



Development of English Picture Vocabulary Test as an Assessment Tool for Very Young EFL Learners' Receptive and Expressive Language Skills

Burcu Güngör & Alev Önder

To cite this article: Burcu Güngör & Alev Önder (2022): Development of English Picture Vocabulary Test as an Assessment Tool for Very Young EFL Learners' Receptive and Expressive Language Skills, *Early Education and Development*, DOI: [10.1080/10409289.2022.2043134](https://doi.org/10.1080/10409289.2022.2043134)

To link to this article: <https://doi.org/10.1080/10409289.2022.2043134>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 06 Mar 2022.



Submit your article to this journal [↗](#)



Article views: 947



View related articles [↗](#)



View Crossmark data [↗](#)

Development of English Picture Vocabulary Test as an Assessment Tool for Very Young EFL Learners' Receptive and Expressive Language Skills

Burcu Güngör ^a and Alev Önder^b

^aEarly Childhood Education, İstanbul Sabahattin Zaim University; ^bEarly Childhood Education, Bahçeşehir University

ABSTRACT

Research Findings: The aim of this study was to construct and validate “English Picture Vocabulary Test (EPVT)” that aimed to assess the very young learners’ (VYLs) receptive and expressive vocabulary knowledge for specific content areas in English as a foreign language (EFL). In this context, EPVT was created in several stages. One of them was initial construct identification with the literature review on various aspects of assessment in early year’s FL learning and development aspects of very young EFL learners. The other stages were “pre-piloting” with 4 experts and 20 VYLs to obtain their perspectives on the test construction and administration and “final piloting” implementation with 5–6 years old private pre-primary school children (251) to ensure reliability, validity, and test quality. After designing the process of the EPVT was clarified considering the ethical and effectiveness issues, the item analysis with Kuder-Richardson 20 and Point-Biserial Correlation was conducted. Besides, the distribution of the items among the ranges of difficulty and the discrimination index values were specified. *Practice and Policy:* The findings support that EPVT is child-friendly and effective in assessing preschoolers’ English vocabulary receptively and expressively. Furthermore, it could also serve to provide feedback for teaching and supporting for learning L2.

The issues of English language assessment at pre-primary level have been growing interest all over the world in parallel with the early English language education. Despite some tiny improvements, it is possible to say that almost no standards have been clearly stated outlining the required level of language skills at this level and suitable assessment tools accordingly in the early language learning assessment (Nikolov, 2016). *Very Young Learners (VYLs)* is a well-used term referring to pre-primary level EFL language learners who are below the age of formal entry into compulsory education, usually under 6 years (Mihaljević Djigunović, 2016; Nikolov, 2016). The other terms we have opted to employ throughout the study are *L1* which refers to mother tongue or first language and *L2* that refers to foreign language where children learn English as a subject in the school in limited hours. The ones who are English learners in Turkey set an example for EFL learners. English that holds the status of a foreign language mostly in Turkey is compulsory at all levels of education; however, L2 education at pre-primary level is not compulsory. It is clear that there is a growing interest in L2 education at pre-primary level. More intensive programs introduce English as the lessons in the school curriculum or less intensive ones as extracurricular English courses and club activities after school hours. Although assessment is fundamental to the success of early English learning, how teachers assess progress and attainment in English at pre-primary level is still something of a blind spot in Turkey.

CONTACT Burcu Güngör  burcu.gungor@izu.edu.tr  Early Childhood Education, İstanbul Sabahattin Zaim University, İstanbul, Turkey

This study is derived from Burcu Güngör’s PhD dissertation and part of it was presented in 6th International Early Childhood Education Congress in Kars, 2–5 October, 2019

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

It is indicated in the young L2 English literature that the VYLs have distinctive characteristics from adult learners (Cameron, 2001). For this reason, early L2 assessment needs to take into consideration factors, such as VYLs' shorter attention span (Robert et al., 2009), developing literacy, and limited experience of the world. The other factors that have effect on developing reliable and age-appropriate assessment instruments can be listed as their low-level foreign language knowledge in EFL contexts and developmental characteristics (McKay, 2006).

Since the quality and quantity of exposure to L2 is varying at a certain degree due to the amount of time spent in L2 introduction and teacher's qualifications in pre-primary schools, it is unlikely to measure learning outcomes successfully in a range of very different situations (Nikolov, 2016). Rixon (2013) emphasized a variety of L2 education models and programs in pre-primary settings in which children have various L2 experiences in quantity and in quality. In the early L2 context, three types of curricula are mentioned in the literature (Edelenbos et al., 2006; Johnstone, 2009). Initially, the *Awareness Raising Programs* which are not concentrated on one additional language alone, then *Traditional FL Programs* offering one to a few classes per week around midway, *Content and Language Integrated Learning (CLIL)* curricula toward the end. In these programs, children are not expected to achieve native level (Inbar-Lourie & Shohamy, 2009). Based on these curricula types, it appears to be particularly problematic to develop a global assessment which fits the richness of content aimed at very young learners and with different learning needs worldwide. However, it is notable that the common age-appropriate achievement targets which are in conformity with very young learners' characteristics and literacy level are related with their basic vocabulary knowledge in different types of curricula and course books for VYLs. It is well known that for the language areas assessed, most of the assessment literature prioritize testing young children's vocabulary knowledge (Nikolov, 2016).

In the English as a foreign language (EFL) context, children start to learn a new language with vocabulary that has been considered the major resource for language use. Related to this, Hestetræet (2019) and Nation (2013) stated that building up a useful vocabulary is central for foreign language learning at earlier ages. For these reasons, vocabulary which is seen as a priority area in L2 education at pre-primary level requires tests to monitor the VYLs' progress in vocabulary learning. Consideration of the range of the vocabulary assessment practices in the developing field of teaching English to pre-primary school-aged children is therefore surely of high relevance. Testing vocabulary knowledge in early L2 programs is mostly often concerned with how VYLs progress and what levels of proficiency they achieve in their L2 vocabulary development by the end of certain periods (Nikolov, 2016).

When it comes to assessing preschool children's performance on L2 vocabulary knowledge effectively to provide valuable information to researchers, educators, parents, and administrators, it should be considered what kind of assessment formats and methods are age-appropriate to measure their ability to use and understand theme-specific vocabulary in their oral language at this level. Under the wide umbrella of best assessment practices, receptive and expressive vocabulary tests in which children are asked to recognize or express any target vocabulary with the help of pictures occupy a primary role in pre-primary level English language education.

Assessment of the Vocabulary Knowledge of Preschool Language Learners

According to Harmer's (2015) and Common European Framework of References for Languages (CEFR; Council of Europe, 2018) model of language level, pre-primary children can be labeled as false beginners or absolute beginners with insufficient previous exposure to English at all in Turkey and most of other countries where English is taught as a foreign language. Among the communicative and linguistic aims, the acquisition of basic vocabulary in context and in a meaningful way has been accentuated as a priority in early language learning at this age (Černá, 2015; European Commission, 2011). In a discussion of the effective and relevant L2 vocabulary instruction at pre-primary level, the starting point is to make VYLs become conscious of theme-specific English vocabulary and its meaning and feel encouraged to activate the words to a reasonable degree.

As well known, it is necessary to understand more accurately what vocabulary is to be measured in early L2 assessment and to what extent the VYLs will be expected to perform to demonstrate mastery of the objective before a test can be organized to measure the learning objective. Based on this, it can be said that the selection of theme and target vocabulary is the cornerstone of the development of appropriate assessment tools based on the young children's characteristics. Analysis of objectives to determine the level of understanding is commonly done by constructing a table of specifications (Linn & Miller, 2005). According to this table, there are four levels – recall (recite, recall, tell, state, memorize), skill/concept (separate, classify, construct, summarize), strategic thinking (assess, compare, critique, formulate), extended thinking (design, create, prove) – which demonstrate the depth of the knowledge levels. In reference to this, the first level points out the measurement of the extent VYLs can recall and retell the target vocabulary.

