



Children's Picture Books as a Source of L2 Literacy Curriculum: A Preliminary Corpus-based Study

Eun-Young Choi (Chung-Ang University) Younghee Cheri Lee (Ajou University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: June 06, 2022
Revised: August 14, 2022
Accepted: August 31, 2022

Eun-Young Choi (co-first author)
Graduate Student,
Dept. of English Education,
Chung-Ang Univ.
Tel: 02-820-5391
E-mail: cey1981@cau.ac.kr

Younghee Cheri Lee (co-first author/ corresponding author)
Teaching Professor,
Dasan University College,
Ajou Univ.
Tel: 031-219-3053
E-mail: cheriberry@ajou.ac.kr

ABSTRACT

Choi, Eun-Young and Younghee Cheri Lee. 2022. Children's picture books as a source of L2 literacy curriculum: A preliminary corpus-based study. *Korean Journal of English Language and Linguistics* 22, 797-820.

It has been a widely accepted claim that reading a children's picture book lays a foundation for young learners' literacy development. Despite meaningful efforts to promote picture-book programs at public schools in the domestic education context, a feasible, tailored picture-book literacy program is still mostly confined to private education sectors, not public English classroom settings. By employing a corpus-based approach to lexical difficulty, thus, this study purports to delve into whether children's picture books in native L1 English can be utilized as a primary source of an L2 literacy curriculum accessible in Korean public class settings. To this end, using the Project Gutenberg archive, we built a corpus of children's picture books entitled the COCH, which encompasses 600 stories with 452,219 tokens. Then, we computed the token coverage rate of each picture book using two-tiered analyses. One was to employ VP-Kids levels 1 to 10 (Roessingh and Cobb 2008) as baseword level lists, and the other was to adopt the KEBWL lists (Shin 2015) compiled based on the national English curriculum of Korea. The results evidenced that sixty children's picture books reached the threshold level of token coverage when the VP-Kids lists were applied. When the KEBWL lists were based, though, it was found that only ten books could survive to fit into the literacy curriculum for elementary schoolers and twenty-nine picture books for secondary schools in the Korean teaching context. During analyses, this study was attentive to how children's picture books that exceed the minimum threshold level of token coverage could have been adjusted and how off-list words could have been revalued to take in more picture books in the shortlist of the L2 picture-book-based literacy curriculum. Implications and future research directions concerning a children's picture book as a source of L2 linguistic and cultural literacy will be discussed.

KEYWORDS

picture book, children's book, young learner, L2, reading, token coverage, vocabulary profile, literacy curriculum, L2 literacy, cultural literacy, BNC/COCA, VP-Kids, corpus

1. Introduction

According to the achievement level of the 2015 revised national curriculum of English proposed by the Korea Institute for Curriculum and Evaluation (KICE) in 2016, Korean elementary schoolers in grades 3 and 4 tend to lack an understanding of basic vocabulary words presented in national English textbooks. The same pattern was also observable in the upper grades. Research has shown that fifth and sixth graders have difficulty grasping the gist of content in English textbooks. The Ministry of Education in Korea has made a wide range of efforts to enhance the quality of the English curriculum in public education settings, such as covering reading and writing skills in national English textbooks earlier than its initial plan and revising the national curriculum as necessary. The reading achievement standards presented in the 2015 revised national curriculum state that the third and fourth graders are expected to read and understand simple words, while the fifth and sixth ones are required to read short texts and understand the main content of them (Ministry of Education 2015). Despite the effortful endeavors made to date, however, Korean English classrooms face the reality that children still lag behind such achievement standards.

On that account, to reach such academic achievement standards guided by the government and to guarantee the effectiveness of English teaching and learning experiences, a novel approach that has never been attempted before seems to be of great importance. Such approaches may include not only the improvement of the L2 learning environment but also that of learning materials. In Korea, the national English curriculum for elementary schoolers has gone through five times of curriculum revisions since 2007. One of the most significant changes regards the increase in the written language portion: learning to read and write. Likewise, controversy over covering a written language in the primary school curriculum has been followed by the debate over its optimum timing of introduction. Consequently, Korea's revised national English curriculum finally shifted its adoption one semester earlier by covering reading and writing from the first semester of the third grade in 2009; the written language form was originally introduced in the second semester of grade 3 (Kim 2016, 2019, Yoon, Y. 2019). Moreover, in a teacher survey about improving English education in the school context, many participants responded that additional book reading programs pertinent to each grade level are of great necessity to expand a wide variety of English teaching and learning experiences (KICE 2018).¹

In school settings, where national English textbooks are utilized as the primary teaching and learning materials, reading and understanding the text significantly influences learners' overall learning achievement. In particular, L2 learners' literacy skills significantly count in the English as a Foreign Language (EFL) environment, where excellent exposure to the target language is limited in both quantity and quality. A feasible alternative to improving literacy in the EFL context is no other than to offer meticulously selected language input by considering learners' proficiency levels and encompassing rich reading materials. In the same vein, reading children's literature with various story plots, accessible vocabulary words, and simple sentence structures has long been welcomed in a classroom setting with insufficient language input (Chae and Lee 2010, Jang and Kim 2021, Park and Park 2007, Rha 2011, Yoon and Maeng 2011, Yoon, T. 2019).

The benefits of reading children's literature also exist in vocabulary learning. Learning to read using children's books can give learners a positive motivation for learning a foreign language. By encountering various stories in children's books, decoding and encoding meanings and inferencing skills can be achieved. Through repetitive reading and listening activities, various sentence patterns can also be recycled (Brewster, Ellis and Girard 2002,

¹ While several extra after-reading programs are accessible in public English classroom settings, it should be noted that some programs (e.g., *Reading Gate*) are not suitable for maintaining early L2 learners' reading motivation due to challenging words and burdensome length. Others (e.g., *Little Fox*) are inappropriate for linking learning after reading (See Yoon and Yang 2021).

Choi and Park 2017, Donelson and Nilsen 2005, Ellis and Breswster 1991). Learning vocabulary by guessing its meaning in the context is far more effective than any other method for learning vocabulary in their L2; more importantly, which applies when learning an L2 as in an L1 (Elley 1991, Krashen 2004, Nation 1997).

Meanwhile, Nation (1990) emphasized that one of the most critical factors to be considered when selecting reading materials is the lexical difficulty of the text. In particular, to design an optimum early literacy curriculum, reading materials should be arranged according to L2 learners' vocabulary levels or their learning goals set in the reading curriculum. Derived from corpus-based empirical studies, if carefully screened additional reading materials are provided according to the reading goals of the target language, it is no wonder that such materials will significantly help improve learners' reading fluency and performance.

Nevertheless, many prior studies have been only confined to delineating the quantitative aspects of vocabulary itself, most of which include: the number of essential vocabulary words that Korean L2 learners need (Maeng and Kim 2011), an optimum number of vocabulary words that L2 learners need (Nation 2001, Schmitt 2008, Webb and Rodgers 2009), and adequate vocabulary learning volume (Ellis and Girard 1992, Nation 1990, Scott and Ytreberg 1990). Further, corpus-informed research into token coverage geared towards young L2 learners has been insufficient, thereby failing to offer additional reading materials as so-called 'after-school makeup learning content' arranged by their lexical difficulty.

This study, therefore, attempts to conduct a preliminary analysis to design an L2 literacy curriculum for young learners of English. The ultimate goal of this study lies in carefully selecting additional reading materials tailed for elementary schoolers in the domestic teaching context, whereby analyzing the lexical profiles of children's picture books embedded with visual information. To this end, copyright-free children's picture books archived in the Project Gutenberg Digital Library were chosen as a primary data source so that every child can have an equal opportunity for extra reading materials.

Indeed, English as a mandatory school subject is signified by one of Korea's broadest educational gulfs driven by income gaps (Lee 2014). Regardless of one's financial status, as long as the Internet is accessible, anyone should be given a chance to access free e-Books in Project Gutenberg across the nation. From the section of 'Frequently Viewed or Downloaded: Top 100 eBooks,' children's picture books with high popularity and preference were downloaded for initial screening. Then, the lexical difficulty of the initially selected picture books was further analyzed using two different benchmark level lists—Cobb's (2021) VocabProfilers lists for Kids (VP-Kids) and Shin's (2015) level lists extracted from the 2015 revised English curriculum (KEBWL). The VP-Kids baseword level lists were adopted to measure children's texts using the benchmark vocabulary lists suited to young learners to produce reliable and consistent results, concurrently considering genre attributes. The KEBWL level lists were utilized additionally to measure the feasibility of children's picture books tailored to the 2015 national English curriculum.

This study purports elementary learners in Korea to experience optimum learning by using additional reading materials in varying but optimum vocabulary levels. In addition, to enhance young L2 learners' early literacy, this study aims to trigger the achievement of young learners' reading habits and decoding skills by supplementing reading materials that are lacking in the national curriculum alone. The research goals stated above lead to the following research questions.

- (1) Of the top 100 children's picture books downloadable from the Project Gutenberg archive, how many books would satisfy the threshold level (95%) of token coverage when the VP-Kids level lists were applied?
- (2) Among the picture book candidates, how many children's picture books would survive when the KEBWL lists were adopted?

2. Literature Review

2.1 Children's Picture Books as a Source of Literacy Development

A well-established claim is that children's picture books hold learning value in linguistic and cultural literacy development concurrently; indeed, children's picture books written in native L1 English have been loved across L2 reading classes around the globe. Here, children's picture books are distinctive from simply children's books in that the pictures given are as meaningful as the text (Norton 1999). It should also be noted that conventional literacy development pertains to phonological and phonemic awareness, word recognition, sentence comprehension, and reading fluency (National Reading Panel 2000). Literacy development in the aspects of picture books is no different from the conventional definition in that it also flows from word recognition accompanied by pictures to sentence comprehension supported by meaningful graphical representations.

Children's picture books are composed of a short story plot typically leading to 200-word tokens, and sentence structures are likely to be quite simple and direct. Children's picture books include common sub-genres, such as fictional novels, true stories, songs, rhymes, poems, fairy tales, and folklore. Those pictures accompanied by their texts are often called illustrations, some of which are presented using artistic techniques of photographs, drawings, and collages (Jalongo 2004). Children's picture books can also be categorized based on the portion between texts and pictures weighed: illustrated books without texts, books with texts and pictures of equal weight, and finally, texts taking the primary portion but pictures for marginal decorations (Temple, Martinex, Yokota and Naylor 1998).

Pictures and texts in picture books do not simply convey the content but intermingle the two elements effectively and explicitly so as to provide the reader with a potential imagination (Lukens 1990). The images presented in the books reflect the story, thereby helping children use the target language more creatively (Huck, Hepler, Hickman and Kiefer 2001). Furthermore, illustrations and visual representations in children's picture books embody young learners' daily experiences and stories delivering historical facts or social phenomena (Russell 2009).

It is well-established that children can better understand the lives of others by being aware of and developing their daily lives as part of their experience through the process of understanding the storyline presented with pictures in the books. L2 learners of English with low proficiency, thus, can quickly access the stories through pictures (Kim and Lee 2012). These findings demonstrate that pictures in children's stories enable young learners, either L1 or L2, to read, decode, and understand, thus providing them with meaningful experiences to become part of the characters in the story and finally rendering them soak in reading pleasure.

In Sipe and Bauer's (2001) study, L1 preschool children were presented with a fairy tale story using a picture book. While reading, it was found that children well understood the text concerning real-life context by linking the pictures with the texts. In addition, children actively intervened in the pictures and texts to manipulate the story into their own ones. A follow-up study conducted by Sipe and McGuire (2006) found that reading children's picture books enabled children to promote autonomy and independence in telling stories through pictures. In addition, a range of activities—chasing pictures in the books, drawing in a different form, or decorating a story—were found to expand into post-reading activities, so fostering children's creative thinking skills.

Research in the national context also supports the prior findings. After reading picture books for two classes, respectively, sixth graders' appreciation levels increased along with reading interest, target cultural awareness, and pleasure in creating new stories (Lee and Bae 2007). Overall, it can be inferred that incorporating picture books into the reading curriculum promotes early literacy, improves reader comprehension, and gives young learners the pleasure of reading.

