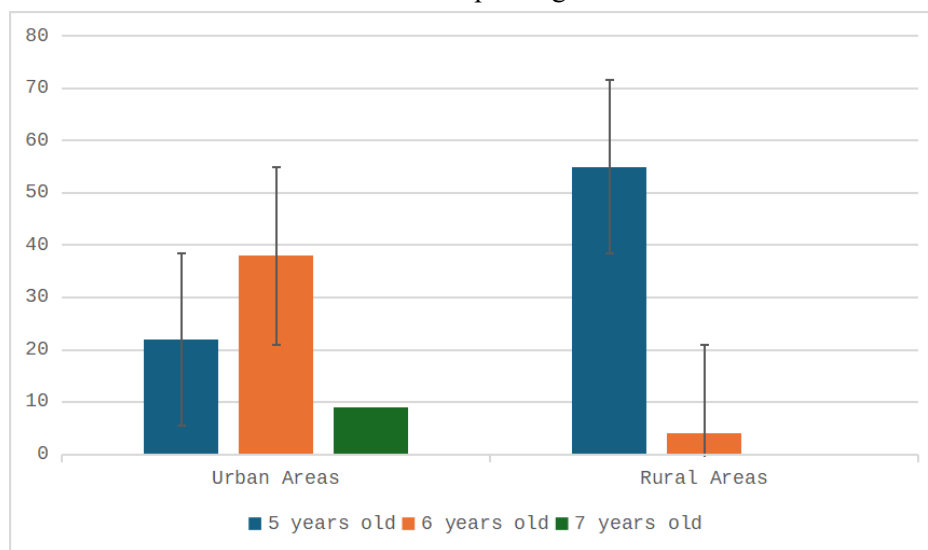
- 1) Is it possible to develop EFL vocabulary exams for children according to the Costa Rican EFL preschool curriculum? If so,
- 2) What is the EFL receptive vocabulary knowledge of preschool children as determined by the newly designed EFL vocabulary test?
- 3) Does student test performance vary according to sociodemographic characteristics, such as the participant age and geographical region of the study?

Participants and Setting

The study comprised 128 participants ($M_{age} = 5$, $SD = 6.8$ months) from schools in Costa Rica. A convenience sampling method was used to select the sample based on school accessibility. In urban areas, schools were selected based on their affiliation with the universities hosting the project. In rural areas, preschool teachers had been previously contacted to determine if their schools were willing to participate in the study. Three schools agreed to do so.

The majority of students ($N = 103$) were enrolled in the national preschool curriculum, and $n = 25$ in the primary school program. Participants were divided into two groups based on geographical study region. One group ($n = 69$) studied in urban areas (San José and Heredia) ($M_{age} = 6$, $SD = 6.5$ months), and the other group ($n = 59$) in a rural area (San Carlos) ($M_{age} = 5$, $SD = 3.6$ months) (Figure 1). All of the rural area participants were all enrolled in preschools. All participants were native speakers of Spanish and did not speak English at home. Most of the students received EFL support at school.

Figure 1.
Participant Age



Source: Author interpretation of the obtained results.

Test Design

In order to respond to research question 1: *Can an EFL vocabulary test for children be designed based on the Costa Rican EFL preschool curriculum?* the Preschool EFL Vocabulary Test (PEVOT) for preschool children in Costa Rica was developed and structured based on the following TEVOPREESC sections: 1) naming vocabulary related to basic concepts, antonyms, and categories, 2) classification of nouns according to their function, use, and characteristics and 3) comprehension of activities, spatial notion, position, quantifiers, and temporalization (Murillo-Rojas, 2012).

The PEVOT consists of four stages in keeping with Güngör and Önder (2023).

Stage 1: Format Definition

In the first stage, an in-depth analysis of the last available version of the MEP national EFL curriculum developed for preschool in Costa Rica was conducted, along with an analysis of teaching and assessment of ELF practices in the preschool context. The MEP curriculum is divided into four study blocks: Study block I: *Who am I?* Study block II: *I communicate with others in different ways*, Study block III: *I discover and enjoy my surroundings*, and Study block IV: *My relationship with people and objects around me*. The contents of each study block can be found here: https://osf.io/4x7p2/?-view_only=8f3dabf456814bae84fc2eca4a841b5c).

According to the program, teaching sessions should include examples of the content and language items related to the study block. For instance, a lesson based on the study block called *Who am I?* must include contents, such as *Self and other people's appearance, feelings, and emotions, cultural aspects, values, and assessment*. After a close examination of the curriculum and drawing from the structure of the TEVOPREESC, a four-picture test format was selected: a picture representing the target word, two pictures as semantic distractors, and one picture semantically unrelated to the target as a filler distractor.

