


Article

Extensive Reading and Science Vocabulary Learning in L2: Comparing Reading-Only and Reading-While-Listening

Raquel Serrano 

Department of Modern Languages and Literatures and English Studies, University of Barcelona,
08007 Barcelona, Spain; raquelserrano@ub.edu

Abstract: This paper presents a study analyzing second language vocabulary gains after an extensive reading program that included non-fiction graded readers of scientific content in English. The study was conducted in a Spanish primary school ($N = 96$) and implemented in two different modalities: reading-only and reading-while-listening, which included audiobooks. The study lasted one school year and involved 39 science graded readers, making it unique in its duration and scope. The findings indicate that the practice of extensive reading resulted in notable improvements in vocabulary acquisition during the first half of the school year; however, the advantages were less evident in the second half. Different factors intrinsic to the program but also related to students' motivation will be discussed in order to explain the findings.

Keywords: extensive reading; graded readers; academic vocabulary learning

1. Introduction

The current study presents an analysis of immediate and long-term vocabulary gains after an extensive reading (ER) program that employed non-fiction books of scientific content in English. The target books were graded readers, which were adapted to the learners' proficiency. The ER program was implemented in a Spanish primary school in two different modalities: reading-only and reading-while-listening. As Nation [1,2] suggests, ER should be an important component of second language (L2) learning programs. There is research that supports the benefits of ER for a variety of L2 areas, mostly, vocabulary learning, reading comprehension, and reading fluency [3]. Most studies on ER have focused on older learners, mainly college students, and programs that used fiction graded readers. In spite of this, it has been suggested that ER programs could also be beneficial for young learners [4] and that non-fiction books should also be included in order to cater to the diverse reading preferences of students [5]. Although including a variety of genres is usually recommended, a study focusing solely on non-fiction books related to a particular subject, for example natural sciences, would be a promising avenue for research. On the one hand, such a study could potentially add to the existing body of literature on ER and its impact on L2 acquisition. Additionally, it could provide valuable insights into how science graded readers could contribute to the acquisition of scientific vocabulary and enhance science learning. This type of investigation would be particularly informative for schools that follow the Content and Language Integrated Learning (CLIL) approach and offer science education in English.

There are several variables that have been demonstrated to impact the extent to which vocabulary acquisition can occur through reading. One such variable is the reading mode, and several studies have compared reading with audio support, also known as assisted reading or reading-while-listening (RWL), and reading only (RO). These studies indicate that RWL tends to be more advantageous [6,7], although in some instances, the differences between the two reading modes have not been found to be statistically significant [8,9]. Therefore, additional research is necessary to gain further insight into how the presence of audio affects the degree of vocabulary learning from reading.



Citation: Serrano, R. Extensive Reading and Science Vocabulary Learning in L2: Comparing Reading-Only and Reading-While-Listening. *Educ. Sci.* **2023**, *13*, 493. <https://doi.org/10.3390/educsci13050493>

Academic Editors: Graham Crookes and Martin Howard

Received: 22 March 2023

Revised: 12 April 2023

Accepted: 10 May 2023

Published: 13 May 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The present study contributes to the literature on the effects of ER for L2 development, and offers new insights by comparing RO and RWL, including an under-researched population (i.e., young learners in primary school) reading graded readers of scientific content instead of fiction, as is common in the literature. Another novelty of the present study is that it involves a program that lasted one school year and that included 39 graded readers, in contrast to most studies that have examined shorter periods and fewer books.

In the following subsections, a review of the literature will be presented with the aim of contextualizing the current study and identifying the research gaps that motivated this investigation.

1.1. Extensive Reading

Nation [1] asserts in his influential paper on “The Four Strands” that a well-structured L2 curriculum must maintain an equilibrium among four essential components: meaning-oriented input, meaning-oriented output, language-focused activities, and fluency practice. For the first strand, he advocates ER as an effective means to promote language learning. In a subsequent paper primarily addressed to language teachers, Nation [2] further suggests that “The single most effective change a teacher can make to a language course is to include an extensive reading program” (p. 6). Similarly, after performing a meta-analysis on the benefits of extensive reading, Nakanishi [4] (p. 6) concludes: “In sum, the available research to date suggests that extensive reading improves students’ reading proficiency and should be a part of language learning curricula”.