For very young EFL children who are at the concrete stages of cognitive development, the themes at this level should be concrete (i.e., “body parts, animals, fruit” instead of the structure of the brain; Bourke, 2006). It means that young children had a schema in their head when they encountered curricula-related or theme-related vocabularies that were included in the test. Regarding this issue, Kail (2010) indicated that salient or familiar themes are helpful in focusing very young EFL children's attention. Research shows that foreign language education programs that include familiar themes from VYLs' mainstream early childhood education result in more successful learning outcomes because VYLs make some connections and their prior knowledge also makes the L2 learning and assessment easier (Bacsa & Csikos, 2016; Nikolov, 2009a).

Regarding the suitable test formats used for VYLs' assessment by considering their characteristics and EFL context, Hestetræet (2019) discussed the need for young children to develop an age-appropriate vocabulary through focusing on the use of word receptively and expressively. With reference to Nation's (1990) framework about the types of vocabulary knowledge, first two levels – receptive and expressive – are more practical to present the basic level concepts and items. Considering all these issues, the assessment of L2 vocabulary knowledge at pre-primary level by allowing children to be exposed to English receptively and expressively is one of the wide-ranging and influential practices (Connery et al., 2010). Regarding the young children's early stages of L1 and L2 learning process, it is stated that the first stage is word recognition in which children learn the skills to recognize words and the second stage is the production in which they activate and express the words (Gibson et al., 2012; Kokla, 2013; Mondria & Wiersma, 2004). These points so far clarified that VYLs – beginning EFL learners – are expected to be able to recognize familiar words and recall them (Thornbury, 2016).

The Challenge of Developing Vocabulary Test for Preschool Language Learners

To develop age-appropriate tools and implement best practices in early EFL assessment, it is essential to have solid understanding of what special principles and approaches are needed. They are mainly listed by Conteh (2012) as:

- deciding the purpose of assessment, distinguishing summative, formative and diagnostic aims.
- selecting/designing the developmentally and culturally appropriate and familiar themes, tools, and methods in assessment.
- taking into consideration the amount of time allocated for English instruction in early childhood education.
- considering the developmental characteristics, skills, biological predisposition, and motivation of VYLs who are quite different from adults (Pinter, 2012; Zandian, 2012).
- focusing on English learning goals at the very earliest stages of L2 education.

Many recent researches have demonstrated that early vocabulary knowledge predicts later reading achievement (Stæhr, 2008) and forms the basis for communicative skills in a foreign language (Richards & Rodgers, 2014). These results suggest that providing developmentally appropriate

frameworks for L2 vocabulary learning and assessment become a priority in early year. Considering the importance of developing a large and functional L2 vocabulary in early years English language instruction (Hestetraet, 2019; Webb & Nation, 2017), it seems vital to start learning English from age-appropriate vocabulary.

In case of vocabulary assessment at the beginning of children's English educational experience, some criteria should be taken into consideration. First, the process of setting the assessment task and selecting vocabulary to the level of VYLs is essential for useful outcomes to emerge. The English level of children at preschool age is generally at the beginner's stage in EFL contexts where the programs generally provide children a limited amount of exposure to the foreign language in pre-primary school setting. Second, VYLs' special characteristics which are collected into three headings by McKay (2006) as *growth*, *literacy*, and *vulnerability* need to be taken into consideration.

For instance, VYLs' well-known characteristics are their having short attention spans (Brewster et al., 2002; Cameron, 2001; Ellis, 2014; Slattery & Willis, 2001). For this reason, children at this stage cannot concentrate on one task for long periods of time. Therefore, they need variety and time limitation in assessment tasks (Garton & Copland, 2019). Another criterion is that preschool language learners have an imprecise mastery of their first language. The other significant point to consider when deciding the target vocabulary is selecting age-appropriate vocabulary that relates to the children's cognitive development. Cameron (2001) clarified this by offering three different levels which are basic, superordinate, and subordinate vocabulary levels in the early L2 learning field. Cameron (2001) indicated that the concepts from the former one are more likely to have been mastered than others. Based on this, to start from the basic level items which include words from children's environment (cat, table, book, shoe, water, etc.) while introducing L2 vocabulary for beginners is more practical and suitable than general vocabulary (i.e., vegetables and animals) and more specific words (i.e., ragdoll cat, danvers). The other crucial factor concerns choosing age-appropriate vocabulary in the sense that the children find it meaningful (Richards & Rodgers, 2014).

Whereas they have developed voluntary attention which allows them to focus on assessment, involuntary attention can be easily triggered by internal or external stimuli, such as hunger, light, color, noise, and tiredness, and may quickly distract children from a set task. Regarding this, Wesson (2011) asserted that the maximum time for VYLs' focused attention during instruction is up to 15–20 min duration, providing the task is engaging and commands their interest. It is also implied that VYLs are in the pre-literacy period; that's why, VYLs' assessment requires special consideration at the time of deciding on the appropriate assessment tasks, such as picture-matching and multiple-choice tasks. Put simply, it is much better to convey the messages with the help of photos, pictures, or flashcards in the early years instead of writing.

As for the vulnerability, positive reinforcement after their response even in their incorrect responses during the assessment process plays a critical role in valid and fair results in early L2 assessment. Regarding this, Mistry and Sood (2015) asserted that young children having a high vulnerability to approval, praise and criticism should be taken into consideration in L2 assessment process at pre-primary school level. They are also vulnerable to ineffective and culturally unsupportive instruction and assessment. Considering these issues, some assessment procedures – giving time to think, repeating the questions twice and encouraging and motivating young learners – need to be taken into consideration in implementing age-appropriate assessments. Related to this issue, the findings of Biricik and Özkan's study (2012) revealed that VYLs need more external motivation contrary to adults who have a pre-existing set of motivations in the FL assessment process because of their immediate or pragmatic need of the L2. According to the relevant ethical guidelines on research with young language learners' assessment, young children should not be reprehended for the wrong answers.

Although there are many standardized tests – *Peabody Picture Vocabulary Test (PPVT)*; Dunn & Dunn, 2007), *Expressive One-Word Vocabulary Test (EOWPVT)*; Brownell, 2000), and *Receptive One-Word Vocabulary Test (ROWPVT)*; Brownell, 2000) – aiming at providing a comprehensive assessment

of learners' general vocabulary knowledge, they are not sufficient to meet the requirements of an assessment tool that can be used to measure the competence in children's specific L2 vocabulary knowledge at pre-primary EFL learning contexts. One of the main reasons for this is because the themes and items in those tests are beyond the scope of generic themes and vocabulary offered in English curricula at pre-primary level. Regarding this, many research findings emphasized that the PPVT is inappropriate for use with L2 learners with limited L2 proficiency (Goriot et al., 2021). In other words, the finding indicated that it is not a reliable measure for pupils who were young unexperienced L2 learners in English. At this point, it is useful to clarify some of the terminology to understand different assessment contexts in early years language learning.