Despite their varied previous findings and different reading goals, either intensive or extensive reading activities

can be performed using picture books (e.g., Birketveit and Rimmereide 2017, Choi and Park 2017, Kusanagi, Kobayashi and Fukaya 2020, Rha 2011). In the English Language Teaching (ELT) discipline, children's picture books have been provided to date in the format of graded readers by several commercial publishers. Practitioners in various EFL teaching contexts worldwide have adopted such leveled readers as primary reading resources for various reading activities. However, graded readers are far from the concept of *so-called* original, authentic picture books in that syntactic complexity of the texts is adjusted, and post-reading tasks for young learners are tailored to meet the needs of L2 learners in the EFL classroom environment.

Many researchers further asserted that there are advantages of using original picture books, distinguished from storybooks mainly intended for language learning purposes in L2 instructions (see Enever and Schmid-Schönbein 2006, Nikolajeva and Scott 2000). Authentic picture books usually comprise songs, poems, and descriptions of various activities. Texts embedded with pictures allow young learners to experience a variety of vocabulary words that would fit L2 learners' proficiency level and, more importantly, words with cultural elements.

Moreover, such rich contexts can promote learner comprehension of the text. Learning a foreign language without being exposed to cultural elements, either directly or indirectly, would be meaningless. In L2 learning contexts, providing learners with an integrated learning experience of language and culture so as to inspire L2 learners to understand the target language culture means far more than simply teaching the target language vocabulary or linguistic codes or symbols (see Brown 1986, Dogancay-Aktuna 2006, Fang 2011, Nostrand 1989). In particular, several assertions have been made that such cultural learning experience positively contributes to learners' personal growth by increasing learners' self-awareness (e.g., Kramersch 2013).

Over the past decades, the assertion that learners should pick up knowledge of the target culture and raise cultural awareness through cultural literacy has been voiced in a multitude of earlier research in L2 teaching and learning. Therefore, in teaching young learners reading literacy, rather than setting culture as an isolated learning goal, attention needs to be paid to the target culture through access to language and culture integrated learning opportunities using forthcoming picture books for effective and natural exposure to the target language.

2.2 Token Coverage for Acceptable Comprehension

One of the most critical factors to consider when providing reading materials to L2 learners is the difficulty of vocabulary (Lewis 1997). In order to give learners pleasure in learning a foreign language through reading, it is of paramount importance to provide reading materials suitable for learners' vocabulary size (Lee 2021, Schulz 1981). Regarding the optimum threshold of token coverage, in particular, although there have been slight differences among the previous findings, one of the most conventional claims is that acceptable comprehension could be made when a text coverage threshold reaches at least 95% (Han and Lee 2022, Lee and Kim 2022). In other words, to understand a text properly, one must know more than 95% of a given vocabulary (Hirsh and Nation 1992, Laufer 1989, Nation 1990). In the same continuum, it has also been argued that L2 learners need to know more than 98% of the text's vocabulary to 'enjoy' the text for themselves without additional assistance (Hu and Nation 2000).

According to recent studies (e.g., Horst and Schmitt 2017, Nation 2006, Schmitt, Cobb, Schmitt, Jiang and Grabe 2011), even 95% token coverage is insufficient for a complete comprehension of a particular text. In a similar vein, it appears inevitable that young L2 learners of English be provided with books that reach the threshold level of token coverage to boost their literacy through reading.

In the domestic research context, it has been argued that it is critical to provide supplementary reading materials suitable for L2 learners' proficiency level. In particular, it was noted that Korean L2 learners' vocabulary levels tend to differ depending on whether or not they have had private tutoring experience (Jung and Kim 2008, Choi

and Park 2017). Sung and Kim (2021) drew notable research findings in L2 learning; though not covering original children's picture books, the research was performed using 25 English animation films containing visual representations. Based on 800 core vocabulary words in the national English curriculum devised for elementary schoolers by the Ministry of Education, the token coverage rate of each animation film was calculated respectively.

The results evidenced that only one movie reached 95% of the token coverage for young Korean learners of English, which means only one out of twenty-five movies could be adopted as class materials to enjoy and watch in elementary school settings. It was also found that 138 mandatory vocabulary words out of 800 were high-frequency words recycled across the movie. Films are instructional materials supported by audiovisual aids. Thus, it was argued that those movies that did not reach 95% of the vocabulary level could also be used as learning aids if learners focused on the high-frequency vocabulary presented before watching the movie.

3. Methods

3.1 Corpus Construction

In an attempt to pursue the proposed research goals, compiling a purpose-end corpus was of vital necessity. In multiple steps, the Corpus of Children's Picture Books (COCH) was constructed using the Project Gutenberg Digital Library as a primary data source. First, raw data selection was performed on the Project Gutenberg site to search children's picture books using keywords like *children's picture book*, *children's book*, *children's literature*, and *picture book*; through careful screening, the children's picture book candidates were selected during the first stage of analyses. As mentioned in the previous section, a *picture book* is defined as a children's picture book if the picture is as essential as the text, so we did not limit the text length during the initial data collection stage.

During the second stage, among the children's picture book candidates, we selected picture books ranked at the top 100 downloads in the Project Gutenberg archive for initial screening. Among the selections, we excluded those written in languages other than English, like Russian or French, the ones unavailable in domestic online bookstores, and finally the ones repeatedly released in multiple versions, such as a series or collection by the same author. We then shortlisted 74 items. Since a particular series contained more than one picture book, the final list had 600 picture book stories in total and consisted of 452,219 tokens with 14,981 types. As the Project Gutenberg site provides plain texts essential to constructing machine-readable files, we saved each story in a plain text file encoded with UTF-8 following the encoding standard of the vocabulary profiling program.

3.2 Baseword Lists for Vocabulary Profiling

This study conducted vocabulary profiling in two-tiered methods. To this end, we adopted two different baseword level lists to compute the token coverage of children's picture books downloaded from the Gutenberg Project archive. According to Cobb (2021), those two baseword lists are considered similar level-wise. Then, we investigated if those could become a viable source for the L2 literacy curriculum applicable in the Korean context. As the first baseword level lists, we employed VP-Kids L1 to L10, offered by the Vocab Profilers suit in the Compleat Lexical Tutor (Cobb 2021). The original version of the VP-Kids level lists was created by Stemach and Williams (1988). Later, Roessingh and Cobb (2008) developed the VP-Kids profiler in that conventional level lists like GSL, AWL, and BNC/COCA have been producing inconsistent and weak results when dealing with children's texts (see Horst and Collins 2006). The lists are based on native children's L1 oral productions and compiled ten

lists in the form of a lemma containing 250 headwords.

A children's picture book is ambivalent in that it both covers textual and verbal attributes as part of genre characteristics. For example, a nursery rhyme as one type of sub-genres holds abundant oral expressions. However, the problem lies in that the classic vocabulary level lists (e.g., GSL, NGSL, and BNC/COCA) do not contain such spoken language, meaning that those level lists might not perfectly measure the lexical range of children's texts. In order to avoid such problematic scenarios in mind, this study employed VP-Kids lists as lexical level measurement due to their further optimized and more delicate baseword vocabulary selections well-suited to children's picture books.

Specifically, when comparing BNC/COCA 1K and VP-Kids L1 lists, we could quickly notice substantial differences between the two baseword level lists. As displayed in Table 1, there were 189-word families detectable unique to the VP-Kids L1 list. These words are likely to be salient in children's literature or picture books in which both narrative and spoken discourses are pervasive. One noteworthy is that these words are rarely detected in the adult-learner-centered BNC/COCA 1K list. As shown, there found only 17-word families that both lists shared.

Table 1. Level List Comparison: BNC/COCA 1K vs. VP-Kids L1

Level List	Tokens (#)	W. Families (#)	Unique to BNC/COCA 1K (#)	Unique to VP-Kids L1 (#)	Tokens Shared (#)	W. Families Shared (#)
BNC/COCA 1K	12,823	1,069	12,716 tokens 1,052 families	732 tokens 189 families	216	17
VP-Kids L1	841	206				

**Word families shared in both lists:*

general • thick • age • meaning • mistake • address • date • expense • worth • young • east • list • weather • autumn • shall • beneath • several

**Word families unique to VP-Kids L1:*

absent • adventure • africa • alarm • antique • apron • aquarium • archer • attend • audience • backpack • badge • baptise • bead • beak • berry • bib • bible • biscuit • blackboard • blame • blind • bloom
blouse • booth • buck • cactus • cafeteria • calf • carpenter • caterpillar • cemetery • checkerboard • claw • cobra • concrete • contain • cork • cough • counter creek • crow • cruel • crumble • deliver • dent
dessert • diamond • disguise • divorce • dome • drill • envelope • erupt • eskimo • exit • experiment • eyebrow • eyelid • faint • fancy • fasten • faucet • fern • fin • fist • flame • flashlight • flavour • fog
footprint • fossil • gallop • grapefruit • grasshopper • grin • hatchet • haystack • headlight • homework • horseback • horseshoe • ignorant • infect • insect • iron • janitor • jaw • jewellery • july • kleenex
lamb • lantern • lawnmower • ledge • lily • lobster • mansion • map • meadow • medal • mood • mustard • napkin • nightgown • oar • overalls • palace • passenger • peacock • pear • peel • petal • pickle
pill • pillowcase • pineapple • platter • playmate • plumb polar • polish • quilt • raisin • ranch • receive • repair • rhino • rhinoceros • ribbon • ripe • robe • rust • salute • sandpaper • saucer • scab • scout
search • season • separate • shovel skateboard • slipper • smokestack • snail • soda • sour • splinter • stab • stadium • steam • steep • steeple • stiff • stir • task • teepee • thumb • toenail • toothpaste • traffic
trampoline triangle • trophy • tug • tulip • typewrite • valley • vase • vest • violin • volcano • wade • wallet • walnut • weak • weapon • wheelbarrow • whisker • whisper • wig • windmill • wise • workbook
wrist • yardstick • yam • zebra

During the second tier of the analyses, the KEBWL lists (Shin 2015) were utilized to judge whether the VP-Kids-based picture book candidates still reach the comparable lexical standards raised by the 2015 national English curriculum in Korea. Based on the results of the first-tier analysis, thus, we further identified whether the picture book candidates are something affordable and enjoyable to elementary schoolers in the domestic school context.

Meanwhile, in the design of the two-tiered analyses, the treatment of proper nouns was paid particular attention. According to the VocabProfilers tool built in the Complete Lexical Tutor, if proper nouns are counted as target vocabulary, the amount of vocabulary that learners need to know will increase. Therefore, the default setting is intentionally set as proper nouns counted in VP-Kids L1. Likewise, off-list unknown words appearing in the VocabProfilers tool are operated the same way as the off-list computed by the AntWordProfiler (Anthony 2021), without excluding proper nouns. In other words, off-list vocabulary refers to the words that are excluded from Kids' lists. Consequently, in this study, like said, proper nouns were included in Level 1K and calculated as such.

4. Results and Discussion

4.1 VP-Kids-based Token Coverage

This study conducted preliminary analyses to see if authentic children's picture books can be utilized as a source of L2 literacy curriculum applicable in the Korean public education settings. The corpus-based empirical studies were performed using two different level lists for vocabulary profiling, which mainly estimate the lexical difficulty of the target texts. The VP-Kids level lists compiled only for young learner texts were thus adopted to delve into the first research question by calculating the token coverage of 594 picture book candidates in the COCH.

As a result of analyzing the token coverage rate of children's picture book candidates based on VP-Kids L1 to L10, it was found that fifty-four out of 594 books reached the threshold level of 95%, assuming that L2 learners know the meaning of vocabulary words in the baseword lists. Only eight picture book candidates in the COCH corpus did not satisfy the threshold level of token coverage. Specifically, those stories with less than 95% were due to the names of story characters and exclamations (i.e., *Tom*, *Moppet*, *Sprat*, *Leerie*, *Ella*, *Polly*, *Mamma*, and *Oh*), which came out as off-list words. On the presumption that L2 learners know those character names, the rest picture book candidates were again gone through the token coverage calculation, which, in turn, enabled eight more stories to survive and finally ended up having sixty picture book stories shortlisted. Table 2 depicts the sixty candidates satisfying the threshold level of token coverage, organized concurrently with sub-genres and themes.