At the heart of an ER program is the objective of providing L2 learners with the opportunity to read extensively, typically at a rate of one book per week, silently and independently over an extended period [5]. The main objective is for students to read fluently and focus on understanding the meaning of the texts, which is why it is crucial that the books are at the right level for the students. These principles are what Waring and McLean [10] considered the “core elements” of ER, as it has been operationalized in the literature. However, according to the authors, there are other “variable elements”, many of which were included in the Top Ten Principles of ER [5], that may or may not be present, such as reading exclusively for pleasure or allowing students to choose books.

Research suggests that ER is beneficial for different L2 aspects, including reading fluency [11–14], and reading comprehension [15,16]. There is also evidence that reading books in an L2 contributes to vocabulary learning. Nakanishi’s meta-analysis on ER research [4] found that the effect of ER for vocabulary learning was larger ($d = 1.25$) than for reading comprehension ($d = 0.72$) or reading speed ($d = 0.61$).

Numerous studies have demonstrated that L2 learners can expand their vocabulary through reading. Empirical support is derived, in part, from controlled investigations in which participants read one or a few books containing a subset of target words that were not readily available beyond the confines of the book(s). Waring and Takaki [17] examined the learning of 25 non-words that replaced real words in one graded reader (5872 words) in the case of 15 Japanese college students. The results of the immediate post-tests showed that on average, participants were able to recognize the form of 15.3 of the target words, the meaning of 10.6 and were able to translate 4.6. These gains decayed over time to 8.4 target words on the form-recognition test, 6.1 on the meaning-recognition test and 0.9 in the translation test.

Pellicer-Sánchez and Schmitt [18] also performed a controlled study to analyze how much vocabulary a group of 20 university EFL students from Spain gained after reading a single non-adapted novel in English (around 67,000 words), which included some African words, 34 of which were selected as target. The results of the study are interesting, not only because they show that vocabulary can be learned through reading but also because they demonstrate that some components of vocabulary knowledge are more easily learned by reading a book than others (see also [19]), with participants being able to recognize the meaning of 43% of the target words in a multiple-choice test, but being able to actively

recall the meaning of only 14%. Spelling recognition and word class recall fell in between, with 34% and 20% gains, respectively.

Although these studies provide insights as to how vocabulary can be learned from reading while controlling for several potentially confounding variables, such as exposure to the target words outside the program, such manipulations in research do not reflect real learning conditions. In ER programs, books are not meant to be the only source of input, as ER is only one part of a larger vocabulary-learning program that also includes other types of meaning-focused input, as well as opportunities for deliberate learning of words in isolation [2]. Additionally, ER programs promote reading of many different books, which is why research is necessary in order to examine how vocabulary knowledge develops thanks to reading multiple books through an extended period of time in authentic classroom-learning situations.

The study by Webb and Chang [20] examines vocabulary learning gains in an ER program that included audiobooks, in a Taiwanese high school. The participants ($N = 82$) read and listened to the same 10 graded readers over the course of 13 weeks. After RWL, the students were given the opportunity to engage in various post-reading activities, such as writing book reports and maintaining a learning journal. The authors administered a meaning recognition test, including 100 words that appeared in the graded readers, at three testing times: pre-test, post-test, and delayed post-test. Approximately half of those words were already known at the time of the pre-test and the authors calculated the relative gains, considering the words that the students learned out of those that were unknown. The results showed relative gains of 44.06% of the target words on the immediate post-test and 36.66% on the three-month delayed post-test. These results were in sharp contrast with those of a control group that followed their regular L2 classes with no ER, who showed 5.19% gain on the post-test and no gain on the delayed post-test.