When the countries' assessment procedures of children's L2 learning at pre-primary and primary level are examined, three different scenarios have emerged. In the first scenario, English is introduced intensely through Content and Language Integrated Learning (using L2 as a vehicle to deliver content knowledge and target domain-specific skills) or Content-based Instruction (integrating age/level of schooling with language and academic content in core subject areas, such as maths, science, social studies, and language arts). A range and variety of alternative assessment tools, such as the collection of children's works in a portfolio (Cyprus; Rixon, 2013), self-assessment, peer-assessment and observation and written description of learner performance (Germany) (Kubanek-German, 2000), (France; Rixon, 2013), are used here. In the second scenario, with the advent of standards-based assessment, some South American and European countries have developed their own national EFL examinations for young learners in Germany (Rupp et al., 2008); in Norway (Hasselgren, 2005); in Slovenia (Pizorn, 2009); in Hungary (Nikolov & Szabó, 2012); in Switzerland (Haenni Hoti et al., 2009); in Poland (Szpotowicz & Campfield, 2016); in Uruguay (Fleurquin, 2003).

In the last scenario, the ones introducing L2 as a school subject in EFL contexts in limited hours tend to use formal assessment procedures, such as tests produced by the class teacher, tests given in the textbook used in class or classroom assessment. Based on this, it can be said that the assessment can vary in early years language learning contexts. To put it more explicitly, some large-scale standardized tests (PPVT, EOWPVT, ROWPVT) mentioned above can be used to compare children's vocabulary breadth in various European countries, using English as the language of instruction in addition to the official language in education (Goriot et al., 2021). However, they are not helpful in gauging children's vocabulary development in receptive and productive ways in content areas specified in many pre-primary level course-books and curricula in EFL contexts (National Institute of Child Health and Human Development (NICHD), 2000). The aforementioned vocabulary tests can be used as reliable tool to test English vocabulary in very young learners when they have more knowledge of English but not in VYLs who have limited exposure to English.

Despite the growing importance of VYLs' assessment, it is noticeable that the assessment is a relatively new issue even for children who are learning English at the pre-primary level. It is apparent that the use of assessment at pre-primary level for VYLs' English level is not given the place it deserves in the literature (Nikolov, 2016). There is unwillingness among teachers and curriculum planners to administer tests or describe progress in a systematic fashion at this level. Despite this indifference to early L2 assessment, it can be said that it has a critical role to gain information about the progress and attainment of children to be responsive both to learner needs and curricular demands whether the assessment is informal, formal, classroom-based, or large-scale.

In the process of designing, piloting, and validating a picture vocabulary test in accordance with the literature review on various aspects of developmental framework of VYLs' assessment, the issues taken into consideration are (1) how to create entertaining and age-appropriate test items bearing in mind the volume of common vocabulary suggested in the coursebooks, textbooks for VYLs and early English language curricula and CEFR; (2) how to reconcile VYLs' developmental characteristics with their low level of foreign language knowledge in order to develop and implement vocabulary test; (3) how to encourage willing participation and intellectual engagement with vocabulary test during the implementation process. As selecting or developing suitable assessment tools is extremely important in the process of early childhood foreign language learning, some criteria should be taken

into consideration. One of them is that the tests for VYLs need to have desirable test qualities, validity, positive impact, reliability. Based on a review of the literature on validity, some criteria such as checking the appropriateness of the purpose, the practicality of test to improve learning, the universality of the test to make reliable assessment decisions, the utility of the test for children to show the required knowledge, understanding knowledge and skills are suggested (Kane, 2013). Another issue is assessment's fitness for purpose, which includes how well it motivates learners to learn English and/or how well it gives information about the VYLs' L2 learning process. In achieving this, realistic achievement targets related to reception, production are required to be set to measure VYLs' vocabulary knowledge.

The Current Study

Prosi-Santovac and Rixon (2019) asserted that the vocabulary assessment tools are needed at preschool level to have some understanding of the progress VYLs are making. In examining this, it should be kept in mind that *assessment in English as foreign language contexts* and *assessment in English as a second language or content-based instruction* is different from each other. The former refers to assessment conducted in EFL countries where the tests and alternative assessment tools include well-designed restricted response items (e.g., “multiple choice questions, short answer questions, matching”) and aim at measuring both their linguistic knowledge at word and phrase levels (Papp, 2019).

Even though assessing very young learners of an FL is fundamental to the success of early English learning, it is a complex area requiring knowledge of English language acquisition, early childhood education, and research methodology in the national and international area (Nikolov, 2016). This might be due to the lack of clear policy decisions and effective studies on the assessment in English and the scarcity of assessment tools at this level. The other reason can be ethical considerations that should be forefront of all the assessment activity involving very young learners (Rixon, 2013). The deficit in reliable and comprehensible assessment tools measuring the children's vocabulary knowledge at pre-primary level is the reason why the researcher has decided to deal with the issue of “assessing VYLs” of English' which has been underrated in the field of L2 assessment.

Contrary to some standardized tests aiming at providing a comprehensive assessment of learners' general vocabulary knowledge and verbal ability by including more items, the general aim of the current study is to develop an “English Picture Vocabulary Test” and to evaluate its validity and reliability for normal 60–72 months Turkish children keeping in mind both the VYLs' characteristics, learning needs and desirable test qualities with ethical guidelines.

Methods and Materials

Participants

Pre-primary education, which is provided in both state-run and private institutions in Turkey, is not compulsory. In many EFL contexts like in Turkey, English has not yet been established as part of the state pre-primary curriculum. A relatively small percentage of the state pre-primary schools provide English instruction as a free-time, after-school, or club activity. On the other hand, private pre-primary school L2 programs in Turkey can be grouped as “High-Intensity,” “Moderate-Intensity,” and “Low-Intensity” that differ in intensity and exposure to the target language.

The setting for this study was 16 private pre-primary schools from different socio-economic backgrounds (low, middle high) in three regions of Istanbul. Doğançay-Aktuna and Kızıltepe (2005) illustrated that the quality and extent of English instruction is mostly related with socio-economic factors. Based on this, to obtain a representative target group for final piloting, 251 children aged 5 and 6 years (112 males and 139 females) ($M = 5$ years and 6 months) from high, moderate, and low English levels were included evenly according to the data coming from Turkish Statistics Institution. Therefore, the research population were private pre-primary school children who have been exposed

to English in various degrees in Turkey. To meet the ethical requirements of research studies, all parents of the participants were asked for informed consent. In doing this, parents were sent an information letter including a broad description of the study and its aims and were asked to consent to their child taking part in the study.

The Design of the English Picture Vocabulary Test and Pilot Studies

EPVT was developed, piloted, and validated in several stages. To test the validity and reliability of EPVT, the following techniques – literature review, content validity, expert opinion, target age group validity, and item analysis – were ordered.

Stage 1

Firstly, a receptive and expressive one-word picture vocabulary test format was decided to assess VYLs' vocabulary knowledge after a close scrutiny of literature review. This is because young children's developmental realities pose unique and challenging considerations. To illustrate, pre-primary school children have illiteracy problems, lack of opportunities to hear and use the language outside the classroom (insufficient input/output), limited amount of time dedicated to the English language in EFL contexts where they mainly focus on developing and practicing target vocabulary receptively and productively (Nikolov, 2016). As for the construct of early language learning, Inbar-Lourie and Shohamy (2009) suggest that L2 assessment should be closely linked to overall early L2 program (focusing on language or content), should be closely linked to overall early L2 program (focusing on language or content) and learners' L1. Related to this, McKay (2006) and Taylor and Saville (2002), construct validity should be ensured by in-depth analysis of curricula and course-books, teaching and assessment practices in EFL context, and young children's cognitive and affective development. In doing this, the researcher decided the constructs are to be assessed by in-depth analysis of the research (Beck et al., 2002; Cameron, 2001; Nikolov, 2016) related to vocabulary instruction and assessment in foreign language, National English Curriculum for Private Preschools in Turkey, which was accepted in 2016, *can do statements* in the CEFR (Council of Europe, 2018), compatible with VYLs' performances, various course books, and word list picture books.