Stage 2: Target words to be included

Based on the analysis of the last version of the EFL preschool curriculum, words from the four study blocks of the MEP (2007) program were incorporated. The first study block, *Who am I?* addresses various topics about students as a unique individuals, their family, and school. It also covers personal choices, their likes and dislikes, values, feelings, emotions, and appearance. Relatedly, words such as *tall* and *happy* were included. The second study block, *I communicate with others in different ways*, develops topics about communication, symbolic and body language, commands, signs, positive image through body health, and forms of expression through art, literature, music, and culture. Words such as *shoes, eyes, black, gray, brown, blue, and yellow* were selected for this block.

The third study block, *I discover and enjoy my surroundings*, explores topics related to services in the community, institutions, different elements in the surroundings, nature, and the environment. Thus, the words *cat, dog, flowers, doll, ball* were chosen for this study block. The fourth study block, *My relationship with people and objects around me*, includes topics concerning shapes, numbers, school objects, and the location of objects and people. The words *friends, small, big, round, house, and table* were selected (Figure 2) for this study block. The final version of the test included twenty target words (ten concrete nouns and ten adjectives) from the MEP program.

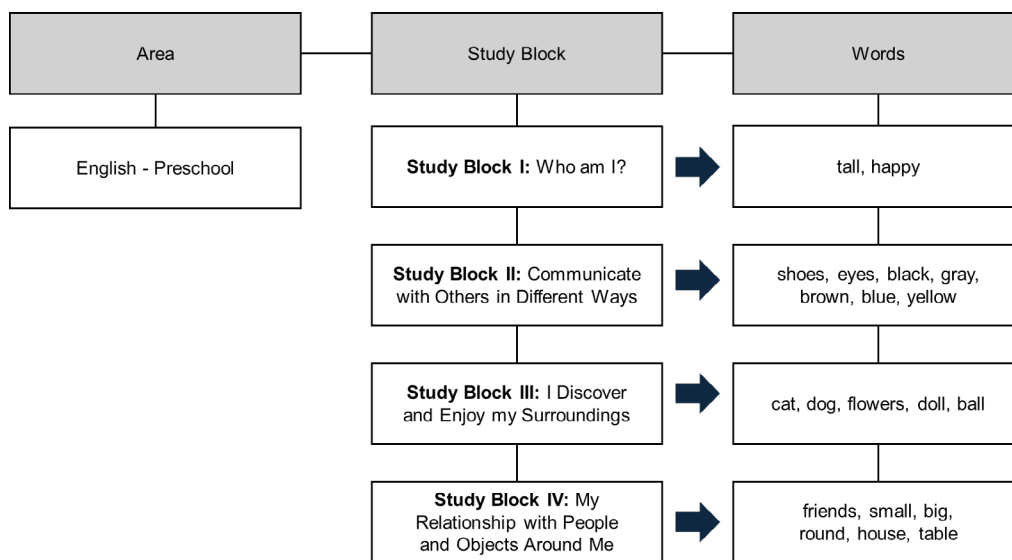
Stage 3: Visual design and support

The design included one testing sheet per target word ($N = 20$) and one training sheet for a total of 21 sheets. The test format was adapted from the TEVOPREESC. Firstly, a pattern of four spaces per

horizontal page was included in the test design. Then, a picture, with linear art, was placed in each space. Pictures were open-access and taken from Google Images for academic purposes.

Figure 2.

MEP Preschool EFL Program Study Blocks



Source: Author interpretation based on MEP preschool EFL Program data.

Each page had a target word and a picture that represented it as well as three semantically related distractor images. However, certain adjective target words, such as black, gray, brown, blue, yellow, round, big, tall, and small, adhered to a different pattern. Color adjectives consisted of four different pictures in color that did not repeat across the sheets. Additionally, one sheet depicted four different shapes to represent the target word round while other sheets displayed images in varying sizes, allowing participants to compare them for the target words small, big and tall. 1. For example, it was decided to use a larger picture of a frog for the target word big and a smaller picture of the frog for small. An additional sheet was created to appropriate distractors for the target word ball, which included three round-shaped distractors.

All of the pictures used for the PEVOT were in black-and-white linear art, except the pictures of adjectives describing colors based on Murillo-Rojas' (2012) TEVOPREESC. Requirements are consistent with a group agreement to only display images with parallel, bi-colored and minimalistic designs. The authors also decided to include distractions and fillers that fall into the same categories as the target words. Each target word was intentionally positioned differently on each subsequent to avoid learning effects. A full list of materials can be found here: https://osf.io/4x7p2/?view_only=8f-3dabf456814bae84fc2eca4a841b5c

Stage 4: Content validation and piloting

The pilot was developed in two stages:

Pre-piloting

An initial pre-piloting meeting with language learning researchers ($n = 3$) and research assistants ($n = 5$) was conducted to determine test difficulties and attractiveness. Based on feedback, modifications to the test were made. A second pre-piloting meeting was carried out to discuss test modifications. It was agreed to change how certain pages were ordered in order to match the semantic context. In a third pre-piloting meeting, the latest test version was discussed. According to this discussion, it was decided to rearrange some sheets to suit the target word. Colors were chosen in different shades for target words of this type to avoid creating obvious repetitions. Sheets for target words referring to colors were the only ones with colored illustrations in the test. Additionally, pictures of animals were chosen in different art styles. Moreover, the distractors were arranged in a pattern where the images belonged to a semantic group to prevent noticeable answers.