Table 2. VP-Kids-based Picture Book Candidates (Cut-off: 95%)

Level List	Headword	Picture Book Title	Series Title	Sub-Genre	Theme
VP-Kids L1(0)	250	NA	NA	NA	NA
VP-Kids L2 (2)	500	<i>Don't Be Good</i>	<i>More Goops and How Not to Be Them</i>	Etiquette	Daily Experience
		<i>There Was an Old Woman</i>	<i>Mother Goose or the Old Nursery Rhymes</i>	Rhyme	People
VP-Kids L3 (1)	750	<i>Street Show</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>Happy Thought</i>	<i>The Kitten's Garden of Verses poetry</i>	Poetry	The World of Nature
VP-Kids L4 (3)	1000	<i>My Mother and Your Mother</i>	<i>Mother Goose or the Old Nursery Rhymes</i>	Rhyme	Family
		<i>The Wind</i>	<i>A Child's Garden of Verses</i>	Poetry	The World of Nature
		<i>Teeny-Tiny</i>	<i>A Picture-book of Merry Tales</i>	Fiction	Daily Experience
VP-Kids L5 (3)	1200	<i>Ring-A-Ring</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>Bed in Summer</i>	<i>A Child's Garden of Verses</i>	Poetry	Daily Experience
		<i>The Counting Lesson</i>	<i>Finger Plays for Nursery and Kindergarten</i>	Song	Daily Experience
VP-Kids L6 (2)	1500	<i>The Little Jumping Girls</i>	<i>Marigold Garden</i>	Poetry	People
		<i>On the Wall Top</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
VP-Kids L7 (10)	1750	<i>System</i>	<i>A Child's Garden of Verses</i>	Poetry	Daily Experience
		<i>Ball</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>My Little Girl</i>	<i>Marigold Garden</i>	Poetry	Family
		<i>Trotty's Lesson</i>	<i>Pages for Laughing Eyes</i>	Poetry	Daily Experience
		<i>The Wee Nest</i>	<i>A Story Garden for Little Children</i>	Story	Personified Animal
		<i>The Dog</i>	<i>Tom Thumb and Other Stories</i>	Fairy Tale	Personified Animal
		<i>Baby Mine</i>	<i>Marigold Garde</i>	Poetry	Family
		<i>The Hungry Wolf</i>	<i>More Russian Picture Tales</i>	Folklore	Personified Animal
		<i>The Story of the Three Little Pigs</i>	NA	Fairy Tale	Personified Animal
		<i>I saw Three Ships</i>	<i>The Baby's Opera</i>	Rhyme	Daily Experience
VP-Kids L8 (9)	2000	<i>The Four Princess</i>	<i>Marigold Garden</i>	Poetry	Cultural Diversity
		<i>The Snowman</i>	<i>A Story Garden for Little Children</i>	Story	The World of Nature
		<i>The Moon</i>	<i>The Kitten's Garden of Verses</i>	Poetry	The World of Nature
		<i>On the Bridge</i>	<i>Pages for Laughing Eyes</i>	Poetry	Daily Experience
		<i>The Wind's Fun</i>	<i>A Story Garden for Little Children</i>	Story	The World of Nature
		<i>The Shadow Kitten</i>	<i>The Kitten's Garden of Verses</i>	Poetry	Personified Animal
		<i>King Frost</i>	<i>More Russian Picture Tales</i>	Folklore	Cultural Diversity
		<i>Jack Sprat Could Eat No Fat</i>	<i>Mother Goose or the Old Nursery Rhymes</i>	Rhyme	Daily Experience
		<i>The Squirrel</i>	<i>Finger Plays for Nursery and Kindergarten</i>	Song	Personified Animal
		<i>Three Children</i>	<i>The Baby's Opera</i>	Rhyme	People

VP-Kids L9 (13)	2250	<i>To Market, to Market to Buy a Plum Cake</i>	<i>Mother Goose or the Old Nursery Rhymes-Series</i>	Rhyme	Daily Experience
		<i>The Story of Miss Moppet</i>	NA	Fiction	Personified Animal
		<i>Willy and His Sister</i>	<i>Marigold Garden</i>	Poetry	Family
		<i>Rain</i>	<i>A Child's Garden of Verses</i>	Poetry	The World of Nature
		<i>Over the Hills and Far Away</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>A Letter to Frank</i>	<i>Tom Thumb and Other Stories</i>	Fairy Tale	Daily Experience
		<i>Little Red Riding-Hood</i>	<i>Favorite Fairy Story</i>	Fairy Tale	Daily Experience
		<i>The Cock and the Bean</i>	<i>More Russian Picture Tales</i>	Folklore	Personified Animal
		<i>The White Snake</i>	<i>Household Stories</i>	Fairy Tale	Personified Animal
		<i>At the Seaside</i>	<i>A Child's Garden of Verses</i>	Poetry	Daily Experience
		<i>Wishes</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>Blue Shoes</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		VP-Kids L10 (18)	2500	<i>Windy Nights</i>	<i>A Child's Garden of Verses</i>
<i>The Lovely Moon</i>	<i>A Story Garden for Little Children</i>			Story	The World of Nature
<i>Natural History</i>	<i>The Baby's Opera</i>			Rhyme	The World of Nature
<i>There Was a Little Boy and a Little Girl</i>	<i>Mother Goose or the Old Nursery Rhymes-Series</i>			Rhyme	People
<i>When You and I Grow Up</i>	<i>Marigold Garden</i>			Poetry	Family
<i>The Saucer Pie</i>	<i>A Story Garden for Little Children</i>			Story	Daily Experience
<i>The Stick Horse</i>	<i>A Story Garden for Little Children</i>			Story	Personified Animal
<i>A Good Play</i>	<i>A Child's Garden of Verses</i>			Poetry	Daily Experience
<i>Little Phillis</i>	<i>Marigold Garden</i>			Poetry	People
<i>A Busy Street</i>	<i>Pages for Laughing Eyes</i>			Poetry	Daily Experience
<i>First Arrivals</i>	<i>Marigold Garden</i>			Poetry	Daily Experience
<i>The Duty of the Strong</i>	<i>More Goops and How Not to Be Them</i>			Etiquette	Daily Experience
<i>Girls and Boys Come Out to Play</i>	<i>Mother Goose or the Old Nursery Rhymes-Series</i>			Rhyme	People
<i>The Lion and the Fox</i>	<i>Tom Thumb and Other Stories</i>	Fairy Tale	Personified Animal		
<i>Cat and Mouse in Partnership</i>	<i>Household Stories</i>	Fairy Tale	Personified Animal		
<i>Catching Blackbirds</i>	<i>Tom Thumb and Other Stories</i>	Fairy Tale	Personified Animal		

*Notes. The figure in parentheses () indicates the number of picture book candidates depicted in each level category. The number next to the VP-Kids levels indicates the number of headwords that L2 learners need to know to read books in each category. Picture book candidates with smaller numbers in VP-Kids levels are more accessible than those with higher numbers.

Of the sixty picture book candidates, the portion of sub-genres in the COCH corpus came out as follows: a fairy tale (13.3%, eight stories), poetry (46.7%, twenty-eight stories), rhyme (15.0%, nine stories), song (3.3%, two stories), etiquette (3.3%, two stories), fiction (3.3%, two stories), folklore (5.0%, three stories), and true story (10.0%, six stories). Likewise, the composition of themes includes daily experience (41.7%, twenty-five stories), personified animal (21.7%, thirteen stories), the world of nature (15.0%, nine stories), family (8.3%, five stories), people (10%, six stories), and cultural diversity (3.3%, two stories).

In order to finalize the picture book candidates, we followed a guiding principle that the books that L2 learners can enjoy with a relatively small number of words than others are to be placed in the upper slot. Table 3 below shows part of the analysis results, only listing those with the highest token coverage in each VP-Kids category.²

As shown, Table 3 suggests that two books are available with a threshold level of token coverage satisfied at VP-Kids L2. The results show that if L2 learners know the meaning of 250 headwords respectively in VP-Kids L1 and L2, they can understand at least 95% of the vocabulary words given in the target picture books. If two or more picture books belonged to the same slot, the one with a higher portion of L1 tokens was leveled higher. If cumulative tokens were the same in the same slot, those were arranged by order of a higher to a lower number of tokens per volume. Finally, if a picture book was part of a series, it was listed in the order of the book title and the series title. As demonstrated, overall, the results seemingly indicate that no book could be understood to a degree of more than 95% only with the words in VP-Kids L1.

² Due to the space constriction, the remaining analysis results covering the token coverage of the rest picture book candidates can be found in Appendix.

Table 3. VP-Kids-based Token Coverage (Cut-off: 95%)

Baseword Level Lists	Type (#)	Token (#)	Token Cov. (%)	Token Cov Cum. (%)	Baseword Level Lists	Type (#)	Token (#)	Token Cov. (%)	Token Cov Cum. (%)
<i>Don't Be Good</i>					<i>Street Show</i>				
VP-Kids L1	20	23	88.46	88.46	VP-Kids L1	25	32	80.00	80.00
VP-Kids L2	1	2	7.69	96.15	VP-Kids L2	3	3	7.50	87.50
VP-Kids L3	0	0	0.00	96.15	VP-Kids L3	1	3	7.50	95.00
VP-Kids L4	1	1	3.85	100.00	VP-Kids L4	0	0	0.00	95.00
Off-List Unknown	0	0	0.00	0.00	Off-List Unknown	2	2	5.00	100
Off-List Words	NA				Off-List Words	trumpets (BNC/COCA 5K)			
<i>My Mother and Your Mother</i>					<i>Ring-A-Ring</i>				
VP-Kids L1	14	19	86.36	86.36	VP-Kids L1	26	44	77.19	77.19
VP-Kids L2	0	0	0.00	86.36	VP-Kids L2	2	2	3.51	80.70
VP-Kids L3	1	1	4.55	90.91	VP-Kids L3	2	5	8.77	89.47
VP-Kids L4	1	1	4.55	95.46	VP-Kids L4	2	2	3.51	92.98
VP-Kids L5	0	0	0.00	95.46	VP-Kids L5	2	2	3.51	96.49
Off-List Unknown	1	1	4.55	100	Off-List Unknown	1	1	1.75	100
Off-List Words	NA				Off-List Words	NA			
<i>The Little Jumping Girls</i>					<i>On the Wall Top</i>				
VP-Kids L1	30	65	81.25	81.25	VP-Kids L1	19	27	71.05	71.05
VP-Kids L2	3	3	3.75	85.00	VP-Kids L2	3	4	10.53	81.58
VP-Kids L3	3	3	3.75	88.75	VP-Kids L3	1	1	2.63	84.21
VP-Kids L4	1	1	1.25	90.00	VP-Kids L4	1	1	2.63	86.84
VP-Kids L5	2	2	2.50	92.50	VP-Kids L5	2	3	7.89	94.73
VP-Kids L6	2	2	2.50	95.00	VP-Kids L6	0	0	0.00	94.73
VP-Kids L7	0	0	0.00	95.00	VP-Kids L7	1	1	2.63	97.36
Off-List Unknown	2	2	2.50	97.50	Off-List Unknown	1	1	2.63	100
Off-List Words	evil (BNC/COCA 2K)				Off-List Words	NA			
<i>I Saw Three Ships</i>					<i>The Squirrel</i>				
VP-Kids L1	27	92	79.31	79.31	VP-Kids L1	41	168	68.02	68.02
VP-Kids L2	3	10	8.62	87.93	VP-Kids L2	10	22	8.91	76.93
VP-Kids L3	2	4	3.45	91.38	VP-Kids L3	4	12	4.86	81.79
VP-Kids L4	0	0	0.00	91.38	VP-Kids L4	3	14	5.67	87.46
VP-Kids L5	1	1	0.86	92.24	VP-Kids L5	3	6	2.43	89.89
VP-Kids L6	2	2	1.72	93.96	VP-Kids L6	2	8	3.24	93.13
VP-Kids L7	0	0	0.00	93.96	VP-Kids L7	1	2	0.81	93.94
VP-Kids L8	1	4	3.45	97.41	VP-Kids L8	1	2	0.81	94.75
VP-Kids L9	0	0	0.00	97.41	VP-Kids L9	2	8	3.24	97.99
Off-List Unknown	2	2	1.72	100	Off-List Unknown	3	5	2.02	100
Off-List Words	joy (BNC/COCA 2K)				Off-List Words	feast (BNC/COCA 5K), oh (BNC/COCA 32K)			