Stage 2

The researcher decided on the quality and quantity of the target words included and assessed in the test. In doing the item construction in EPVT, the criterion suggested by Nikolov (2016) was guided and target vocabulary with which children were mostly familiar from English instructions in the pre-primary schools or their course-books. In this context, among the three vocabulary levels as *Tier one*, *Tier two*, and *Tier three* in terms of the words' commonality (more to less frequently occurring) and applicability (broader to narrower; Beck et al., 2002), basic English vocabulary items addressing body parts, basic colors and numbers, common fruits and vegetables, common domesticated animals and prevalent wild animals, and food items from Tier 1 were included. The National Framework English Curriculum, which was published in 2016 with the intention of supporting the already growing initiatives in private pre-primary schools in Turkey, was also taken into consideration, with a view to determining what VYLs should know about the target language. For the purpose of the EPVT, content validity is ensured by creating items that are age-appropriate, familiar, and adequately represent the vocabulary used by children (Cohen, 2010).

Nikolov (2016) and Rixon (2016) emphasized the importance of early language vocabulary assessment in terms of how children progress and what levels they achieve in their vocabulary learning. When EPVT is used at the beginning and at the end of different early FL programs, it can also be a predictor of how children progress in their L2 vocabulary learning, how and what extent different types of curricula and vocabulary instruction contribute to early L2 vocabulary learning, how children achieve specific curricular expectations. In other words, the results of EPVT can depict a picture about the designs of pre-primary English programs, vocabulary instruction besides assessing the children's

L2 vocabulary progress. Keeping all these concerns in mind, EPVT has two components, which are Receptive and Expressive. This measure is unique in that the target words ($N = 48$) were included into two different parts of the test separately. Related this, Shin and Crandall (2014) emphasized that the most appropriate testing method to assess proficiency in vocabulary is within the context of assessing oral language skills for young children, where they reflect their ability to comprehend and express English vocabulary with the help of the pictures in meaningful way.

In doing this, children's *breadth of vocabulary knowledge* indicating how much children know and *depth of vocabulary knowledge* how well they know it should be taken into consideration to gain insight into what level VYLs achieved basic vocabulary in English receptively and expressively (Szpotowicz & Campfield, 2016; Vinco, 2013). The fact that children can receive either the receptive or the expressive format of a given word with certain intervals in a test allows for comparisons of children's performance between test formats and learning outcomes of different curricular subjects. This comparison also allows for the researchers or educators to control the children's *effect of chance success* which means that children choose the correct answers randomly and accidentally in the receptive part. Based on this, it can be concluded that although expressive and receptive vocabulary are distinct constructs, they are supplementary to each other in this study.

One of the findings of Güngör's (2020) study is that the comparisons of the results of expressive vocabulary tests aiming at assessing children's ability to express a concept by means of an L2 word and receptive vocabulary tests aimed at measuring children's knowledge of the meaning of the same word can also be the predictor of vocabulary instruction. In this study (Güngör, 2020), the children who learned the target vocabulary didactically in a decontextualized manner by repeating chorally and individually many times to memorize it can achieve in Receptive part at a degree but have more difficulty in picture-naming test which is a measure of expressive vocabulary. On the other hand, children who acquire the target language with more constructive, communicative, and learner-centered vocabulary instruction (through songs, stories, thinking skill activities, art & craft activities, games, role-plays) in contextualized way outperform both in the receptive part and the expressive part. Thus, EPVT including 48 items both in receptive and expressive component was designed to give the best representation of VYLs' L2 word knowledge by comparing the children's understanding and learning of vocabulary receptively and expressively.

An easy-to-use EPVT Receptive Language Test (Appendix A), looking like a desk calendar, provided four images for each question. It has been developed to measure children's listening and understanding of single-word vocabulary on pre-determined subjects. Each image plate contains four colorful pictures, one of which best represents the meaning of the corresponding target word. For each target word, three distorters of the same category are identified, and the test items are created as one target and three distorters for each item. The EPVT Expressive Language Test (Appendix B) consists of the same 48 words, providing one target image on each page. Pictures for each item were presented in the test. For this age group, illustrations were also considered good promoters for motivation to complete the task.

Stage 3

After the target words were drawn by a professional artist, the piloting processes to ensure reliability, validity, and test quality were conducted in two stages. In the first stage, initial piloting was conducted via pre-pilot meetings with 20 children (8 males and 12 females, aged 5–6 years) from 3 private pre-primary schools (high-intensity English program) in three distinct İstanbul regions (Kadıköy, Beşiktaş, and Ümraniye). In the second stage, children's performance and perceptions on EPVT were triangulated by interviewing with four experts, consisting of two specialists in the field of ECE and two specialists in ELT. While the ones in ECE checked the test paper in terms of their formats, order, suitability for children's developmental characteristics, level of difficulty, appropriateness of illustrations, convenience between the curricular objectives, the others in ELT assessed in terms of the appropriateness of the length of the test, the avoidance of incorrect English and the quality and appropriateness of the other test materials like EPVT

Implementation Guide and Record Forms. Thus, the results are analyzed in depth regarding the triangulated data. Use of merged results produces better understanding and mutually confirms findings and ultimately provides validation for EPVT.

Initial piloting was to explore how children (1) understood instructions: to ensure they had been formulated in an age-appropriate and comprehensible way (2) responded to test items: in order to estimate their level of difficulty (3) had an idea about the pictures: in order to check if the style and esthetics appealed to young learners' tastes and background (4) commented on the difficulty and user-friendliness of the whole test and individual items. Szpotowicz and Campfield (2016) defined the pre-pilot meetings as "cognitive laboratories which were fundamental to the process of test construction." They expressed their advantages by saying that "they provided information on children's understanding, perception of the language and the visual materials or types of tests." Similarly, Nikolov and Szabó (2012) indicated that these meetings provide insight to the researcher or test developers on the test difficulty, familiarity, and attractiveness. School and parental consent for the interviews were previously obtained. After the researcher encouraged the children to attempt the EPVT (Receptive) and EPVT (Expressive) separately, interviews were carried out by the researcher and took place with groups of four children in quiet classrooms. During the interviews, some questions like "Was the test easy or difficult?" "Was the test interesting or boring?" "Did you like the pictures and its layout and design of the page?," "Were the instructions clear?," "Would you change anything in the test?" were asked to the children.

In content validation process, the experts' judgments were used as the criterion on which the content-related evidence of validity. In brief, the researcher's aims in the pre-pilot meetings with children and experts were to explore (1) age-appropriateness of tests; (2) developmentally appropriateness; (3) children's performance on the test could be measured; (4) measurability of the tests in terms of receptive and expressive vocabulary knowledge; (5) attractiveness of tests; (6) picture quality of the test.

Some of the feedback given and the improvements made as a result of these initial pilot meetings were explained as follows: The colors and esthetics of some illustrations caused some changes in the test. Some cognate words which were easy words to remember such as "tomato" and "potato" in the fruit & vegetable themes were reworded with "pear" and "grape." Related to this issue, it is known that form and meaning overlaps help children derive the meaning of the word even when children have limited exposure to the L2 (Potapova et al., 2016). As a result of the experts' comments on the ambiguity of picture-words relationship, the vocabulary about feelings was drawn again by paying more attention to using the illustrations that were same in size, color, and shape. It was decided to use emojis to provide the commonality in illustrations. In addition to this, children's practical advice about pictures like short skirt advice instead of a long one, smaller nose instead of the big one, long trousers instead of short one (causing ambiguity with shorts), one cherry instead of two cherries, various colors for each item in "clothes" instead of using the same color for all provide test to be age-appropriate and high reliability free from complexity. To provide clarity, the instruction was given in their mother tongue. Both the order of the themes and the items in the themes were ordered randomly to avoid item interdependence.