In a fourth pre-pilot meeting, the team agreed to modify sheets to suit the target words referring to sizes (small, big, and tall) and the colors of some of the pictures in the sheets, as they did not seem to present clear differences with the target color. As a result of modifications and improvements, five test versions were created before the piloting stage. The fifth version, which included all the modifications, was used in the pilot stage.

Piloting

Before piloting, the study was approved by the Ethics Committee of the University of Costa Rica (Protocol Number CEC-43-2023). Piloting was conducted with two children, aiming to assess whether instructions were clear, if the pictures were appealing, and to measure the duration of the test. Overall, the goal was to guarantee a positive testing experience for participants (Nikolov, 2016).

The first piloting session was conducted in a university context. Parents brought their children to the university and was present at all times during the test. The child was asked whether the items were easy or difficult to complete and if they understood the instructions and the pictures. Piloting lasted approximately twelve minutes. The second piloting session was conducted in a home environment with the parent present at all times. The child was asked the same questions as the first session. The second session lasted approximately ten minutes. Both children completed the test successfully.

Content Validation

The PEVOT underwent content validation and specific testing procedures to respond to the research questions: *What is the preschooler EFL receptive vocabulary knowledge based on the newly designed EFL vocabulary test and does test performance vary by sociodemographic characteristics such as participant age and geographical study region?* (Section 3.5).

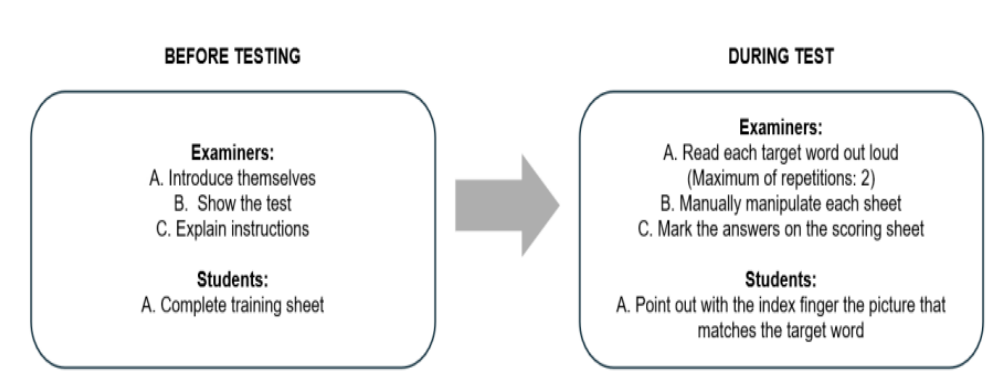
The content validation process with expert reviewers in preschool education ($n = 2$), and English teaching ($n = 3$) aimed to assess how the test measures receptive vocabulary knowledge, picture quality, and attractiveness for children (Güngör & Önder, 2023). To assess the measurability of receptive vocabulary knowledge, reviewers were asked if each sheet contributed to recognizing the target word. Additionally, they were asked if the picture for the semantic distractor worked as an accurate semantic distractor. For each picture, quality reviewers had to indicate if visual target on each sheet provided an accurate visual representation of the word, and if the pictures for the distractors also represented it accurately (See Annex). The attractiveness of the test for children was examined through one question asking if, overall, the item was appealing to children. Answers had to be provided on a 5-point Likert scale: *strongly disagree*, *disagree*, *neither agree nor disagree*, *agree*, *strongly agree* on a Google Form. Overall, reviewer responses validated the test content (please see Section 4.1 and Annex for complete results).

Testing Procedure

The test was administered in one session. Participants were tested individually in a school environment with no time constraints. The examiners introduced themselves in a friendly manner, showed the test, explained the instructions, and repeated them if necessary. Immediately after the instructions were provided, participants completed a practice sheet to become familiarized with the test format. After the practice sheet was completed, the test began with the examiner reading each target word aloud. The participant would then point to the picture that matched the target word with their middle index finger (Figure 3).

Figure 3.

Testing Procedure



Source: Author interpretation of the obtained results.

During the test, the examiner manually manipulated each sheet so that the participant only had to point with their index finger to the stipulated word. To make sure that participants understood the target word, the examiner could repeat the word twice. Each examiner had a scoring sheet to record the answers of each participant. A correct answer was scored with 1 and a wrong answer with 0. An

example of the scoring sheet can be found on the OSF page of the study. Each test lasted approximately 10 minutes. Parents provided informed consent before testing.