That a text satisfies a threshold level (95%) of token coverage leads to the idea that L2 learners can recognize up to 95% of the vocabulary given in the text and can understand its content by making an educated guess of the remaining 5% of unknown words using the context. More specifically, the BNC/COCA 1K to 34K benchmark lists are composed of high-frequency vocabulary words used by native speakers of English. Conventionally, the BNC/COCA 1K to 3K are considered core vocabulary words recommended for L2 learners in an EFL environment. Compared with the level of vocabulary required in the national English curriculum of Korea, vocabulary words in BNC/COCA 1K are comparable to those covered in the elementary curriculum (grades 1 to 6). BNC/COCA 2K is for middle schoolers (grades 7 to 9); BNC/COCA 3K containing part of the Academic Word List (AWL), can be applicable to the high school curriculum (grades 10 to 11) in the Korean teaching context. Therefore, if a word is taken from a way above the core BNC/COCA lists, it may cause learners to suffer from predicting the meaning of words from the context.

From the first-tier analysis, not a small number of words were detected as off-list words. As shown in Table 3, words like *trumpets*, *evil*, *joy*, *feast*, and *oh* are listed in the off-list words slot, but to Korean L2 learners, there can

be a big chance that those words can be easily decodable. According to the BNC/COCA benchmark lists, the word *trumpets* was listed in BNC/COCA 5K; *evil* in BNC/COCA 2K, *joy* in BNC/COCA 2K, *feast* in BNC/COCA 5K, and *oh* in BNC/COCA 32K (marginal word list). Suppose such words are treated as target vocabulary for explicit learning during the class. If so, the present version of vocabulary profiling in Table 3 can be revised to include more accessible picture books in the literacy curriculum design.

Therefore, the off-list words extracted from the picture book candidates were further investigated using the BNC/COCA 1K to 34K as benchmark lists to grasp where those words are placed across the lists.³ Moreover, we attempted to seek further possibilities for more picture books to be included in the final list; thus, we conducted vocabulary profiling of those off-list words from the picture book candidates satisfied with over 90% token coverage. As listed in Table 4, the results found that 229 different types of words were recurring across the BNC/COCA 1K, which can be accessible by young learners in the Korean context. It can be inferred that the VP-Kids baseword level lists primarily created for young learners might not be comprehensive enough to encompass all children's texts.

Table 4. BNC/COCA-based Vocabulary Profiling: Off-list Words

Benchmark List	Type (#)	Type (%)	Benchmark List	Type (#)	Token (%)
BNC/COCA 1K	229	7.40	BNC/COCA 16K	16	0.52
BNC/COCA 2K	514	16.61	BNC/COCA 17K	19	0.61
BNC/COCA 3K	325	10.5	BNC/COCA 18K	11	0.36
BNC/COCA 4K	307	9.92	BNC/COCA 19K	15	0.48
BNC/COCA 5K	295	9.53	BNC/COCA 20K	6	0.19
BNC/COCA 6K	221	7.14	BNC/COCA 21K	10	0.32
BNC/COCA 7K	177	5.72	BNC/COCA 22K	8	0.26
BNC/COCA 8K	127	4.10	BNC/COCA 23K	3	0.10
BNC/COCA 9K	124	4.01	BNC/COCA 24K	4	0.13
BNC/COCA 10K	82	2.65	BNC/COCA 25K	1	0.03
BNC/COCA 11K	62	2.00	BNC/COCA 31K	130	4.20
BNC/COCA 12K	53	1.71	BNC/COCA 32K	14	0.45
BNC/COCA 13K	40	1.29	BNC/COCA 33K	65	2.10
BNC/COCA 14K	30	0.97	BNC/COCA 34K	6	0.19
BNC/COCA 15K	24	0.78	Words not in the BNC/COCA	176	5.69
BNC/COCA 16K	229	7.40	Total	3094	557

Worse, besides 514 different types of words from BNC/COCA 2K and 325 types from BNC/COCA 3K, most word types were found to be positioned way off the target level of benchmark lists. By looking at the off-list words profiled in Table 4, we can quickly figure out that those words are relatively high. It is unlikely that the picture book texts can be comprehensible with only learners' predictions.

Furthermore, 176 words not in the BNC/COCA benchmark lists were marked in the off-list words slot. In order to delve into the off-list words qualitatively, additional analyses were performed using the Vocabulary Profile Tool built in the AntWordProfiler program. The following table represents the Top 10 high-frequency off-list words across the picture book candidates with 91% to 95% token coverage rates. Among the off-list words, words of exclamations (i.e., ooh and ah) and proper nouns (i.e., characters' names) were paid particular attention. For example, words like *Kywitt*, *Grethel*, *Rapunzel*, *Havery* were not observable across the proper noun list in BNC/COCA 31K and ended up being

³ According to Nation (2016), the BNC/COCA baseword lists 31K to 35K provide additional words that are different from standard level lists in that they are compiled with some particular, purpose-end words for each category: BNC/COCA 31K for proper nouns (e.g., *Alice*), BNC/COCA 32K for marginal words such as onomatopoeia and mimetic words (e.g., *huh*), BNC/COCA 33K for compound words such as (e.g., *zookeeper*), and finally BNC/COCA 34K for acronyms (e.g., *ASAP*). These baseword lists are also formed based on the word family.

positioned in the off-list words slot. Proper nouns such as main characters' names and settings are not commonly picked up by L2 learners, especially in the early stages of forming literacy. However, they are unique words that are supposed to be acknowledged as people's names or places around the globe. By excluding proper nouns from explicit learning goals and teaching them using pictures or character webs intentionally, it would be plausible that the final literacy curriculum may contain more picture book candidates reaching the threshold level of token coverage. In addition, the word *waked* should have been covered under the headword of *wake* as a word family to mean an alternative spelling of *woke* in BNC/COCA 1K even though it is now considered a nonstandard and archaic form.

Table 5. Top 10 High-frequency Off-list Words

Rank	Off-list Word	Freq (#)	Range (#)	Token Coverage					Notes
				91%	92%	93%	94%	95%	
1	<i>kywitt</i>	48	2	24	24	0	0	0	not listed in BNC/COCA 31K (proper noun)
2	<i>grethel</i>	34	1	0	34	0	0	0	not listed in BNC/COCA 31K (proper noun)
3	<i>rapunzel</i>	23	1	23	0	0	0	0	not listed in BNC/COCA 31K (proper noun)
4	<i>havery</i>	17	4	9	5	1	0	2	not listed in BNC/COCA 31K (proper noun)
5	<i>goldenlocks</i>	13	1	0	13	0	0	0	miss-spelled <i>goldilocks</i> in the original story
6	<i>tickity</i>	13	1	0	13	0	0	0	a suffix added to make an affectionate nickname
7	<i>waked</i>	12	4	1	4	2	5	0	meant to be the past tense of <i>wake</i> ; need to be listed in BNC/COCA 1K
8	<i>dickory</i>	10	2	0	9	1	0	0	mimetic word; need to be covered in BNC/COCA 32K
9	<i>fagot</i>	9	2	0	8	1	0	0	a bundle of sticks or twigs tied up to be used as fuel (Middle English)
10	<i>rampion</i>	9	1	9	0	0	0	0	a species of wildflower

Overall, the results indicate that the more complete the baseword level lists for token coverage calculation, the more accessible picture books the final literacy curriculum can contain. Provided that baseword level lists are organized more meticulously and comprehensively, in that case, it will be far more resourceful and reliable in understanding the difficulty level of children's texts in developing the L2 literacy curriculum to fit into the domestic English classroom settings.

4.2 KEBWL-based Token Coverage

For the second-tier analysis, vocabulary in the COCH corpus was profiled based on the KEBWL lists, consisting of six different baseword level lists. KEBWL L1 comprises 800 core headwords for grades one up to six recommended by the national English curriculum of Korea. KEBWL L2 is targeted for grades seven to nine, containing 1800 core headwords; KEBWL L3 is for grades ten to twelve, composed of 400 headwords pertinent to high schoolers' career paths. KEBWL L4 is about 200 foreign words; KEBWL L5 contains 129 miscellaneous supplementary words like months, dates, and counting units. Finally, 12,289 proper nouns are contained in KEBWL L6. KEBWL L4 up to L6 are not customarily set as learning goals when designing language lessons or curricula; hence in this study, those lists were assumed to be 'known words' to L2 learners rather than 'unknown words.' To compute precise results in this study, though, BNC/COCA 31K replaced KEBWL L6, which does not encompass enough proper nouns as in the BNC/COCA 31K list holding 22,327 types.

Evaluating KEBWL-based token coverage rates is paramount when identifying whether authentic children's picture book candidates can be adopted as a source of the literacy curriculum design applicable in Korean public education. In the present study, KEBWL L1, L2, L4, and L5, except for L3, were applied along with BNC/COCA 31K instead of KEBWL L6 to complement insufficient proper nouns in the list.

According to Lee's (2014) research into comparing the vocabulary index of American school textbooks (e.g., Language Arts) and English textbooks for Korean middle school students, the vocabulary level in Korean middle school textbooks was equivalent to that of U.S. kindergarteners or first graders. Therefore, the present study further

adopted the secondary-level vocabulary the Ministry of Education (2015) recommended.

With the KEBWL L1 list, only ten picture books, leading to the one-sixth of the candidates from the first phase analysis, almost reached the token coverage threshold of 95%, suggesting that those picture books can be included in the literacy curriculum for Korean L2 learners of English. Also, with KEBWL L2, additional twenty-nine books surpassed the cutoff line. Nevertheless, there are still found twenty-one picture books positioned below the threshold level of token coverage. Five of them were placed below 90%, and the remaining sixteen books were found to fall between 90% and 95%. Table 6 shows picture book titles arranged in the order of the easiest to most complex and maps out the ten picture book candidates in KEBWL L1 and an additional twenty-nine candidates in KEBWL L2 that could reach the threshold level of token coverage across the COCH corpus.