In accordance with the corrections and improvements as a result of the initial piloting, the related pictures in the test were drawn by the artist several times. The Receptive and Expressive vocabulary test framework including the aim, themes, and items is shown in [Figure 1](#).

In the last stage, the other test materials like "EPVT Implementation Guide" and "Record Forms" apart from "Receptive and Expressive Vocabulary Test Books" designed as table calendars were also developed to make the test easy to administer and score. The sequence of events followed leading to the final version of the test is demonstrated in [Table 1](#).



Figure 1. Receptive and expressive vocabulary test framework.

Table 1. English picture vocabulary test development sequence.

Stages of test development and administration	Additional tasks
1 Test conceptualization (Specification of Test Purpose)	
2 English National Curriculum in Turkey and course books for VYLs analysis (target vocabulary and structure)	Consultation with external experts
3 Selection of types of test and test items	Consultation with external experts
4 Test plan and specification	
5 Drawings of illustrations	
6 Evaluation of test items (drawings)	Consultation with 20 children and 4 experts
7 Correction and modification of test items	
8 Preparing recording and scoring keys	
9 Final pilot-test administration	Sample audio recordings
10 Recording pilot-test data	
11 Analysis (KR20 and Point biserial correlation)	Consultation with external experts
12 Proofreading of the test	

Data Collection Process

EPVT was carried out in quiet classrooms to each child by two researchers to eliminate the experimenter's bias and discrepancy in application. The researchers spent enough time with the whole class by attending their daily school activities and with each child individually before the test by having a conversation to establish rapport and children's sense of security. EPVT was administered individually in each pre-primary school in two sessions. Each session took approximately 10–15 min to complete. The administration process lasted for 2 weeks as there were two test formats including receptive and expressive EPVT. One of the possible reasons is stated by Papp and Walczak (2016) in their study that if there are various test formats that will be conducted consecutively, the results of the tests can be negatively affected due to some factors, such as inattention, fatigue, and boredom. The other reason is that the answers in Receptive part can affect children's answer in Expressive part. For this reason, the test order threat was controlled by implementing the tests in two different weeks.

The assessment process was started by greeting and accompanying the child to the testing room from their classroom. Then, the researchers introduced the task verbally in their first language to each child at the beginning. Every subject began with Receptive EPVT in which the child was expected to point to the correct picture among four pictures. The important thing is for them to show the correct picture not repeating the word. In the expressive part, the child expressed the name of the picture

loudly. After completion of the test, the child was escorted out of the testing room and given a 5- to 10-min rest period while a report of test results was recorded on the form. The report included subject and test information (date of test, date of birth and age) and the number of correct and incorrect responses. After completion of the test, each was given a sticker and then dismissed. This process was administered in the same way for all children for 2 weeks.

In administering the test to the children, the researchers tried to be consistent with the relevant ethical guidelines on research with young language learners' assessment. In this sense, the researchers were very sensitive to reinforcement and motivational processes. For instance, the researchers supported the children by saying "Great!," "Okey," "Good!," "Perfect!," "Go on!" after their answers. With respect to this, Nikolov (2016) asserted that positive feedback or incentives activated the acquired skills to actual performance.

The researchers recorded the answer on the performance record paper as 1 if it was correct and 0 if it was wrong. It was worth highlighting that the assessors' responsibilities – establishing rapport, administering the items according to instructions, keeping the materials ready, responding appropriately to the child, precisely recording the child's responses, keeping the child engaged, and scoring the child's responses were high. Therefore, another assessor was included to record the children's answer on the forms to make the test application more practical and time saving. The overall scores of the EPVT were entered into the Statistical Package for the Social Sciences (SPSS) for analysis and assessed separately by two independent-raters. A kappa measure of the two rater's assessment was greater than 0.85, indicating acceptable inter-rater reliability (Landis & Koch, 1977). On the other hand, the researchers are aware of the limitation of the use of video/audio-based observation research method as a data collection tool because of confidentiality and privacy issues.

Data Analysis

As for item analysis, it refers to the specific methods used to evaluate items on a test both qualitatively and quantitatively (Krishnan, 2013). The aforementioned corrections and improvements as a result of pre-pilot meetings with children and four experts were the example for the qualitative review for item development. As for the quantitative analysis (statistical analysis), the reliability and validity analysis, item discrimination indices and point biserial correlation coefficients were examined for each test including 48 multiple-binary choice items. The parameters obtained included a difficulty index, discrimination index, point biserial correlation, and reliability and validity indexes.

Results

The reliability of the EPVT (Receptive) and EPVT (Expressive) was measured with the help of the Kuder-Richardson Formula 20 (KR20), and the reliability coefficients were found to be 0.89 and 0.91, respectively. KR-20 values of 0.8 or higher are considered good reliability (Salkind, 2010). Besides, EPVT were administered to the same group of pre-primary school children by selecting 30 children twice after 3 weeks later. Once completed, the pairs of scores for each child were lined up in two columns, and correlation coefficient was calculated between the two sets of scores by using the test-retest method. The test-retest reliability measures which were 0.93 for EPVT (Receptive) and 0.94 for EPVT (Expressive) show that the scales can be said to have acceptable internal consistency. They are shown in Table 2.

Table 2. Pilot reliability indices: test versions (KR20).

	Number the of items	KR20	Test-retest
Receptive EPVT	48	.89	.93
Expressive EPVT	48	.91	.94

Table 3. Distribution of the items in expressive and receptive EPVT among the different ranges of difficulty indices.

Difficulty index	Degree of difficulty	Number of questions in Expressive	Number of questions in Receptive
0–0.2	Very difficult	0	0
0.21–0.4	Moderately difficult	1	0
0.41–0.6	Intermediate difficulty	25	10
0.61–0.8	Moderately easy	22	38
0.81–1.0	Very easy	0	0
Total		48	48

In addition to this, point-biserial correlation was conducted to measure how much predictive power an item had and how the item contributed to predictions by estimating the correlation between each test item and the total test score. Point-biserial correlations employed for item analysis were used to determine difficulty levels of the items by measuring the proportion of children who answer the question correctly (McCowan & McCowan, 1999). Related to this, Schwarz (2011) states that high p -values mean the item is easy and low p -values mean the item is difficult. Based on this, the item discrimination index (p -values) ranging from 0.223 to 0.626 for EPVT (Expressive) and from 0.24 to 0.65 for EPVT (Receptive) was considered “acceptable.” In the literature, it is indicated that items with point biserial correlation above 0.20 are accepted (Sotaridona et al., 2013). The results of point biserial correlation analysis are shown in Tables 3 and 4.

Table 3 shows the distribution of the questions among the ranges of difficulty index values for the 48 items included in both receptive and expressive EPVT. Items were classified as very difficult ($\rho \leq 0.20$), moderately difficult ($\rho > 0.20$ and ≤ 0.40), intermediately difficulty ($\rho > 0.40$ and ≤ 0.60), moderately easy ($\rho > 0.60$ and ≤ 0.80), or very easy ($\rho > 0.80$). The highest number of questions (25) falls in the intermediate difficulty index ranges 0.41–0.6, while 22 items fall in the moderately easy index ranges 0.61–0.8 and 1 question is in the moderately difficult category in the Expressive EPVT. On the other hand, the highest number of questions (38) falls in the moderately easy index ranges 0.61–0.8, while 10 items fall in the range 0.41–0.6. The items falling in the very difficult index ranges (0–0.2) and very easy index ranges (0.81–1.0) are not acceptable. However, the degree of difficulty of the items in Receptive and Expressive EPVT was found to be between 0.2 and 0.8 which is acceptable. Therefore, the scales can be said to have ideal difficulty in terms of discrimination potential. In addition to this, discrimination indices of the items in Expressive and Receptive EPVT are given in Table 4.