Data Analysis

The test was validated using inter-rater reliability analysis conducted in R with the irr package (Gamer et al., 2012). Inter-rater reliability was assessed using Fleiss' Kappa to measure agreement among test raters with regard to the content and appearance of the items. Five raters assessed the items using a 5-point Likert scale. Raters assessed (1) the accuracy of each target image in representing the target word (2) the accuracy of each distractor word, and (3) the extent to which the distractor word functioned as a semantically appropriate distractor for the target. To assess preschooler EFL vocabulary knowledge using the newly created EFL vocabulary test, we conducted t-tests in R (R Core Team, 2019).

RESULTS

This section presents the results organized according to the research questions.

Research Questions 1 and 2

Results of content and test validation and participant EFL receptive vocabulary knowledge are mentioned below to answer the research questions: *Can an EFL vocabulary test for children be designed based on the Costa Rican EFL preschool curriculum? If so, what is the receptive EFL vocabulary knowledge of preschoolers as measured by the newly developed test?*

Content and test validation

Fleiss' Kappa analysis was performed to analyze the results of PEVOT content and validation outcomes, revealing minimal agreement among the raters ($\kappa = 0.07$, $p < 0.01$). Given that the p -value < 0.05 , the null hypothesis is rejected, meaning that the agreement is statistically significant. However, higher consensus was observed among various elements (refer to the supplemental materials) which led to the Kappa result. The response to the question, "How appealing is the test for children, overall?" revealed that two raters considered the test very appealing, two found it appealing, and one rater indicated that it was somehow appealing. Raters generally determined the test to be appealing to the EFL study participants.

Vocabulary assessment of the full sample

Subsequently, preschooler EFL receptive vocabulary knowledge was evaluated using the newly designed assessment. Participants recognized approximately 55% of the PEVOT target words ($M = 11.02$, $SD = 4.09$) without any significant differences observed across blocks, $F(2, 2368) = 0.01$, $p = 0.99$, $\eta^2 = 0.01$. Table 1 shows the descriptive data for EFL vocabulary results.

A one-sample t-test was used to evaluate participant performance relative to chance. The chance level represents the probability of accurately choosing the correct answer by guessing. The PEVOT

design, which features four options with a single correct answer, results in a probability level of 25%. The results demonstrated that the mean score was markedly higher than chance $t(127) = 16.97, p < .0001$. This suggests that test performance exceeded the level anticipated from random guessing and that participants were actually able to identify EFL words on the vocabulary assessment.

A t-test was performed to examine potential variations in EFL vocabulary scores with regard to adjectives and nouns. The results indicated no significant differences $t(117) = 0.02, p < .9844$, demonstrating that participants could identify EFL words irrespective of the word type (i.e., adjective or noun).

Table 1.

Descriptive Statistics of the EFL Vocabulary Results per Study Block

Study Block	Mean	SD	Min	Max
Study Block I	10.68	3.99	2	20
Study Block II	10.68	3.98	2	20
Study Block III	10.69	3.97	2	20
Study Block IV	10.70	3.98	2	20

Source: Author interpretation of the obtained results.

Research Question 3

The impact of sociodemographic characteristics

Sections 4.3.1 and 4.3.2 reference the results that respond to research question 3: *Does test performance vary according to sociodemographic characteristics such as participant age and geographical region of the study?*

Age

Significant disparities in test scores based on the age of the participants were also examined. An independent-samples t-test was performed to compare test scores between 5 and 6-year-old children. The findings revealed a substantial disparity between the two groups, $t(117) = 5.78, p < .0001$, with 6-year-olds ($M = 14, SD = 3.28$) achieving higher scores compared to 5-year-olds ($M = 10, SD = 3.58$). This indicates that the test effectively reflects age-related influences potentially associated with developmental and educational variables..