Table 6. KEBWL-based Picture Book Candidates (Cut-off: 95%)

Level List	Headword	Picture Book Title	Series Title	Sub-Genre	Theme
KEBWL L1 (10)	800	<i>Rain</i>	<i>A Child's Garden of Verses</i>	Poetry	The World of Nature
		<i>On the Wall Top</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>Don't Be Good</i>	<i>More Goops and How Not to Be Them</i>	Etiquette	Daily Experience
		<i>My Mother, and Your Mother</i>	<i>Mother Goose or the Old Nursery Rhymes</i>	Rhyme	Family
		<i>There Was an Old Woman</i>	<i>Mother Goose or the Old Nursery Rhymes</i>	Rhyme	People
		<i>Baby Mine</i>	<i>Marigold Garden</i>	Poetry	Family
		<i>Happy Thought</i>	<i>A Child's Garden of Verses</i>	Poetry	The World of Nature
		<i>Ring-A-Ring</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>The Counting Lesson</i>	<i>Finger Plays for Nursery and Kindergarten</i>	Rhyme	Daily Experience
		<i>The Dog</i>	<i>Tom Thumb and Other Stories</i>	Fairy Tale	Personified Animal
KEBWL L2 (29)	1800	<i>The Duty of the Strong</i>	<i>More Goops and How Not to Be Them</i>	Etiquette	Daily Experience
		<i>When You and I Grow Up</i>	<i>Marigold Garden</i>	Poetry	Family
		<i>The Little Jumping Girls</i>	<i>Marigold Garden</i>	Poetry	People
		<i>The Wind</i>	<i>A Child's Garden of Verses</i>	Poetry	The World of Nature
		<i>The Moon</i>	<i>The Kitten's Garden of Verses</i>	Poetry	The World of Nature
		<i>First Arrivals</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>Blue Shoes</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>The Wee Nest</i>	<i>A Story Garden for Little Children</i>	Story	Personified Animal
		<i>Over the Hills and Far Away</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>On the Bridge</i>	<i>Pages for Laughing Eyes</i>	Poetry	Daily Experience
		<i>Bed in Summer</i>	<i>A Child's Garden of Verses</i>	Poetry	Daily Experience
		<i>Cat and Mouse in Partnership</i>	<i>Household Stories</i>	Fairy Tale	Personified Animal
		<i>System</i>	<i>A Child's Garden of Verses</i>	Poetry	Daily Experience
		<i>The Lion and the Fox</i>	<i>Tom Thumb and Other Stories</i>	Fairy Tale	Personified Animal
		<i>Catching Blackbirds</i>	<i>Tom Thumb and Other Stories</i>	Fairy Tale	Personified Animal
		<i>There Was a Little Boy and a Little Girl</i>	<i>Mother Goose or the Old Nursery Rhymes-Series</i>	Rhyme	People
		<i>The Lovely Moon</i>	<i>A Story Garden for Little Children</i>	Story	The World of Nature
		<i>The Wind's Fun</i>	<i>A Story Garden for Little Children</i>	Story	The World of Nature
		<i>The Story of Miss Moppet</i>	<i>The Story of Miss Moppet</i>	Fiction	Personified Animal
		<i>The White Snake</i>	<i>Household Stories</i>	Fairy Tale	Personified Animal
		<i>At the Seaside</i>	<i>A Child's Garden of Verses</i>	Poetry	Daily Experience
		<i>A Good Play</i>	<i>A Child's Garden of Verses</i>	Poetry	Daily Experience
		<i>Little Phillis</i>	<i>Marigold Garden</i>	Poetry	People
		<i>The Four Princesses</i>	<i>Marigold Garden</i>	Poetry	Cultural Diversity
		<i>Ball</i>	<i>Marigold Garden</i>	Poetry	Daily Experience
		<i>My Little Girlie</i>	<i>Marigold Garden</i>	Poetry	People
		<i>The Shadow Kitten</i>	<i>The Kitten's Garden of Verses</i>	Poetry	The World of Nature
		<i>Willy and His Sister</i>	<i>Marigold Garden</i>	Poetry	Family
		<i>Natural History</i>	<i>The Baby's Opera</i>	Rhyme	The World of Nature

*Notes. The figure in parentheses () next to the KEBWL levels indicates the number of picture books found in each level list category. Picture book candidates with smaller numbers in KEBWL levels are considered easier than those with higher numbers.

As provided above, it can be noticeable that the complete picture book lists are full of various genres and themes. Its sub-genres comprise poetry (50.0%, five stories), rhyme (30.0%, three stories), etiquette (10.0%, one story), and a fairy tale (10.0%, one story). When the KEBWL L1 list was based, the themes of the corpus were found to cover daily experience (40.0%, four stories), family (20%, two stories), people (10.0%, one story), personified animal (10.0%, one story), and the world of nature (20%, two stories).

Likewise, it was also found that twenty-nine stories with a good number of different sub-genres and themes were additionally readable with the KEBWL L2 list: there were six different sub-genres observable, including etiquette (3.4%, one story), fiction (3.4%, one story), poetry (58.6%, seventeen stories), rhyme (6.9%, two stories), true story (10.3%, three stories), and a fairy tale (13.8%, four stories). In the KEBWL L2 section, the themes were classified into cultural diversity (3.4%, one story), daily experience (48.3%, fourteen stories), family (13.8%, four stories), people (17.2%, five stories), personified animal (24.1%, seven stories), and the world of nature (27.6%, eight stories). Table 7 below displays the final complete picture book candidates that can be utilized as a source of a literacy curriculum aligned with the national English curriculum for grades one to six.⁴

As well-represented in Table 7, of the ten picture book candidates listed in the KEBWL L1 baseword category, not all the ten books surpassed the cutoff level. Seven of them were perfectly satisfied with the threshold level of token coverage using only the KEBWL L1 list. The other three candidates, however, were slightly below 95%, so those picture books went through additional vocabulary profiling to examine the texts qualitatively. For example, the story *Baby Mine* reached 97.88% with the help of three different word types—*sweet*, *such*, and *shower*—taken from the KEBWL L2 list. Indeed, to Korean L2 learners of English, words like *sweet* and *shower* are pretty familiar.

Those words are likely to be transcribed in the Korean language and recycled very often, meaning that those common words do not need to be set as challenging language goals in the literacy program. Therefore, it can be assumed that young Korean L2 learners of English can easily recognize the meaning of those words; thereby, those words would not affect the overall difficulty of picture books chosen. As Richards (2001) argues, when setting a language goal, a factor of similarity across the cultures needs to be considered. In addition, as depicted in the picture book *The Counting Number*, even though words like *hide* or *count* may not belong to a high-frequency vocabulary range, the words can still be taught using pictures or demonstration. Thus, such content words are worthwhile to learn through reading picture books.

The results also evidenced that only one picture book satisfied the threshold token coverage on its own without holding any proper nouns as in BNC/COCA 31K, foreign words as in KEBWL L4, or supplementary words as in KEBWL L5, which is titled *Rain*. One thing notable here is that the book *Rain* was rated as VP-Kids L7 in the first-tier analysis, indicating that there seems to be a quite wide gulf between VP-Kids baseword lists and those of KEBWL lists. It can be inferred that the 800 core vocabulary words recommended for elementary schoolers in Korea might not have been covered in the high-frequency core vocabulary lists for young native readers of English. Also, there were no off-list words noticeable from additional vocabulary profiling output.

⁴ The lexical profiling results of the remaining picture book candidates can be found in Appendix.

Table 7. KEBWL-based Token Coverage (Cut-off: 95%)

Baseword Level Lists	Type (#)	Token (#)	Token Cov. (%)	Token Cov Cum. (%)	Baseword Level Lists	Type (#)	Token (#)	Token Cov. (%)	Token Cov Cum. (%)
Rain					Don't Be Good				
KEBWL L1	18	25	100.00	100.00	KEBWL L1	23	29	93.55	93.55
					BNC/COCA 31K	1	1	3.23	96.78
Off-List Unknown	0	0	0.00	0.00	Off-List Unknown	1	1	3.23	100.01
Off-List Words	NA				Off-List Words	circus			
There Was an Old Woman					My Mother and Your Mother				
KEBWL L1	916	23	92	92	KEBWL L1	15	25	92.59	92.59
BNC/COCA 31K	1	1	4	96.00	BNC/COCA 31K	1	1	3.7	96.29
Off-List Unknown	0	0	0	0	Off-List Unknown	0	0	0	0
Off-List Words	NA				Off-List Words	NA			
On the Wall Top					Baby Mine				
KEBWL L1	25	38	90.48	90.48	KEBWL L1	16	40	85.11	85.11
KEBWL L5-S	2	2	4.76	85.24	BNC/COCA 31K	1	1	2.13	87.24
BNC/COCA 31K	1	1	2.38	97.62	KEBWL L2	3	5	10.64	97.88
Off-List Unknown	0	0	0	0	Off-List Unknown	1	1	2.13	100.01
					KEBWL L2 Words	such, sweet, shower			
Off-List Words	NA				Off-List Words	NA			
Happy Thought					Ring-A-Ring				
KEBWL L1	19	22	91.67	91.67	KEBWL L1	33	56	93.33	93.33
BNC/COCA 31K	1	1	4.17	95.84	BNC/COCA 31K	1	1	1.67	95.00
KEBWL L2	1	1	4.17	100.01	KEBWL L2	1	1	1.67	96.67
Off-List Unknown	0	0	0	0	Off-List Unknown	2	2	3.33	100
Off-List Words	NA				Off-List Words	merry, twirls			
The Counting Lesson					The Dog				
KEBWL L1	40	104	73.24	73.24	KEBWL L1	33	63	85.14	85.14
KEBWL L5-S	5	18	12.68	85.92	KEBWL L4-F	1	1	1.35	86.49
BNC/COCA 31K	1	1	0.7	86.62	BNC/COCA 31K	2	2	207	89.19
KEBWL L2	7	13	9.15	95.77	KEBWL L2	3	5	6.76	95.95
Off-List Unknown	3	6	4.23	100	Off-List Unknown	2	3	4.05	100
KEBWL L2 Words	hear, hide, nobody, once, soon, then, count				KEBWL L2 Words	bit, gun, hip			
Off-List Words	beehive, creeping, hive				Off-List Words	hen, hay			

*Notes. The KEBWL L5 baseword level list contains other supplementary words, which does not indicate that KEBWL L5 is more challenging than KEBWL L2. Likewise, the KEBWL L4 list contains foreign words, and again the KEBWL L4 and L5 lists, along with BNC/COCA 31K, are meant to be considered KEBWL L1 in the present study. The BNC/COCA 31K list was loaded to complement the proper noun list in KEBWL L6.

The following story is an excerpt from *Rain*, which can be readable with 18 types of vocabulary words pertaining to only KEBWL L1.

*The rain is raining all around.
It falls on a field and tree.
It rains on the umbrellas here,
And on the ships at sea.*

There found four books besides six picture book candidates where some off-list words were noticeable. To name a few, there found words like a *circus* in *Don't Be Good*, *merry* and *twirls* in *Ring-A-Ring*, *beehive*, *hive*, and *creeping* in *The Counting Lesson*, and *hen* and *hay* in *The Dog*. As listed, these off-list words represent cultural experiences somehow, either old or new. Given that not all these cultural activities are common in Korea, these words may need to be taught explicitly to help Korean L2 learners grasp those off-list words' linguistic and cultural meanings. How well L2 learners understand such particular cultural elements may vary depending on L2 learners'

life experiences. In the same sense, Goodman (1975) claims that young learners understand texts based on their life experiences more efficiently. Therefore, even if the number of word tokens in one story can be small, words that are embedded in cultural elements associated with young learners' lives should never be dropped in the L2 literacy curriculum. The gap between the two different baseword level lists (i.e., VP-Kids vs. KEBWL) should be understood as the disparity between the two cultures.

When setting up language goals like core vocabulary in the literacy program, thus, it can be interpreted that utilizing pictures associated with cultural elements can help promote their understanding and pick up authentic vocabulary and cultural schemata (see Webb and Rodgers 2009). Therefore, reading materials suitable for learners' vocabulary levels and enjoyable to any young learners of English irrespective of differences in economic wealth are thought to help learners be equipped with linguistic and cultural literacy skills more efficiently and have a positive motivation to learn further.

There is no alternative but to debate over whether priority should be given to vocabulary that learners will encounter throughout the early stages of literacy development or vocabulary that will assist them in learning the target language culture they wish to attain. Nevertheless, an EFL context in which L2 learners learn a target language while living in an L1 culture differs from an ESL setting wherein L2 learners learn a target language while living in a target culture (Littlewood and William 1984). In this respect, it is vital to have EFL learners exposed to additional vocabulary input on cultural elements, as language and culture are intrinsically tied.

L2 learners are likely to have constrained exposure to target culture words that are not frequently faced in daily life in an EFL setting. It might even be viable to have such learners acquire vocabulary by actively incorporating images in picture books that native-speaking children would enjoy reading to facilitate them in learning the vocabulary of cultural dimensions.

All in all, profiling vocabulary to generate complete picture book resources was conducted in an orderly manner. The raw COCH corpus compiled before profiling comprises 452,219 tokens with 14,981 types. After the first-tier analyses using VP-Kids baseword level lists, the COCH corpus has been reduced to one-tenth of the original corpus in many lexical and syntactic aspects. The comparable scale of multiple versions in the COCH corpus is mapped out in Table 8.