Table 4 shows the distribution of the questions among the ranges of difficulty index values for the 40 items included in both receptive and expressive EPVT. Test discrimination values were classified as poor items ($\rho \leq 0.19$), fairly good items ($\rho > 0.20$ and ≤ 0.29), good items ($\rho > 0.30$ and ≤ 0.39), and very good items ($\rho > 0.40$; Brown, 1996; Carroll & Sapon, 1958; Guilford & Fruchter, 1973). In general, it was seen in the discrimination index that 62.5% of total questions in both tests were classified as with good or excellent discrimination index.

Table 5 shows the examples of item difficulty and point-biserial correlation values for receptive and expressive test items separately. EPVT (Receptive) and EPVT (Expressive) developed for measuring very young learners’ receptive and expressive vocabulary in English consist of 48 items in each test, and they

Table 4. The distribution of the items in expressive and receptive EPVT among the different ranges of discrimination indices.

Discrimination index	Percentage of questions (Receptive)	Percentage of questions (Expressive)
≤ 0.09	0	0
0.1–0.19	0	0
0.2–0.29	6.25%	6.25%
0.3–0.39	31.25%	31.25%
> 0.39	62.5%	62.5%
Total (%)	100	100

Table 5. EPVT difficulty and discrimination indices for each item in the pilot data.

Item	Exp. Mean	Rec. Mean	Rec. Std	Exp. Std	Exp. Rpbi	Rec. Rpbi	Item difficulty (Exp.)	Item difficulty (Rec.)
yellow	36.32	34.48	0.44	0.49	0.429	0.609	0.62	0.73
green	37.08	35.45	0.42	0.49	0.474	0.543	0.58	0.78
blue	35.94	34.70	0.42	0.50	0.312	0.506	0.52	0.76
black	35.93	29.27	0.44	0.49	0.239	0.326	0.39	0.73
red	36.64	35.33	0.43	0.47	0.517	0.588	0.67	0.76
purple	36.31	34.86	0.43	0.50	0.312	0.574	0.47	0.75
white	37.00	34.86	0.43	0.49	0.478	0.532	0.59	0.76
pink	36.21	34.57	0.42	0.50	0.315	0.512	0.49	0.78
grapes	36.03	35.15	0.50	0.43	0.526	0.318	0.75	0.49
orange	35.14	34.00	0.47	0.41	0.455	0.323	0.74	0.68
pear	31.21	35.68	0.50	0.45	0.469	0.353	0.73	0.55
cherry	30.73	35.19	0.48	0.41	0.490	0.340	0.75	0.63
corn	30.55	35.12	0.47	0.41	0.447	0.448	0.76	0.66
carrot	31.89	35.54	0.49	0.47	0.513	0.382	0.68	0.62
banana	31.76	34.71	0.43	0.46	0.522	0.368	0.70	0.75
apple	31.81	34.41	0.43	0.46	0.537	0.432	0.70	0.75
eye	32.31	36.95	0.48	0.45	0.627	0.633	0.71	0.63
foot	33.39	37.48	0.48	0.49	0.605	0.670	0.59	0.63
head	32.73	36.12	0.49	0.49	0.552	0.478	0.61	0.62
ear	30.11	36.75	0.48	0.41	0.392	0.584	0.75	0.65
nose	33.71	36.33	0.46	0.50	0.505	0.583	0.47	0.69
hand	32.50	36.53	0.47	0.50	0.457	0.577	0.55	0.67
hair	31.65	36.72	0.50	0.45	0.511	0.453	0.71	0.53
mouth	33.01	36.96	0.48	0.50	0.514	0.589	0.55	0.63
happy	32.08	36.16	0.46	0.50	0.386	0.549	0.54	0.69
scared	30.42	36.17	0.49	0.48	0.300	0.306	0.65	0.41
surprised	30.96	36.46	0.50	0.46	0.409	0.388	0.70	0.48
tired	30.76	36.66	0.50	0.50	0.234	0.484	0.47	0.57
hungry	31.10	35.72	0.46	0.50	0.260	0.487	0.46	0.70
sad	30.97	35.76	0.48	0.50	0.223	0.427	0.50	0.63
angry	31.88	37.31	0.48	0.49	0.450	0.655	0.61	0.64
thirsty	31.02	35.85	0.49	0.50	0.255	0.413	0.46	0.60
shoe	31.96	35.18	0.45	0.49	0.459	0.463	0.61	0.72
hat	31.72	36.06	0.47	0.47	0.496	0.525	0.68	0.68
skirt	31.93	35.30	0.50	0.50	0.335	0.303	0.46	0.54
coat	32.59	35.89	0.50	0.50	0.397	0.387	0.47	0.57
trousers	32.57	35.59	0.50	0.50	0.376	0.351	0.44	0.57
sock	31.71	35.18	0.48	0.50	0.306	0.345	0.45	0.63
dress	32.55	34.88	0.49	0.49	0.363	0.273	0.43	0.59
scarf	25.85	35.48	0.48	0.47	0.234	0.406	0.67	0.63
monkey	29.25	35.40	0.44	0.48	0.439	0.475	0.63	0.73
chicken	32.14	34.51	0.44	0.50	0.355	0.359	0.46	0.73
horse	29.38	35.25	0.44	0.49	0.437	0.506	0.62	0.74
snake	32.96	33.08	0.44	0.49	0.556	0.261	0.59	0.73
lion	33.31	36.02	0.47	0.49	0.581	0.490	0.58	0.66
dog	33.35	35.26	0.44	0.50	0.558	0.332	0.56	0.74
cat	32.99	33.66	0.42	0.50	0.409	0.240	0.44	0.76
elephant	33.38	35.32	0.44	0.50	0.548	0.482	0.55	0.76

are answered with true/false and said/couldn't say (1/0 scored), respectively. The correlation coefficient being calculated here is between a naturally occurring dichotomous nominal scale (the correct or incorrect answer on each item usually coded as 1 or 0) with an interval scale test. Findings support that the English Picture Vocabulary Test is child-friendly and practicable way of assessing very young EFL children's receptive and expressive foreign language vocabulary related to specified curricular subjects.

Discussions and Conclusions

Contrary to some mostly used tools providing more comprehensive evaluation of the breadth of receptive and expressive vocabulary knowledge, English Picture Vocabulary Test aims at assessing some basic theme-related vocabulary receptively and expressively. Within this scope, the newly

developed assessment tool – EPVT – including familiar common target vocabulary items which are appropriate to basic levels might be useful and relevant to measure young children’s receptive and expressive vocabulary knowledge in English.

In the field of early *English language assessment* that refers to any method used for measuring the learning and performance of young children to observe and obtain information about their vocabulary knowledge, linguistic, and communicative skills (Espinoza & Lopez, 2007), *picture vocabulary tests* play a more important role for young children than for adults. In the early stages of foreign language learning, the assessment of VYL’s vocabulary knowledge in terms of both its comprehension and expression with the help of pictures can be used as only a starting point. It should be kept in mind that L2 learning at pre-primary level is the *first meeting* with a foreign language for children who start to learn this language in EFL context. In this sense, language teachers, policymakers, teacher educators, and researchers who place emphasis on children’s first impressions in their first meetings should pay more attention to include meaningful and useful assessment tools and methods; therefore, the positive effects of this “impressive meeting” sustain for a long time.