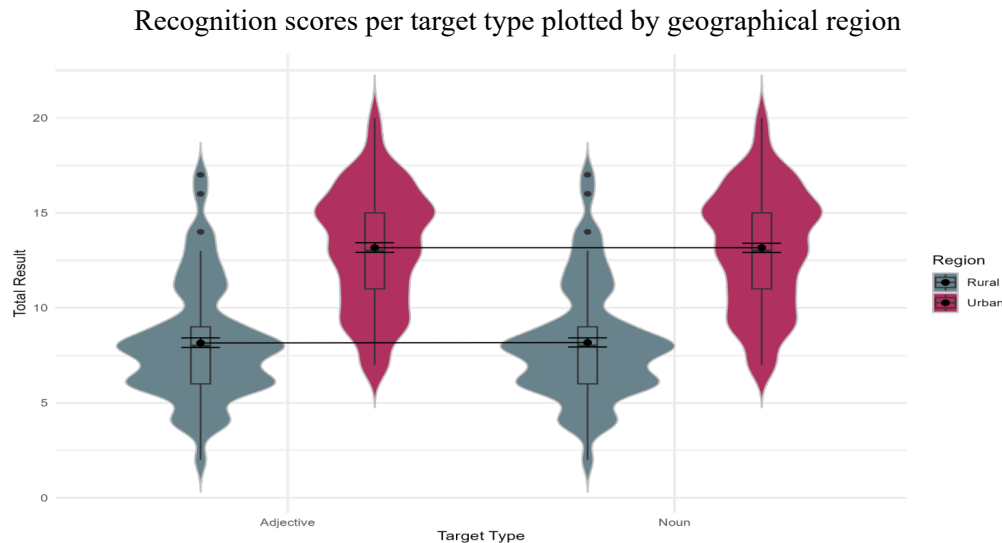
Geographical study region

Participants in urban areas recognized 65% of the target words on the EFL vocabulary test compared to 40% in rural areas. Two different one-sample t-tests were used to see if the PEVOT test results were substantially greater than the chance level for each region. The results show that the mean score in the urban region ($M=14, SD=3.2$) was considerably higher than chance $t(68) = 23.36, p < .0001$. Similarly, the mean score in the rural region ($M = 8, SD = 3.14$) was also considerably higher than chance $t(58)=7.33, p < .0001$. An independent-sample t-test was used to compare test scores between the

two locations. The results indicated that there was a significant difference between the regions $t(126) = 10.72, p < .0001$, with children from urban areas scoring higher scores than those from rural areas.

Figure 4 displays EFL vocabulary recognition scores by geographical region and target word.

Figure 4.



Note: Error bars surrounding the means represent 95 % confidence intervals; violin plots show data distributions.

Source: Author interpretation of the obtained results.

DISCUSSION

This study sought to design an EFL vocabulary test for preschoolers in Costa Rica and administer it to assess their EFL receptive vocabulary knowledge. Results showed that it is possible to create an EFL receptive vocabulary test for preschoolers in Costa Rica aligned with the Ministry of Education EFL preschool program. Although no significant age-related differences were identified in our sample, geographical region did influence the test performance of the participants.

Research Questions 1 and 2

A series of stages, including pre-piloting, piloting procedures (section 3.3.), and content validation (section 3.4) were conducted to address the research questions, *Can an EFL vocabulary test for children be designed based on the Costa Rican EFL preschool curriculum? If so, what is the receptive EFL vocabulary knowledge of the preschooler as measured by the newly developed test?* The results showed that it is possible to create an EFL vocabulary curriculum-based test that is appealing and comprehensible for children according to standard criteria (Nikolov & Timpe-Laughlin, 2021).

In terms of EFL vocabulary assessment, PEVOT results highlighted that children in Costa Rica can recognize EFL words that belong to the MEP current preschool EFL program. The four study

blocks included in the EFL curriculum provide vocabulary that children may acquire through in-class teaching activities in urban and rural areas. Ample range performance (min 2, max 20) indicates the potential of the test to identify individual variations. Monitoring the EFL progress of the students with an EFL vocabulary test provides an extraordinary opportunity to evaluate their language needs and development. No significant differences were found between noun and adjective learning.

Children begin to acquire vocabulary in their L1 by decoding linguistic and contextual input that they eventually transform into output (Tomasello, 2010). The first words that children typically articulate are mostly nouns since they are tangible and present in their surroundings (Tomasello, 2003). Children may be able to identify and verbally express themselves through words associated with food and toys, given their prevalence the everyday life of the child. Children tend to use words to designate object categories (Davies et al., 2022), potentially exposing them to more exposure to nouns compared to verbs and adjectives. Upon entering school, children begin to learn vocabulary from various parts of speech.

Participants in this study most likely shared a similar vocabulary knowledge of concrete nouns and adjectives in their first language (L1) that contributed to their L2 similarity in noun and adjective recognition. One may also argue that EFL didactic teaching strategies have improved EFL learning in children to such an extent that they can comparably recognize nouns and adjectives on an EFL test. More research into this area is needed to further explore the didactic knowledge of the ESL instructor and its possible implications in EFL teaching and learning in Costa Rica.

Research Question 3

Independent sample t-tests (section 4.3.1) and one-sample t-tests (section 4.3.2) were conducted to address research question 3, *Does test performance vary by sociodemographic characteristics such as participant age and geographical study region?*

Significant differences among 5 and 6-year-old children indicate that PEVOT adequately reflects age-related effects indicative of developmental changes, enhancements in cognitive abilities, heightened social engagement, instruction and accumulated exposure to foreign language. This suggests that the test may be used in the future to establish normative data to compare the individual performance of the child against that of their peers.

Consistent with our expectations, urban students performed better on the test than those in rural areas. Studying in urban areas increases exposure to the English language, with only 6.8% of the population in rural public preschools receiving English instruction, compared to 23.2% in urban areas (State of the Nation Report, 2021). Greater L1 literacy skills in urban areas (Conejo & Carmiol, 2018) along with limited EFL exposure in rural preschools, may have also contributed to less EFL word recognition in rural areas.