Table 8. The Scale Summary of the COCH Corpus

	Text File (#)	Token (#)	Type (#)	STTR (%)	MWL (#)	MSL (#)
COCH_Raw Corpus	600	452,219	14,981	95.97	4.03	21.64
VP-Kids-based Candidates	60	13,441	1,689	93.95	3.80	18.69
KEBWL L2-based Candidates	29	6,072	1,195	94.60	3.84	19.91
KEBWL L1-based Candidates	10	363	154	83.33	3.48	15.13

In the original raw corpus, the average length of sentences (i.e., MSL) led to 21.64; the average length of words (e.g., MWL) was 4.03. Both results show that comparatively longer sentences and words are commonly thought to be way too burdensome for young L2 learners to enjoy for literacy development. The same pattern was noticeable while going through the second phase of analysis. Whether KEBWEL L2- or KEBWEL L1-based, it can be readily discernible that the size of tokens and types and their word and sentence lengths turned out to be well adjusted to fit into the levels of the national English curriculum for young L2 learners in standard classroom settings.

5. Conclusion

Children's picture books produced in native L1 English have indeed been appreciated across L2 reading classes worldwide in that they promote L2 learners' linguistic and cultural literacy development. However, a viable, customized picture-book curriculum in the domestic school setting seems to have been implemented only in private educational contexts, not in public classrooms. This study thus adopted a corpus-based method of vocabulary profiling to assess the lexical difficulty of authentic children's texts produced in native L1 English. We then attempted to shortlist accessible children's picture book candidates that can be utilized as a primary source of an L2 literacy curriculum suitable for Korean public-school settings.

To this end, we extracted 600 children's picture book texts, the top 100 most frequently downloaded from the Project Gutenberg Digital Library, to assemble a Corpus of Children's Picture Books (COCH), yielding 459,321 tokens. Then, employing two different types of baseword level lists, we measured the token coverage rate of each picture book candidate. One was to assess the lexical difficulty of each text using VP-Kids L1 to L10 (Roessingh and Cobb 2008) as level lists, and the other was to use KEBWL L1 to L6 (Shin 2015), which have been compiled based on the national English curriculum of Korea.

A children's picture book is versatile in that it has both textual and oral elements as genre traits. One form of sub-genres, for example, is a nursery rhyme, which provides narrative and verbal elements. Given that conventional vocabulary level lists created for adult learners (e.g., GSL or BNC/COCA) do not include spoken language, these might not always profile children's texts' lexical range meticulously. Due to their comprehensive and more delicate baseword vocabulary selections well-suited to children's picture books, thus, this study attempted to adopt the VP-Kids lists as token coverage measurement to avoid potentially undesirable instances.

When the VP-Kids lists were applied, the results revealed that sixty children's picture book candidates surpassed the threshold level (95%) of token coverage. However, by using the list of KEBWL L1 pertinent to elementary schoolers in Korea, it was discovered that just ten titles could suit the literacy curriculum for elementary schoolers in the Korean teaching context. With KEBWL L2, twenty-nine picture books were then found to fit into secondary schoolers in Korea.

Furthermore, in respect of further investigation into off-list words, it was observed that Korean learners are unlikely to encounter particular vocabulary words caused by L1 to L2 cultural differences, so the culture-related words were left off the proposed vocabulary list for schoolchildren in Korea. In children's picture books, visual information supports textual information. Therefore, if culture-related off-list words are taught explicitly using meaningful images merged with cultural elements, the picture books whose token coverage does not exceed 95% may surpass the threshold level of comprehension.

The present study attempted to construct a corpus using free picture books provided by the Project Gutenberg archives, which is advantageous in that anyone can access them. Nevertheless, there is still a limitation that most of the picture books available in the archive were likely to be classics or those released decades ago, thereby missing works of contemporary authors. Therefore, if we could include 'copy-right free' picture books of modern writers in our corpus data, our final picture book candidates would have been more diverse in sub-genres and themes. More importantly, if there could be more thorough baseword level lists available targeted young Korean L2 learners of English, further well-tailored literacy curriculum sources would be accessible.

Early literacy skills in L2 learning are thought to impact long-term reading skills substantially. L2 learners with high initial literacy skills might anticipate strengthening their literacy over time. Those who do not have adequate early literacy skills may end up with low reading and comprehension skills in the long run (see Alexander and Entwisle 1993, Stanovich 2009). In this context, it is indisputable that not all L2 learners perform the same even if

they are taught with the same learning materials. In other words, L2 learners who receive extra support from their families would perform better in reading and vice versa. Likewise, school textbooks would be more challenging for some kids with low literacy skills. Others with strong literacy skills, on the other hand, would be discouraged by relatively easy school texts. Additional reading materials tailored to the vocabulary levels of young L2 learners may be vital to motivate both learners at varying proficiency levels. Children's picture book reading experiences will also boost their experience and performance while concurrently learning a target language and culture.

References

- Alexander, K., D. Entwisle and S. Dauber. 1993. First-grade classroom behavior: Its short-and long-term consequences for school performance. *Child Development* 64(3), 801-814.
- Anthony, L. 2021. AntConc Version 3.5.8 [computer software]. Tokyo, Japan: Waseda University. Accessed 10 Aug 2021 at <https://www.laurenceanthony.net/software>
- Brewster, J., G. Ellis and D. Girard. 2002. *The Primary English Teacher's Guide*. Harlow: Pearson Education.
- Birketveit, A. and H. Rimmereide. 2017. Using authentic picture books and illustrated books to improve L2 writing among 11-year-olds. *The Language Learning Journal* 45(1), 100-116.
- Brown, H. 1986. Learning a second culture. In J. Valdes, ed., *Culture Bound: Bridging the Cultural Gap in Language Teaching*, 33-48. Cambridge: Cambridge University Press.
- Chae, C. and S. Lee. 2010. A study on using pattern poetry to teach English reading and writing to elementary students. *Journal of the Korea English Education Society* 9(3), 275-297.
- Choi, G. and S. Park. 2017. The effects of shared-English-fairy tale reading on elementary school children's vocabulary and reading competency, and affective domains. *Journal of Education and Culture* 23(1), 343-372.
- Cobb, T. 2021. Vocab Stats v.8.3 [computer program]. Accessed 15 Aug 2021 at <https://www.lexutor.ca/stats>
- Cobb, T. and H. Roessingh, 2008. Lextutor for kids: Profiling the vocabulary of k-2 learners. Paper presented at TESOL Conference, New York.
- Dogancay-Aktuna, S. 2006. Expanding the socio-cultural knowledge base of TESOL teacher education. *Language, culture and curriculum* 19(3), 278-295.
- Donelson, K. and A. Nilsen. 2005. *Literature for Today's Young Adults*. Boston: Pearson Education.
- Elley, W. 1991. Acquiring literacy in a second language: The effect of book-based programs. *Language Learning* 41(3), 375-411.
- Ellis, G. and J. Brewster. 1991. *The Storytelling Handbook for Primary Teachers*. London: Penguin Books.
- Ellis, G., J. Brewster and D. Girard. 1992. *The Primary English Teacher's Guide*. London: Penguin.
- Enever, J. and G. Schmid-Schönbein. 2006. *Picture Books and Young Learners of English*. München: Langenscheidt.
- Fang, F. 2011. International cultures in the framework of World Englishes: What should EFL teachers do? *Journal of Asia TEFL* 8(1), 111-137.
- Goodman, Y. 1975. Reading comprehension: A redundant phrase. *Michigan Reading Journal* 9(2), 2.
- Han, J-S. and Y. C. Lee. 2022. A corpus-based analysis of the perceived difficulty increase in the 2022 CSAT English Test. *Korean Journal of English Language and Linguistics* 22, 336-354.
- Hirsh, D. and P. Nation. 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8(2), 689-696.

- Hu, M. and I. Nation. 2000. Vocabulary density and reading comprehension. *Reading in a Foreign Language* 13(1), 403-430.
- Huck, C., S. Hepler., J. Hickman and B. Kiefer. 2001. *Children's Literature in the Elementary School*. The McGraw-Hill Companies.
- Horst, M. and L. Collins. 2006. From fable to strong: How does their vocabulary grow? *Canadian Modern Language Review* 63(1), 83-106.
- Jalongo, M. 2004. *Young Children and Picture Books*. Washington, DC: National Association for the Education of Young Children.
- Jang, Y. and H. Kim. 2021. A study on the implementation and effectiveness of the English extensive reading program in a primary class using children's literature, *The Journal of Korea Elementary Education* 32(2), 85-108.
- Jung, Y. and K. Kim. 2008. A study on the influence of elementary school students' home environment and private education on English learning abilities. *English Language Teaching* 20(2), 273-295.
- Kim, H. and S. Lee. 2012. A study of teaching beginning reading based on rhymes using storybooks. *Primary English Education* 18(3), 65-96.
- Kim, Y. 2016. An analysis of phonics activities in English textbooks for Korean elementary school students. *Primary English Education* 22(4), 237-257.
- Kim, Y. 2019. An analysis of reading and writing activities in 3rd and 4th elementary English textbooks of 2015 revised curriculum. *Teacher Education Research* 58(4), 439-452.
- KICE. 2015. *Curriculum for English Language Education, Science and Technology*. Ministry of Education.
- KICE. 2016. *Analysis of the Status of Korean, Math, and English Curriculum Completion of Elementary School Students*. Ministry of Education.
- KICE. 2018. *Study Presentation on Measures to Improve School English Education*. Ministry of Education.
- Krashen, S. 2004. *The Power of Reading: Insights from the Research*. Portsmouth, NH; Heinemann.
- Kramsch, C. 2013. Culture in foreign language teaching. *Iranian Journal of Language Teaching Research* 1(1), 57-78.
- Kusanagi, Y., M. Kobayashi and M. Fukaya. 2020. An effective approach to inspire readers: Introduction to stimulating picture books in extensive reading. *Journal of Extensive Reading* 5, 95-105.
- Laufer, B. 1989. What percentage of text-lexis is essential for comprehension? In C. Lauren and M. Nordmann, eds., *Special Language: From Humans Thinking to Thinking Machines*, 316-323, Bristol, UK: Multilingual Matters.
- Lee, J. 2014. Educational gap analysis of elementary and middle school students in Seoul: Focusing on academic achievement from 2010 to 2013. *Seoul Education Research and Information Institute* 214, 142-155.
- Lee, Y. C. 2021. The predictability of receptive vocabulary size to assess college students' TOEIC® readiness. *English Language and Linguistics* 27(1), 151-176.
- Lee, Y. C. and T.-Y. Kim. 2022. The paradigm shift in English language teaching in North Korea: A corpus-assisted analysis. *Korean Journal of English Language and Linguistics* 22, 279-299.
- Lee, Y. and H. Bae. 2007. An understanding of the multicultural through elementary English teaching -centered on English multicultural picture story book reading. *Korean Journal of Elementary Education* 18(1), 96-135.
- Lewis, M. 1997. *Implementing the Lexical Approach*. NY: Language Teaching Publications.
- Littlewood, W. and L. William. 1984. *Foreign and Second Language Learning: Language Acquisition Research and Its Implications for the Classroom*. UK: Cambridge University Press.
- Lukens, R. 1990. *A Critical Handbook to Children's Literature*. New York: Harper Collins.