The important finding resulting from this study is the fact that the EPVT which is developed by taking into consideration VYLs’ age, characteristics, context of instruction, amount and type of exposure to English and purpose of assessment is age-appropriate and theme-related assessment tool. At a time of growing interest in pre-primary L2 learning and assessment, EPVT can be incorporated into the ELT national programs for VYLs. Besides, it can be used by early English language teachers in the assessment of young children’s L2 vocabulary knowledge at the end of their pre-primary education.

As a final note, considering the process of development and implementation of EPVT, it can be claimed to be a well-designed early foreign language assessment tool integrating ECE and ELT pedagogy. Following the careful piloting process which include the reliability and validity issues in test development and the pre-piloting procedures which consist of target-age children’s perspective in terms of linguistic, visual, and pragmatic content of test item as well as experts’ opinion, the final product could be accepted as a functional and practical assessment tool in low exposure EFL pre-primary schools. The results of EPVT can be used for giving diagnostic feedback on strengths and weaknesses at vocabulary level, providing information on children’s L2 vocabulary development, monitoring progress, planning future action, and evaluating the effectiveness of vocabulary instruction in programs by individual learners, teachers, classes, schools, regions, or nations.

Implications

The study bears implications for using EPVT to assess very young EFL learners’ receptive and productive language learning at pre-primary levels. It is not limited to measuring L2 learning results but also serves to provide feedback for teaching and supporting for learning. Pre-primary school teachers of English who are still struggling to find their way by mostly drawing on a number of ELT sources such as ministry curricula and textbooks need to benefit from age-appropriate assessment tools to assess children’s development and support their L2 learning potential. In this regard, the study provides a framework to guide researchers and teachers in developing and implementing age appropriate L2 assessments.

The themes and target vocabulary in EPVT, albeit developmentally, linguistically, and culturally relevant in the Turkish context, may be less familiar for young EFL learners in different parts of the world. On the other hand, EPVT may not have one-to-one match with the achievement targets which are included in different language policy documents. Therefore, the use of EPVT is possible in different educational contexts, depending on how early L2 teaching and learning are understood in different contexts, what achievement targets are specified or what the culturally responsive instructions and assessments are placed in their curricula. By taking into account these dimensions, EPVT can be used in different EFL contexts for progress monitoring in preschool’s English receptive and expressive vocabulary.

Limitations and Future Directions

This study has highlighted the critical importance of the children's perspective on how they feel toward the test, what they think is difficult or easy or what they think is confusing about the test items. Future research should also give attention to the need for target-age group consultation and careful piloting of items and test procedures.

There is a scarcity on the national and international research scene of instruments measuring very young learners' progress and attainment in English at pre-primary level. Therefore, more efforts to the integration of different types of assessments – formal and informal, standards-based and performance-based, or standardized and alternative assessment – would be needed to ascertain whether VYLs obtain linguistic and communicative achievement targets in L2. In doing this, considering the criteria and procedures of early childhood assessment and ethical considerations is of great importance.

Furthermore, this study has included young children with similar language and socioeconomic backgrounds. Therefore, the use of EPVT may not generalize to heterogeneous contexts where there are culturally and linguistically diverse children who have a wide variety of language and literacy abilities. Based on this, teachers and researchers may try to design assessment tools which would be more culturally responsive and more aligned with their learners' L2 needs.

Acknowledgments

The authors would like to thank the 16 private pre-primary schools that give us the opportunity to meet with their children. We are also grateful to young children who provide their comments and opinions about the test difficulty, familiarity, and attractiveness in the process of test construction.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Burcu Güngör  <http://orcid.org/0000-0002-3102-8183>

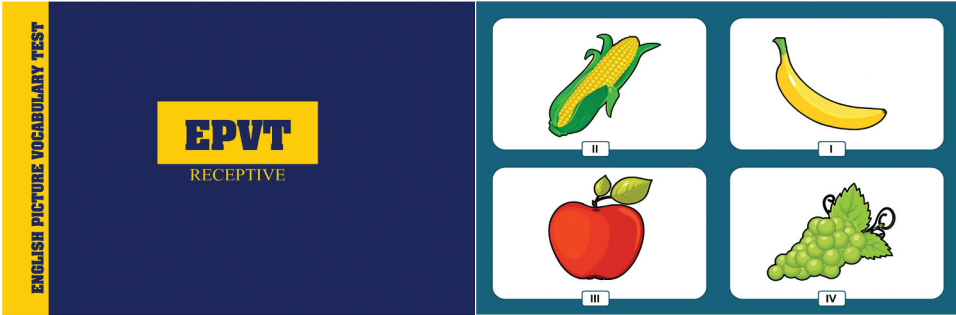
References

- Bacsa, É., & Csikos, C. (2016). The role of individual differences in the development of listening comprehension in the early stages of language learning. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 263–290). Springer.
- Beck, I., McKeawn, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.
- Biricik, E., & Özkan, Y. (2012). The role of teacher attitude in preschool language education. *Çukurova University Faculty of Education Journal*, 41(1), 70–86. <https://dergipark.org.tr/en/download/article-file/46477>
- Bourke, J. M. (2006). Designing a topic-based syllabus for young learners. *ELT Journal*, 60(3), 279–286. <https://doi.org/10.1093/elt/ccl008>
- Brewster, J., Ellis, G., & Girard, D. (2002). *The primary English teacher's guide*. Pearson Education Limited.
- Brown, J. D. (1996). *Testing in language programs*. Prentice Hall.
- Brownell, R. (2000). *Expressive and receptive one-word picture vocabulary tests*. Academic Therapy.
- Cameron, L. (2001). *Teaching languages to young learners*. Cambridge University Press.
- Carroll, J. B., & Sapon, S. M. (1958). *Modern language aptitude test*. The Psychological Corporation.
- Černá, M. (2015). Pre-primary English language learning and teacher education in the Czech Republic. In S. Mourão & M. Lourenço (Eds.), *Early years second language education: International perspectives on theories and practice* (pp. 165–176). Routledge.
- Cohen, R. J. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). McGraw-Hill Higher Education.
- Connery, C., John-Steiner, V., & Marjanovich-Shane, A. (2010). *Vygotsky and creativity: A cultural-historical approach to play, meaning making, and the arts*. Peter Lang.
- Conteh, J. (2012). *Teaching bilingual and EAL learners in primary schools*. Sage and Learning Matters.