It is noteworthy that the effect of geographical setting on foreign language acquisition has been previously identified in Costa Rica with learners in rural areas demonstrating greater engagement towards their learning process (García-Castro & O'Reilly, 2022). Teachers and students in rural areas may be more committed to boosting EFL learning to mitigate the likelihood of experiencing poverty in these communities (State of the Nation Report, 2021), given the social and economic benefits of speaking English in Costa Rica (Abarca & Ramírez, 2017). The current study did not analyze the impact of increased motivation, which underscores the need for additional research on the topic.

CONCLUSIONS

This study involved creating an EFL receptive vocabulary test for preschoolers (PEVOT) and evaluated it with EFL learners from both urban and rural settings in Costa Rica. The study objective was to assess preschooler receptive EFL vocabulary knowledge using the newly designed test. The conclusion asserts that the PEVOT is a reliable instrument for assessing EFL receptive vocabulary knowledge at the preschool level in Costa Rica (research question 1). Developing an EFL vocabulary assessment for preschoolers is an innovative step forward in the field of EFL education in Costa Rica. Prior to this study, there was no tool to measure preschooler EFL vocabulary knowledge that matched the national curricula. Developing and using a contextualized EFL vocabulary test provides valuable insights into national EFL education in Costa Rica.

The PEVOT can be used to test the first stages of EFL vocabulary knowledge (research question 2) among 5- and 6-year-olds given that test performance does not significantly vary by age; however, it should be noted that students in urban areas may perform better than those in rural areas (research question 3). Assessment of preschooler vocabulary knowledge, through the PEVOT, could help establish a solid foundation for EFL proficiency to support the ongoing advancement of EFL education from an early age. For instance, the creation of the PEVOT might help diagnose existing gaps in the national EFL curricula, particularly in rural areas. If approached during initial years of primary education, such gaps could decrease which would better prepare EFL students to thrive in an increasingly interconnected world. It is essential to continue to strengthen early childhood assessments in the country in order to identify gaps, monitor student progress, improve the Costa Rican EFL curriculum, and guide future pedagogical investments.

Studies are not immune to constraints. The number of participants in this study may have limited the results obtained. It is necessary to replicate this study in a larger sample from diverse regions in Costa Rica. The low Fleiss' Kappa value indicates significant but weak agreement among raters. This may stem from several factors, including a limited number of raters and the inherent subjectivity of the rating criteria. Variations in expertise and personal biases likely contributed to discrepancies in assessments, highlighting the challenge of achieving consistent evaluations in subjective judgments of this sort.

Future research should contemplate EFL expert teachers perspectives and didactic knowledge to attain a broader picture of EFL vocabulary development. Findings from this study should be replicated with a larger EFL population in urban and rural areas in Costa Rica. It is highly recommended to create an EFL vocabulary test for primary and high-school populations to better assess the national EFL curriculum and student performance.

Funding

This work was funded by the Universidad de Costa Rica, Regular Fund for Research Projects, C3311, and a grant from the Women's Research Support Fund from the Office of the Vice Rector for Research at the University of Costa Rica, Research Project C 3329.

REFERENCES

- Abarca, A., & Ramírez, S. (2017). *The labor market benefits of bilingualism: The case of Costa Rica*. Observatorio del Desarrollo de la Universidad de Costa Rica.
- García-Castro, V., & O'Reilly, J. (2022). Foreign Language Anxiety and Online Engagement During the COVID-19 Pandemic: a Comparison Between EMI and FMI University Students. *English Teaching & Learning, 46*, 273–291. <https://doi.org/10.1007/s42321-022-00120-x>
- Barquero, K., Carmiol, A. M., León, J., Murillo, D., & Román, M. (2019). *Resumen Estado de la Educación Costarricense*. [Summary of the State of Costa Rican Education]. Programa Estado de la Nación [State of the Nation Report.]. (PEN). <https://estadonacion.or.cr/wp-content/uploads/2019/08/Estado-Educacio%CC%81n-RESUMEN-2019-WEB.pdf>
- Bergström, K., Klätte, M., Steinbrink, C., & Lachmann, T. (2016). First and Second Language Acquisition in German Children Attending a Kindergarten Immersion Program: A Combined Longitudinal and Cross-Sectional Study. *Language Learning, 66*(2), 386-418. <https://onlinelibrary.wiley.com/doi/10.1111/lang.12162>
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology, 96*(1), 31-42. <https://doi.org/10.1037/0022-0663.96.1.31>
- Conejo, L. D., & Carmiol, A. M. (2018). *La lectoescritura emergente en la educación preescolar costarricense y su relación con las concepciones y prácticas docentes: Un estudio representativo en aulas de transición de tres direcciones regionales del Ministerio de Educación Pública: Ponencia preparada para el Séptimo Informe Estado de la Educación* [Emergent literacy in Costa Rican preschool education and its relationship with teaching concepts and practices: A representative study in transition classrooms of three regional directorates of the Ministry of Public Education: Presentation prepared for the Seventh State of Education Report]. Programa Estado de la Nación (PEN). <https://hdl.handle.net/20.500.12337/7764>
- Davies, C., Syrett, K., Taylor, L., Wilkes, S., & Zuniga-Montanez, C. (2022). Supporting adjective learning across the curriculum by 5–7 year-olds: Insights from psychological research. *Language and Linguistics Compass, 16*(11), e12476. <https://doi.org/10.1111/lnc3.12476>