- Maeng, U. and S. Kim. 2011. Vocabulary proficiency of primary school students and their perception on English vocabulary. *Modern English Education* 12(2), 115-139.
- Nation, I. 1990. *Teaching and Learning Vocabulary*. MA: Newbury House.
- Nation, I. 1997. The language learning benefits of extensive reading. *The Language Teacher* 21(5), 13-16.
- Nation, I. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge Univ. Press.
- Nation, I. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63(1), 59-82.
- National Reading Panel. 2000. *Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. MD: National Institute of Child Health and Human Development.
- Nikolajeva, M. and C. Scott. 2000. The dynamics of picture book communication. *Children's Literature in Education* 31(4), 225-239.
- Norton, D. 1999. *Through the Eyes of a Child: An Introduction to Children's Literature*. Columbus, OH: Merrill.
- Nostrand, H. 1989. The beginning teacher's cultural competence: Goal and strategies. *Foreign Language Annals* 22(2). 189-193.
- Park, S. and Y. Park. 2007. Effects of literature-based literacy activities on elementary student's reading and writing proficiency. *Journal of Learner-Centered Curriculum and Instruction* 7(2), 95-117.
- Rha, K. 2011. Effects of literature-based literacy activities on elementary student's reading and writing proficiency. *Studies in English Language & Literature* 53(3), 163-184.
- Russell, B. 2009. *Human Knowledge: Its Scope and Limits*. New York: Routledge.
- Schmitt, N. 2008. Instructed second language vocabulary learning. *Language Teaching Research* 12(3), 329-363.
- Schmitt, N., T. Cobb., M. Horst and D. Schmitt. 2017. How much vocabulary is needed to use English? *Language Teaching* 50(2), 212-226.
- Schmitt, N., X. Jiang and W. Grabe. 2011. The percentage of words known in a text and reading comprehension. *The Modern Language Journal* 95(1), 26-43.
- Schulz, R. 1981. Literature and readability: Bridging the gap in foreign language reading. *The Modern Language Reading* 65(1), 43-53.
- Scott, W. and L. Ytreberg. 1990. *Teaching English to Children*. London: Longman.
- Shin, D. 2015. 2015 Korea English Basic Word List (KEBWL) RANGE program. Accessed 17 December 2021 at <https://m.cafe.daum.net/sdhera>
- Sipe, L. and J. Bauer. 2001. Urban kindergartners' literary understanding of picture storybooks. *The New Advocate* 14, 329-342.
- Sipe, L. and C. McGuire. 2006. Picture book endpapers: Resources for literary and aesthetic interpretation. *Children's Literature in Education* 37(4), 291-304.
- Stanovich, K. 2009. Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education* 189(1-2), 23-55.
- Stemach, G. and W. Williams. 1988. *WordExpress: The First 2500 Words of Spoken English*. Novato, CA: Academic Therapy Publications.
- Sung, M. and K. Kim. 2021. Analysis of the vocabulary level in English-language animated movies based on the vocabulary list in the Korean national curriculum for English. *Korean Journal of English Language and Linguistics* 21, 87-109.
- Temple, C., M. Martinez., J. Yokota and A. Naylor. 1998. *Children's Books in children's Hands: An Introduction to Their Literature*. Boston: Allyn & Bacon.

- Webb, S. and M. Rodgers. 2009. The lexical coverage of movies. *Applied Linguistics* 30(3), 407-427.
- Yoon, M. and U. Maeng. 2011. The comparison study of the effect of vocabulary instruction using storytelling and flash cards. *Journal of the Korea English Education Society* 10(3), 163-185.
- Yoon, T. 2019. Teaching primary English through children's literature with close reading annotation. *Studies in Humanities* 62, 139-158.
- Yoon, Y. 2019. An analysis of phonics in primary English textbooks of the 2015 revised national curriculum. *Journal of the Korea English Education Society* 18(1), 99-116.
- Yoon, T. and J. Yang. 2021. A case study of primary English learners' responses to an online extensive reading program. *STEM Journal* 22(1), 155-177.

Examples in: English
Applicable Languages: English
Applicable Level: Elementary

Appendix

1					2						
VP-Kids-based Token Coverage (Cut-off: 95%)					VP-Kids L5						
VP-Kids L2	Baseword	Type (°)	Token (°)	Token Cov. (%)	Token Cov. Cum (%)	Baseword	Type (°)	Token (°)	Token Cov. (%)	Token Cov. Cum (%)	
	DON'T BE GOOD						TAKEN-TIME				
VP-Kids L1	20	23		88.46	88.46	VP-Kids L1	57	206	53.51	53.51	
VP-Kids L2	1	2		7.69	96.15	VP-Kids L2	7	20	5.19	58.70	
VP-Kids L3	0	0		0.00	96.15	VP-Kids L3	9	18	4.68	63.38	
VP-Kids L4	1	1		3.85	100.00	VP-Kids L4	4	7	1.82	65.20	
VP-Kids L5	0	0		0.00	100.00	VP-Kids L5	0	0	0.00	65.20	
VP-Kids L6	0	0		0.00	100.00	VP-Kids L6	5	121	31.43	65.20	
VP-Kids L7	3	6		1.56	96.63	VP-Kids L7	3	6	1.56	98.19	
VP-Kids L8	0	0		0.00	98.19	VP-Kids L8	0	0	0.00	98.19	
VP-Kids L9	2	2		0.52	98.19	VP-Kids L9	2	2	0.52	98.19	
VP-Kids L10	0	0		0.00	98.71	VP-Kids L10	0	0	0.00	98.71	
VP-Kids L11	5	5		1.30	100	VP-Kids L11	5	5	1.30	100	
VP-Kids L12	0	0		0.00	100.00	VP-Kids L12	0	0	0.00	100.00	
VP-Kids L13	0	0		0.00	100.00	VP-Kids L13	2	2	0.52	98.19	
VP-Kids L14	0	0		0.00	100.00	VP-Kids L14	2	2	0.52	98.19	
VP-Kids L15	0	0		0.00	100.00	VP-Kids L15	2	2	0.52	98.19	
VP-Kids L16	0	0		0.00	100.00	VP-Kids L16	1	1	0.26	98.45	
VP-Kids L17	0	0		0.00	100.00	VP-Kids L17	1	1	0.26	98.71	
VP-Kids L18	0	0		0.00	100.00	VP-Kids L18	1	1	0.26	99.00	
VP-Kids L19	0	0		0.00	100.00	VP-Kids L19	0	0	0.00	99.00	
VP-Kids L20	0	0		0.00	100.00	VP-Kids L20	0	0	0.00	99.00	
VP-Kids L21	0	0		0.00	100.00	VP-Kids L21	0	0	0.00	99.00	
VP-Kids L22	0	0		0.00	100.00	VP-Kids L22	0	0	0.00	99.00	
VP-Kids L23	0	0		0.00	100.00	VP-Kids L23	0	0	0.00	99.00	
VP-Kids L24	0	0		0.00	100.00	VP-Kids L24	0	0	0.00	99.00	
VP-Kids L25	0	0		0.00	100.00	VP-Kids L25	0	0	0.00	99.00	
VP-Kids L26	0	0		0.00	100.00	VP-Kids L26	0	0	0.00	99.00	
VP-Kids L27	0	0		0.00	100.00	VP-Kids L27	0	0	0.00	99.00	
VP-Kids L28	0	0		0.00	100.00	VP-Kids L28	0	0	0.00	99.00	
VP-Kids L29	0	0		0.00	100.00	VP-Kids L29	0	0	0.00	99.00	
VP-Kids L30	0	0		0.00	100.00	VP-Kids L30	0	0	0.00	99.00	
VP-Kids L31	0	0		0.00	100.00	VP-Kids L31	0	0	0.00	99.00	
VP-Kids L32	0	0		0.00	100.00	VP-Kids L32	0	0	0.00	99.00	
VP-Kids L33	0	0		0.00	100.00	VP-Kids L33	0	0	0.00	99.00	
VP-Kids L34	0	0		0.00	100.00	VP-Kids L34	0	0	0.00	99.00	
VP-Kids L35	0	0		0.00	100.00	VP-Kids L35	0	0	0.00	99.00	
VP-Kids L36	0	0		0.00	100.00	VP-Kids L36	0	0	0.00	99.00	
VP-Kids L37	0	0		0.00	100.00	VP-Kids L37	0	0	0.00	99.00	
VP-Kids L38	0	0		0.00	100.00	VP-Kids L38	0	0	0.00	99.00	
VP-Kids L39	0	0		0.00	100.00	VP-Kids L39	0	0	0.00	99.00	
VP-Kids L40	0	0		0.00	100.00	VP-Kids L40	0	0	0.00	99.00	
VP-Kids L41	0	0		0.00	100.00	VP-Kids L41	0	0	0.00	99.00	
VP-Kids L42	0	0		0.00	100.00	VP-Kids L42	0	0	0.00	99.00	
VP-Kids L43	0	0		0.00	100.00	VP-Kids L43	0	0	0.00	99.00	
VP-Kids L44	0	0		0.00	100.00	VP-Kids L44	0	0	0.00	99.00	
VP-Kids L45	0	0		0.00	100.00	VP-Kids L45	0	0	0.00	99.00	
VP-Kids L46	0	0		0.00	100.00	VP-Kids L46	0	0	0.00	99.00	
VP-Kids L47	0	0		0.00	100.00	VP-Kids L47	0	0	0.00	99.00	
VP-Kids L48	0	0		0.00	100.00	VP-Kids L48	0	0	0.00	99.00	
VP-Kids L49	0	0		0.00	100.00	VP-Kids L49	0	0	0.00	99.00	
VP-Kids L50	0	0		0.00	100.00	VP-Kids L50	0	0	0.00	99.00	
VP-Kids L51	0	0		0.00	100.00	VP-Kids L51	0	0	0.00	99.00	
VP-Kids L52	0	0		0.00	100.00	VP-Kids L52	0	0	0.00	99.00	
VP-Kids L53	0	0		0.00	100.00	VP-Kids L53	0	0	0.00	99.00	
VP-Kids L54	0	0		0.00	100.00	VP-Kids L54	0	0	0.00	99.00	
VP-Kids L55	0	0		0.00	100.00	VP-Kids L55	0	0	0.00	99.00	
VP-Kids L56	0	0		0.00	100.00	VP-Kids L56	0	0	0.00	99.00	
VP-Kids L57	0	0		0.00	100.00	VP-Kids L57	0	0	0.00	99.00	
VP-Kids L58	0	0		0.00	100.00	VP-Kids L58	0	0	0.00	99.00	
VP-Kids L59	0	0		0.00	100.00	VP-Kids L59	0	0	0.00	99.00	
VP-Kids L60	0	0		0.00	100.00	VP-Kids L60	0	0	0.00	99.00	
VP-Kids L61	0	0		0.00	100.00	VP-Kids L61	0	0	0.00	99.00	
VP-Kids L62	0	0		0.00	100.00	VP-Kids L62	0	0	0.00	99.00	
VP-Kids L63	0	0		0.00	100.00	VP-Kids L63	0	0	0.00	99.00	
VP-Kids L64	0	0		0.00	100.00	VP-Kids L64	0	0	0.00	99.00	
VP-Kids L65	0	0		0.00	100.00	VP-Kids L65	0	0	0.00	99.00	
VP-Kids L66	0	0		0.00	100.00	VP-Kids L66	0	0	0.00	99.00	
VP-Kids L67	0	0		0.00	100.00	VP-Kids L67	0	0	0.00	99.00	
VP-Kids L68	0	0		0.00	100.00	VP-Kids L68	0	0	0.00	99.00	
VP-Kids L69	0	0		0.00	100.00	VP-Kids L69	0	0	0.00	99.00	
VP-Kids L70	0	0		0.00	100.00	VP-Kids L70	0	0	0.00	99.00	
VP-Kids L71	0	0		0.00	100.00	VP-Kids L71	0	0	0.00	99.00	
VP-Kids L72	0	0		0.00	100.00	VP-Kids L72	0	0	0.00	99.00	
VP-Kids L73	0	0		0.00	100.00	VP-Kids L73	0	0	0.00	99.00	
VP-Kids L74	0	0		0.00	100.00	VP-Kids L74	0	0	0.00	99.00	
VP-Kids L75	0	0		0.00	100.00	VP-Kids L75	0	0	0.00	99.00	
VP-Kids L76	0	0		0.00	100.00	VP-Kids L76	0	0	0.00	99.00	
VP-Kids L77	0	0		0.00	100.00	VP-Kids L77	0	0	0.00	99.00	
VP-Kids L78	0	0		0.00	100.00	VP-Kids L78	0	0	0.00	99.00	
VP-Kids L79	0	0		0.00	100.00	VP-Kids L79	0	0	0.00	99.00	
VP-Kids L80	0	0		0.00	100.00	VP-Kids L80	0	0	0.00	99.00	
VP-Kids L81	0	0		0.00	100.00	VP-Kids L81	0	0	0.00	99.00	
VP-Kids L82	0	0		0.00	100.00	VP-Kids L82	0	0	0.00	99.00	
VP-Kids L83	0	0		0.00	100.00	VP-Kids L83	0	0	0.00	99.00	
VP-Kids L84	0	0		0.00	100.00	VP-Kids L84	0	0	0.00	99.00	
VP-Kids L85	0	0		0.00	100.00	VP-Kids L85	0	0	0.00	99.00	
VP-Kids L86	0	0		0.00	100.00	VP-Kids L86	0	0	0.00	99.00	
VP-Kids L87	0	0		0.00	100.00	VP-Kids L87	0	0	0.00	99.00	
VP-Kids L88	0	0		0.00	100.00	VP-Kids L88	0	0	0.00	99.00	
VP-Kids L89	0	0		0.00	100.00	VP-Kids L89	0	0	0.00	99.00	
VP-Kids L90	0	0		0.00	100.00	VP-Kids L90	0	0	0.00	99.00	
VP-Kids L91	0	0		0.00	100.00	VP-Kids L91	0	0	0.00	99.00	
VP-Kids L92	0	0		0.00	100.00	VP-Kids L92	0	0	0.00	99.00	
VP-Kids L93	0	0		0.00	100.00	VP-Kids L93	0	0	0.00	99.00	
VP-Kids L94	0	0		0.00	100.00	VP-Kids L94	0	0	0.00	99.00	
VP-Kids L95	0	0		0.00	100.00	VP-Kids L95	0	0	0.00	99.00	
VP-Kids L96	0	0		0.00	100.00	VP-Kids L96	0	0	0.00	99.00	
VP-Kids L97	0	0		0.00	100.00	VP-Kids L97	0	0	0.00	99.00	
VP-Kids L98	0	0		0.00	100.00	VP-Kids L98	0	0	0.00	99.00	
VP-Kids L99	0	0		0.00	100.00	VP-Kids L99	0	0	0.00	99.00	
VP-Kids L100	0	0		0.00	100.00	VP-Kids L100	0	0	0.00	99.00	
VP-Kids L101	0	0		0.00	100.00	VP-Kids L101	0	0	0.00	99.00	
VP-Kids L102	0	0		0.00	100.00	VP-Kids L102	0	0	0.00	99.00	
VP-Kids L103	0	0		0.00	100.00	VP-Kids L103	0	0	0.00	99.00	
VP-Kids L104	0	0		0.00	100.00	VP-Kids L104	0	0	0.00	99.00	
VP-Kids L105	0	0		0.00	100.00	VP-Kids L105	0	0	0.00	99.00	
VP-Kids L106	0	0		0.00	100.00	VP-Kids L106	0	0	0.00	99.00	
VP-Kids L107	0	0		0.00	100.00	VP-Kids L107	0	0	0.00	99.00	
VP-Kids L108	0	0		0.00	100.00	VP-Kids L108	0	0	0.00	99.00	
VP-Kids L109	0	0		0.00	100.00	VP-Kids L109	0	0	0.00	99.00	
VP-Kids L110	0	0		0.00	100.00	VP-Kids L110	0	0	0.00	99.00	
VP-Kids L111	0	0		0.00	100.00	VP-Kids L111	0	0	0.00	99.00	
VP-Kids L112	0	0		0.00	100.00	VP-Kids L1					