According to one of the principles proposed by Day and Bamford [5], in ER, reading is its own reward. However, many classroom-based studies, as Webb and Chang [20] above, include post-reading activities, and some authors have even stated that ER should be integrated within the language curriculum and not included as an isolated self-contained activity [21]. The study conducted by Boutorwick et al. [22] further examined to what extent including reading activities associated with the books contributed to L2 learning by comparing vocabulary-learning gains in an ER program that did not include post-reading activities (ER-only) and another one which included post-reading discussions in small groups (ER-plus). The study took place in a university in New Zealand, where the students read the same five graded readers. The results of the study suggest that, while both approaches led to vocabulary gains in word association knowledge, the students in the ER-plus group made more gains in mid-frequency target words that were focused on during language related episodes in their small-group discussions. These findings would support the implementation of post-reading activities in ER programs. Through these activities learners have the opportunity to continue practicing the vocabulary encountered in their reading materials. These results align with those of previous studies that have examined shorter texts and found that “reading plus” conditions, which incorporate post-reading vocabulary exercises, facilitate greater vocabulary acquisition than “reading only” conditions that solely focus on meaning comprehension [23,24].

1.2. Reading Only (RO) versus Reading-While-Listening (RWL)

It has been demonstrated that vocabulary learning can be affected by the input mode through which the students are exposed to the target words. In this respect, learners' exposure could be exclusively aural (listening), written (RO), or a combination of both written and aural (RWL). Additionally, vocabulary learning may occur in multimodal conditions, in which written and aural input is also supported by the presence of images representing the target words, with those images being either static (books with pictures) or dynamic (videos). It has been shown that learners interact with written text differently in multimodal conditions including audio and pictorial support [25–27], and that multimodal

conditions facilitate vocabulary learning [28]. Although the books used in the current study also contained pictures, the main aim of the present investigation only concerns in which way the presence or absence of aural input affected vocabulary learning.

While some studies have investigated the effectiveness of ER programs that incorporate audiobooks (e.g., [20]), there is a lack of research comparing the effectiveness of RO versus RWL in the case of ER. Brown et al. [8] conducted one of the first studies in this direction. The authors examined vocabulary learning gains among Japanese college students who were learning English as a foreign language (EFL) and read three graded readers in either RO, RWL, or listening only (LO) modes. To ensure that vocabulary gains were solely attributed to reading, the authors substituted 28 words in each book with non-words. The multiple-choice test results revealed that 48% of the target non-words were acquired in the RWL mode, compared to 45% in the RO mode and 29% in the LO mode for meaning recognition. However, in the translation test, the gains were smaller, with only 16%, 15%, and 2% in the RWL, RO, and LO modes, respectively. Notably, the differences between LO and the other two modes were statistically significant, while the differences between RO and RWL were not.

Webb and Chang [6] further compared RO and RWL in the context of repeated reading in a Taiwanese high school, in which EFL learners read 28 short stories of approximately 300 words, and around 2 min 30 s in the RWL mode, over two seven-week periods. The students had to read each story at least twice with the main aim of understanding and enjoying the content, although they could also ask questions or use dictionaries if necessary. The results of the vocabulary tests showed that repeated reading was a successful approach that encouraged vocabulary learning in both modes, but RWL was significantly more beneficial than RO.

More recently, several studies have compared the two reading modes for the acquisition of collocations, instead of single words. Webb and Chang [7] provided further evidence in support of RWL versus RO for learning of English collocations through a graded reader, in the case of college students in Taiwan. These results were replicated in Vu and Peters' [29] study with Vietnamese EFL college students, who read three different fiction graded readers. According to both studies, RWL could be more beneficial because the audio support helps learners segment the input into meaningful chunks, which facilitates comprehension. The incorporation of the prosodic features of the audio is claimed to be especially useful in the noticing and processing of collocations. However, the advantages of RWL over RO are still far from being established.

Dang et al. [30] investigated the acquisition of collocations presented in various modes during an academic lecture. These modes included RWL and RO, as well as LO, viewing, and viewing with captions. The study was conducted with Chinese college students. In contrast to the previously cited studies, RO facilitated the learning of academic collocations more than RWL. According to the authors, the content of the lecture might have been challenging for learners, who might have lacked enough resources to notice and focus on the target collocations while doing LO or RWL, in contrast to the self-paced RO mode.