- Dunn, L., & Dunn, D. (2007). *Peabody Picture Vocabulary Test*. Fourth Edition. <https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Peabody-Picture-Vocabulary-Test-%7C-Fourth-Edition/p/100000501.htm>
- Eigsti, I.-M. (2017). Peabody Picture Vocabulary Test. In F. R. Volkmar (Ed.), *Encyclopedia of Autism Spectrum Disorders* (pp. 1–5). Springer New York. https://doi.org/10.1007/978-1-4614-6435-8_531-3
- Ferjan-Ramírez, N., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational turns and advances infant language development. *Proceedings of the National Academy of Sciences*, 117(7), 3484-3491. <https://doi.org/10.1073/pnas.1921653117>
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). *irr: Various Coefficients of Interrater Reliability and Agreement* [R package version 0.84]. CRAN.R. <https://CRAN.R-project.org/package=irr>
- Goriot, C., Van Hout, R., Broersma, M., Lobo, V., McQueen, J. M., & Unsworth, S. (2021). Using the Peabody picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, 24(4), 546-568. <https://doi.org/10.1080/13670050.2018.1494131>
- Güngör, B., & Önder, A. (2023). Development of English Picture Vocabulary Test as an Assessment Tool for Very Young EFL Learners' Receptive and Expressive Language Skills. *Early Education and Development*, 34(2), 572-589. <https://doi.org/10.1080/10409289.2022.2043134>
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55-88. <https://doi.org/10.1016/j.dr.2005.11.002>
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, 61(4), 343-365. <https://doi.org/10.1016/j.cogpsych.2010.08.002>
- Lakshmanan, U. (2013). Child Second Language Acquisition. In P. Robinson (Ed), *The Routledge encyclopedia of second language acquisition* (pp. 71-74). Taylor & Francis.
- Mihaljević-Djigunović, J., Nikolov, M., & Ottó, I. (2008). A comparative study of Croatian and Hungarian EFL students. *Language Teaching Research*, 12(3), 433-452. <https://doi.org/10.1177/1362168808089926>
- Ministerio de Educación Pública [MEP]. (2007). *Programa de estudios inglés preescolar en el ciclo de Transición* [Preschool English Study Program in the Transition Cycle]. MEP. <https://www.mep.go.cr/sites/default/files/media/inglestransicion.pdf>
- Murillo-Rojas, M. (2012). La evaluación del vocabulario en la educación preescolar: El Tevopreesc [Vocabulary assessment in preschool education: The Tevopreesc]. *Revista Káñina*, 36, 151-162. <https://www.redalyc.org/articulo.oa?id=44249254016>
- Nakatsuhara, F., Taylor, L., & Jaiyote, S. (2020). The Role of the L1 in Testing L2 English. In C. J. Hall & R. Wicaksono (Eds.), *Ontologies of English: Conceptualising the Language for Learning, Teaching, and Assessment* (pp. 187–208). chapter, Cambridge: Cambridge University Press.

- Nikolov, M. (2016). Trends, issues, and challenges in assessing young language learners. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 1-18). Springer.
- Nikolov, M., & Timpe-Laughlin, V. (2021). Assessing young learners' foreign language abilities. *Language Teaching*, 54(1), 1-37. <https://doi.org/10.1017/S0261444820000294>
- Programa Estado de la Nación. (2021). *Octavo Estado de la Educación* [Eighth State of Education]. Programa Estado de la Nación (PEN). https://estadonacion.or.cr/wp-content/uploads/2021/09/Lanzamiento_ee.pdf
- R Core Team. (2019). *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Richards, J. C., & Schmidt, R. W. (2013). *Longman Dictionary of Language Teaching and Applied Linguistics* (4th ed.). Routledge. <https://doi.org/10.4324/9781315833835>
- Romeo, R. R., Segaran, J., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Yendiki, A., Rowe, M. L., & Gabrieli, J. D. E. (2018). Language Exposure Relates to Structural Neural Connectivity in Childhood. *The Journal of Neuroscience*, 38(36), 7870-7877. <https://doi.org/10.1523/JNEUROSCI.0484-18.2018>
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363. <https://doi.org/10.1177/1362168808089921>
- Schmitt, N. (2010). *Researching Vocabulary*. Palgrave Macmillan UK. <https://doi.org/10.1057/9780230293977>
- Stein, M. T., & Lukasik, M. K. (2009). Developmental Screening and Assessment: Infants, Toddlers, and Preschoolers. In W. B. Carey, A. C. Crocker, W. L. Coleman, E. R. Elias, & H. M. Feldman (Eds.), *Developmental-Behavioral Pediatrics* (4th Ed., pp. 785-796). W.B. Saunders. <https://doi.org/10.1016/B978-1-4160-3370-7.00079-1>
- Taylor, L., & Saville, N. (2002). *Developing English tests for young learners* (Research Notes No. 7). University of Cambridge Local Examinations Syndicate. <https://www.cambridgeenglish.org/images/23119-research-notes-07.pdf>
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press. <https://doi.org/10.2307/j.ctv26070v8>
- Tomasello, M. (2010). *Cognitive Linguistics and First Language Acquisition*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199738632.013.0041>

ANNEX

Table A1.