11				
VP-Kids L2	4	6	9.52	82.54
VP-Kids L3	1	1	1.59	84.13
VP-Kids L4	1	1	1.59	85.72
VP-Kids L5	1	1	1.59	87.31
VP-Kids L6	0	0	0.00	87.31
VP-Kids L7	1	1	1.59	88.90
VP-Kids L8	1	1	1.59	90.49
VP-Kids L9	2	2	3.17	93.66
VP-Kids L10	1	1	1.59	95.25
OE-Last Unknown	3	3	4.76	100
OE-Last Words		whoop, halloo		
Total	37	60	0.00	100
THE LION AND THE FOX				
VP-Kids L1	23	152	72.04	72.04
VP-Kids L2	17	23	10.90	82.94
VP-Kids L3	3	3	1.42	81.36
VP-Kids L4	5	5	2.37	80.73
VP-Kids L5	4	4	1.90	88.63
VP-Kids L6	3	11	5.21	93.84
VP-Kids L7	1	1	0.47	94.31
VP-Kids L8	0	0	0.00	94.31
VP-Kids L9	1	1	0.47	94.78
VP-Kids L10	1	1	0.47	95.25
OE-Last Unknown	8	10	4.47	100
OE-Last Words		hem, hal, holden, mirth, murther, clever, froggy, sashy, yalla, omelet, 33, 65, 55, 25%		
Total	108	201	0.00	100
CAT AND MOUSE IN PARTNERSHIP				
VP-Kids L1	153	698	71.22	71.22
VP-Kids L2	46	109	11.12	82.34
VP-Kids L3	24	41	4.18	86.52
VP-Kids L4	21	32	3.27	89.79
VP-Kids L5	11	15	1.53	91.32
VP-Kids L6	7	16	1.63	92.85
VP-Kids L7	4	8	0.82	93.77
VP-Kids L8	4	4	1.04	94.18
VP-Kids L9	4	4	0.41	94.59
VP-Kids L10	6	6	0.61	95.20
OE-Last Unknown	35	47	4.80	100
OE-Last Words		acquaintance, evening, expect, night, common, second, longish, refuse, universal, demand, omelet, constant, rational, canon, frock, meander, greasy, slosh, folk, devour, leaped		
Total	315	580	0.00	100
CATCHING BLACKBIRDS				
VP-Kids L1	14	20	8.89	85.33
VP-Kids L2	5	5	2.22	87.55
VP-Kids L3	4	4	1.78	89.33
VP-Kids L4	4	4	1.78	91.11
VP-Kids L5	4	4	1.78	92.85
VP-Kids L6	1	6	2.67	93.78
VP-Kids L7	1	1	0.44	94.22
VP-Kids L8	0	0	0.00	94.22
VP-Kids L9	1	1	0.44	94.66
VP-Kids L10	1	1	0.44	95.10
OE-Last Unknown	9	11	4.89	100
OE-Last Words		blackbirds (*with omitted 92-95, 10%)		
Total	65	93	0.00	100

Appendix B

KBEWL-based Token Coverage (Cut-off: 95%)

Baseword	Type (%)	Token (%)	Token Cov. (%)	Token Cov. Cum (%)
ON THE WALL TOP				
KBEWL L1	25	38	90.48	90.48
KBEWL L5	1	1	4.76	85.74
BNC-COCA 3HK	1	1	2.38	97.62
KBEWL L2	1	1	2.38	100
OE-Last Unknown	0	0	0	0
OE-Last Words			NA	
Total	29	42	0	100
DON'T BE GOOD				
KBEWL L1	23	29	93.55	93.55
BNC-COCA 3HK	1	1	3.23	96.78
OE-Last Unknown	1	1	3.23	100.01
OE-Last Words			circus	
Total	25	31	0	100

12				
MY MOTHER AND YOUR MOTHER				
KBEWL L1	15	25	92.59	92.59
BNC-COCA 3HK	1	1	3.7	96.29
OE-Last Unknown	0	0	0	99.99
OE-Last Words		NA		
Total	17	27	0	100
THERE WAS AN OLD WOMAN				
KBEWL L1	916	23	92	92
BNC-COCA 3HK	1	1	4	96
KBEWL L2	1	1	4	100
OE-Last Unknown	0	0	0	0
OE-Last Words		NA		
Total	18	25	0	100
BABY MINK				
KBEWL L1	16	40	85.11	85.11
BNC-COCA 3HK	1	1	2.13	87.24
KBEWL L2	3	5	10.64	97.88
OE-Last Unknown	1	1	2.13	100.01
OE-Last Words		oh		
Total	21	47	0	100
HAPPY THOUGHT				
KBEWL L1	19	22	91.67	91.67
BNC-COCA 3HK	1	1	4.17	95.84
KBEWL L2	1	1	4.17	100.01
OE-Last Unknown	0	0	0	0
OE-Last Words		NA		
Total	21	24	0	100
RING-A-RING				
KBEWL L1	33	56	93.33	93.33
BNC-COCA 3HK	1	1	1.67	95
KBEWL L2	1	1	1.67	96.67
OE-Last Unknown	2	2	3.33	100
OE-Last Words		meety, twits		
Total	37	60	0	100
THERE WAS A LITTLE BOY AND A LITTLE GIRL				
KBEWL L1	19	23	88.53	88.53
BNC-COCA 3HK	1	1	1.67	90
KBEWL L2	1	1	1.67	91.67
OE-Last Unknown	1	1	3.33	100
OE-Last Words		alley, oh		
Total	24	29	0	100
THE DOG				
KBEWL L1	33	63	85.14	85.14
KBEWL L4	1	1	1.35	86.49
BNC-COCA 3HK	2	2	2.07	89.59
KBEWL L2	3	5	6.76	95.95
OE-Last Unknown	2	3	4.05	100
OE-Last Words		hen, hay		
Total	41	74	0	100

Baseword	Type (%)	Token (%)	Token Cov. (%)	Token Cov. Cum (%)
THE DUTY OF THE STRANG				
KBEWL L1	29	60	91.3	91.3
BNC-COCA 3HK	1	1	1.45	92.75
KBEWL L2	4	5	7.25	100
OE-Last Unknown	0	0	0	0
OE-Last Words		NA		
Total	34	69	0	100
WHEN YOU AND I GROW UP				
KBEWL L1	34	50	87.72	87.72
BNC-COCA 3HK	2	2	3.51	91.23
KBEWL L2	5	5	8.77	100
OE-Last Unknown	0	0	0	0
OE-Last Words		NA		
Total	41	57	0	100
THE LITTLE JUMPING GIRLS				
KBEWL L1	38	52	61.9	61.9
KBEWL L4	2	2	2.37	90.47
BNC-COCA 3HK	1	1	1.19	91.66
KBEWL L2	7	7	8.33	99.99
OE-Last Unknown	0	0	0	0
OE-Last Words		NA		
Total	48	84	0	100
THE WIND				

13				
VP-Kids L1	11	11	100	100
VP-Kids L2	1	1	100	100
VP-Kids L3	1	1	100	100
VP-Kids L4	1	1	100	100
VP-Kids L5	1	1	100	100
VP-Kids L6	1	1	100	100
VP-Kids L7	1	1	100	100
VP-Kids L8	1	1	100	100
VP-Kids L9	1	1	100	100
VP-Kids L10	1	1	100	100
OE-Last Unknown	1	1	100	100
OE-Last Words				
Total	12	12	100	100

14				
VP-Kids L1	1	1	100	100
VP-Kids L2	1	1	100	100
VP-Kids L3	1	1	100	100
VP-Kids L4	1	1	100	100
VP-Kids L5	1	1	100	100
VP-Kids L6	1	1	100	100
VP-Kids L7	1	1	100	100
VP-Kids L8	1	1	100	100
VP-Kids L9	1	1	100	100
VP-Kids L10	1	1	100	100
OE-Last Unknown	1	1	100	100
OE-Last Words				
Total	12	12	100	100

15				
VP-Kids L1	1	1	100	100
VP-Kids L2	1	1	100	100
VP-Kids L3	1	1	100	100
VP-Kids L4	1	1	100	100
VP-Kids L5	1	1	100	100
VP-Kids L6	1	1	100	100
VP-Kids L7	1	1	100	100
VP-Kids L8	1	1	100	100
VP-Kids L9	1	1	100	100
VP-Kids L10	1	1	100	100
OE-Last Unknown	1	1	100	100
OE-Last Words				
Total	12	12	100	100