- Council of Europe. (2018). *Collated representative samples of descriptors of language competences developed for young learners aged 7–10 years*.
- Doğançay-Aktuna, S., and Kızıltepe, Z. (2005). English in Turkey. *World Englishes*, 24(2), 253–265.
- Dunn, L. M., & Dunn, D. M. (2007). *The Peabody picture vocabulary test* (4th ed.). NCS Pearson, Inc.
- Edelenbos, P., Johnstone, R., & Kubanek, A. (2006). *The main pedagogical principles underlying the teaching of languages to very young learners: Languages for the children of Europe*. Final Report of the EAC 89/04, Lot 1 study. Brussels: European Commission.
- Ellis, G. (2014). Young learners: Clarifying our terms. *ELT Journal*, 68(1), 75–78. <https://doi.org/10.1093/elt/cct062>
- Espinoza, L. M., & Lopez, M. L. (2007). Assessment considerations for young English language learners across different levels of accountability. *Paper prepared for The National Early Childhood Accountability Task Force and First 5 LA*. European Commission. (2011). *Language learning at pre-primary school level: Making it efficient and sustainable: A policy handbook*. http://ec.europa.eu/languages/pdf/ellpwp_en.pdf
- Fleurquin, F. (2003). *Development of a standardized test for young EFL learners: Spaan fellow working papers in second or foreign language assessment* (Vol. 1). English Language Institute, University of Michigan.
- Garton, S., & Copland, F. (Ed.). (2019). *The Routledge handbook of teaching English to young learner*. Routledge.
- Gibson, T., Oller, D. K., Jarmulowicz, L., & Ethington, C. A. (2012). The receptive-expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism: Language and Cognition*, 15(1), 102–116. <https://doi.org/10.1017/S1366728910000490>
- 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. doi:10.1080/13670050.2018.1494131
- Guilford, J. P., & Fruchter, S. B. (1973). *Fundamental statistics in psychology and education* (5th ed.). McGraw-Hill.
- Güngör, B. (2020). *The effects of early childhood English language education program on very young learners' vocabulary knowledge and communicative skills* [PhD thesis]. Marmara University.
- Haenni Hoti, A., Heinzmann, S., & Müller, M. (2009). “I can you help?”: Assessing speaking skills and interaction strategies of young learners. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 119–140). Mouton de Gruyter .
- Harmer, J. (2015). *The practice of English language teaching* (5th ed.). Longman.
- Hasselgren, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), 337–354. <https://doi.org/10.1191/0265532205lt312oa>
- Hestetræet, T. I. (2019). Vocabulary teaching for young learners. In S. Garton & F. Copland (Eds.), *The Routledge handbook of teaching English to young learner* (pp. 220–233). Routledge.
- Inbar-Lourie, O., & Shohamy, E. (2009). Assessing young language learners: What is the construct? In M. Nikolov (Ed.), *Contextualizing the age factor: Issues in early foreign language learning* (pp. 83–96). Mouton de Gruyter.
- Johnstone, J. (2009). An early start: What are the key conditions for generalized success? In J. Enever, and J. Moon (Eds.), *Young learner English language policy and implementation: International perspectives*. Reading: Garnet Education and IATEFL.
- Kail, R. (2010). *Children and their development* (5th ed.). Pearson Prentice Hall.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kokla, A. (2013). *Dora the explorer: A TV character or a preschooler's foreign language teacher* [Master's thesis]. Aristotle University of Thessaloniki.
- Krishnan, V. (2013). *The early child development instrument (EDI): An item analysis using classical test theory (CTT) on Alberta's data*. Early Child Development Mapping Project.
- Kubanek-German, A. (2000). Early language programmes in Germany. In M. Nikolov, and H. Curtain (Eds.), *An early start: Young learners and modern languages in Europe and beyond* (pp. 59–70). Strasbourg: Council of Europe Publishing.
- Landis, J. R., & Koch, G. D. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Pearson/Merrill Prentice Hall.
- McCowan, R. J., & McCowan, S. C. (1999). *Item analysis for criterion-referenced tests*. CDHS, SUNY.
- McKay, P. (2006). *Assessing young language learners*. Cambridge University Press.
- Mihaljević Djigunović, J. (2016). Individual differences and young learners' performance on L2 speaking tests. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 243–263). Springer.
- Mistry, M., & Sood, K. (2015). *English as an additional language in the early years: Linking theory to practice*. Routledge.
- Mondria, J. A., & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 79–100). John Benjamins Publishers.
- Nation, P. (1990). *Teaching and learning vocabulary*. Heinle & Heinle Publishers.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

- National Institute of Child Health and Human Development (NICHD). (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). U.S. Government Printing Office
- Nikolov, M. (Ed.). (2009a). *The age factor and early language learning*. Mouton de Gruyter.
- Nikolov, M. (2016). *Assessing young learners of English: Global and local perspectives*. Springer International Publishing.
- Nikolov, M., & Szabó, G. (2012). Developing diagnostic tests for young learners of EFL in grades 1 to 6. In E. D. Galaczi & C. J. Weir (Eds.), *Voices in language assessment: Exploring the impact of language frameworks on learning, teaching and assessment – Policies, procedures and challenges* (pp. 347–363). UCLES/Cambridge University Press.
- Papp, S. (2019). Assessment of young English language learners. In S. Garton, and F. Copland (Eds.), *The Routledge handbook of teaching English to young learner* (pp. 389–409). Routledge.
- Papp, S., & Walczak, A. (2016). The development and validation of a computer-based test of English for children: The computer-based Cambridge English: Young learners tests. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 139–190). Springer.
- Pinter, A. (2012). Is research relevant for teachers of English working with young learners? In M. Allström, & A. Pinter (Eds.), *English for Young learners – Forum 2012* (pp. 11–27). Upsala Universitet.
- Pizorn, K. (2009). Designing proficiency levels for English for primary and secondary school students and the impact of the CEFR. In N. Figueras & J. Noijons (Eds.), *Linking to the CEFR levels: Research perspectives* (pp. 87–100). Cito/ EALTA.
- Potapova, I., Blumenfeld, H. K., & Pruitt-Lord, S. (2016). Cognate identification methods: Impacts on the cognate advantage in adult and child Spanish-English bilinguals. *International Journal of Bilingualism*, 20(6), 714–731. <https://doi.org/10.1177/1367006915586586>
- Prosi-Santovac, D., & Rixon, S. (2019). *Integrating assessment into early language learning and teaching*. Channel View Publications.
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge university press.
- Rixon, S. (2013). *British Council survey of policy and practice in primary English language teaching worldwide*. British Council.
- Rixon, S. (2016). Do developments in assessment represent the “coming of age” of young learners English language teaching initiatives? The international picture. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp.19–41). New York: Springer.
- Robert, C., Borella, E., Fagot, D., Lecerf, T., & De Ribaupierre, A. (2009). Working memory and inhibitory control across the life span: Intrusion errors in the reading span test. *Memory & Cognition*, 37(3), 336–345. <https://doi.org/10.3758/MC.37.3.336>
- Rupp, A., Vock, M., Harsch, C., & Köller, O. (2008). *Developing standards-based assessment tasks for English as a foreign Language: Context, processes, and outcomes in Germany*. Waxmann.
- Salkind, J. N. (2010). *Encyclopedia of research design*. SAGE Publications.
- Schwarz, J. (2011). *Research methodology: Tools. Applied data analysis (with SPSS). Lecture 02: Measurement scales and item analysis*. Lucerne University of Applied Sciences and Arts
- Shin, J. K., & Crandall, J. A. (2014). *Teaching young learners English: From theory to practice*. Boston: National Geographic Learning/Cengage Learning.
- Slattery, M., & Willis, J. (2001). *English for primary teachers*. Oxford University Press.
- Sotaridona, L. S., Wibowo, A., Hendrawan, I., & Pornel, J. (2013, October 1–2). An application of nominal response model to identify erroneously-scored test items. *Invited paper at the 12th National Convention on Statistics*.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139–152. <https://doi.org/10.1080/09571730802389975>
- Szpotowicz, M., & Campfield, D. E. (2016). Developing and piloting proficiency tests for polish young learners. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 109–139). Springer .
- Taylor, L., & Saville, N. (2002). Developing English language tests for young learners. *Research Notes*, 7, 2–5. <https://www.cambridgeenglish.org/Images/23119-research-notes-07.pdf>
- Thornbury, S. (2016). Communicative language teaching in theory and practice. In G. Hall (Ed.), *The Routledge handbook of English language teaching* (pp. 224–237). Routledge.
- Vinco, H. M. (2013). *Assessment of preschool vocabulary: Expressive and receptive knowledge of word meanings* [Master’s thesis]. Florida State University.
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- Wesson, K. (2011). *Attention span revisited*. Retrieved June 3, 2020, from <http://sciencemaster77.blogspot.com/2011/01/attention-spans-revisited.htm>
- Zandian, S. (2012). Participatory activities, research and language classroom practice. In M. Allström, & A. Pinter (Eds.), *English for young learners – Forum 2012* (pp. 133–142). Upsala Universitet.

Appendix A: EPVT Receptive Part



Appendix B: EPVT Expressive Part