Descriptive statistics of the rating assessing target and distractor's content and appearance

Item	<i>M</i>	<i>SD</i>	Min	Max
Training_target_picture_accuracy	4.8	0.45	4	5
Training_distractor_picture_accuracy	4.4	0.89	3	5

Training_distractor_semantic_accuracy	4.6	0.55	4	6
Target_shoes_picture_accuracy	5	0	5	5
Target_shoes_distractor_picture_accuracy	4.8	0.45	4	5
Target_shoes_distractor_semantic_accuracy	4.6	0.55	4	5
Target_friends_picture_accuracy	4.6	0.89	3	5
Target_friends_distractor_picture_accuracy	4.6	0.89	3	5
Target_friends_distractor_semantic_accuracy	5	0	5	5
Target_cat_picture_accuracy	4.8	0.45	4	5
Target_cat_distractor_picture_accuracy	5	0	5	5
Target_cat_distractor_semantic_accuracy	4	1.22	2	5
Target_dog_picture_accuracy	4.6	0.55	4	5
Target_dog_distractor_picture_accuracy	4.4	0.89	3	5
Target_dog_distractor_semantic_accuracy	4.4	0.55	4	5
Target_doll_picture_accuracy	5	0	5	5
Target_doll_distractor_picture_accuracy	5	0	5	5
Target_doll_distractor_semantic_accuracy	3.8	1.30	2	5
Target_ball_picture_accuracy	5	0	5	5
Target_ball_distractor_picture_accuracy	4.8	0.45	4	5
Target_ball_distractor_semantic_accuracy	3.8	1.64	1	5
Target_eyes_picture_accuracy	5	0	5	5
Target_eyes_distractor_picture_accuracy	5	0	5	5
Target_eyes_distractor_semantic_accuracy	4.2	0.84	3	5
Target_house_picture_accuracy	4.4	1.34	2	5
Target_house_distractor_picture_accuracy	4.4	0.55	4	5
Target_house_semantic_accuracy	4.2	0.45	4	5
Target_table_picture_accuracy	5	0	5	5
Target_table_distractor_picture_accuracy	4.8	0.45	4	5
Target_table_semantic_accuracy	4.4	0.55	4	5
Target_flowers_picture_accuracy	5	0	5	5
Target_flowers_distractor_picture_accuracy	4.4	0.89	3	5
Target_flowers_semantic_accuracy	4.2	1.30	2	5
Target_black_picture_accuracy	5	0	5	5
Target_black_distractor_picture_accuracy	3.6	1.95	1	5
Target_black_semantic_accuracy	4.2	1.10	3	5
Target_happy_picture_accuracy	4.8	0.45	4	5
Target_happy_distractor_picture_accuracy	4.8	0.45	4	5
Target_happy_semantic_accuracy	4.2	0.84	3	5
Target_grey_picture_accuracy	5	0	5	5
Target_grey_distractor_picture_accuracy	4.4	1.34	2	5

Target_grey_semantic_accuracy	4	1	3	5
Target_brown_picture_accuracy	4.6	0.55	4	5
Target_brown_distractor_picture_accuracy	4.8	0.45	4	5
Target_brown_semantic_accuracy	4.4	0.89	3	5
Target_small_picture_accuracy	4.4	0.55	4	5
Target_small_distractor_picture_accuracy	4	1.22	2	5
Target_small_semantic_accuracy	4	1.22	2	5
Target_blue_picture_accuracy	5	0	5	5
Target_blue_distractor_accuracy	5	0	5	5
Target_blue_semantic_picture_accuracy	4	1.41	2	5
Target_big_picture_accuracy	4.4	0.55	4	5
Target_big_distractor_picture_accuracy	3.2	1.64	1	5
Target_big_semantic_picture_accuracy	4.6	0.55	4	5
Target_tall_picture_accuracy	4.6	0.55	4	5
Target_tall_distractor_picture_accuracy	4.6	0.55	4	5
Target_tall_semantic_picture_accuracy	4.8	0.45	4	5
Target_round_picture_accuracy	5	0	5	5
Target_round_distractor_picture_accuracy	5	0	5	5
Target_round_semantic_picture_accuracy	4.6	0.55	4	5
Target_yellow_picture_accuracy	5	0	5	5
Target_yellow_distractor_picture_accuracy	5	0	5	5
Target_yellow_semantic_picture_accuracy	4	1	3	5