In a similar vein, Tuzcu [9] did not find the RWL mode to be more advantageous than the RO mode for the acquisition of medical collocations in the case of college students learning English in the US. One of the reasons the author presents for the lack of advantage of the RWL mode is that the learners' proficiency was quite high and they might have read ahead of the audio, which could have slightly disrupted the reading process. In line with claims made by Conklin et al. [31], we need a better understanding of how reading alignment (or misalignment) with the audio affects processing and vocabulary learning in RWL. Another study by Dang et al. [32], focusing on single words, also failed to find differences between RO and RWL to an academic lecture.

The conflicting results from the previously described studies suggest that more research is necessary in order to investigate whether vocabulary learning from reading is encouraged more easily by RO, or by including audio support, as in RWL. Considering previous findings, text genre might also influence whether one input mode is more benefi-

cial than the other; more specifically, it seems that academic vocabulary learning does not benefit as much from the RWL mode.

1.3. Research Gaps and Research Questions

There are some research gaps that the current study aims to fill. First, there is a lack of research on vocabulary learning through ER that exclusively consists of science graded readers. Considering previous findings from studies focusing on academic language, the presence of the audio in ER programs that include non-fiction graded readers might be less beneficial than in studies that have used fiction graded readers. Apart from the theoretical interest in investigating the conditions in which audio support is beneficial for vocabulary learning from reading, examining ER programs that include science graded readers would also be relevant for pedagogical reasons. Obtaining some insights into how vocabulary learning can be fostered through science graded readers would be useful in many contexts in which primary school students learn science in English through CLIL.

Second, not many studies on ER have focused on younger learners in primary school, and the studies that have examined the effect of reading mode (RO vs. RWL) have mostly included adults. Nakanishi's [4] meta-analysis on ER research did not include any ER program with children, but the results suggest that the benefits of ER programs increase with participants' age. Despite this finding, in his recommendations for further research, Nakanishi advocates for further research involving younger learners, positing that ER programs could be particularly motivating for this age group (see also [33,34]). Research is needed, therefore, to shed more light on the potential benefits of ER for a younger population. Since the publication of Nakanishi's meta-analysis, some studies have appeared focusing on primary school L2 learners, analyzing L2 gains as well as students' attitudes towards ER [34–36]. Overall, studies with younger learners indicate that ER is as effective as teacher-fronted instruction and has the added benefit of being particularly motivating for students. In light of the previously mentioned gaps, the present study aims to answer the following research questions:

1. To what extent can primary school students learn science vocabulary through an ER program including graded readers of scientific content?
2. Are there differences between RO and RWL?

The data from the present study is part of a larger project that, apart from vocabulary, examined the development of different aspects of L2 learning, including listening and reading comprehension and fluency, and students' L2 learning motivation [35,36]. The present study focuses on vocabulary learning through the whole academic year, and adds to the findings reported by Tragant et al. [36], which measured different L2 areas during the first half of the year.

2. Materials and Methods

2.1. Participants

For this study, 96 10–11-year-old students in four intact grade 5 classes at the target school were considered, comprising the entire available population for this grade. Within those classes, two were randomly assigned to an extensive RWL program ($N = 47$; 20 girls), one to a RO program ($N = 24$; 11 girls) and another one served as a control group ($N = 25$; 10 girls). As will be seen in the results, not all the students were present at all testing times. The students attended a school in Barcelona, partly funded by the government, where most families could be considered middle-class and highly educated: according to a background questionnaire administered to the students, 70% of the mothers in the student sample had a university degree. Most students had an English proficiency of A1, according to their teacher's estimation. All the students had good knowledge of Catalan and Spanish, which are the official languages in Barcelona.

The school provides more English hours than the typical schools in Spain and has a strong focus on reading (e.g., the students always carry a book in their backpack, which they read when they finish the assigned activities in any class period). All the students

received 7 h of English instruction per week: 3 h were devoted to regular EFL lessons, 2 h to science in English following the CLIL approach, and the remaining 2 h were devoted to extensive RO/RWL or to communicative practice in the case of the control group.

2.2. The ER Program

The ER program ran from October through May, roughly aligning with the school year, which goes from September until June. The program aimed to enhance the children's English language skills by exposing them to a large amount of input in this language. It also supported their learning of science and vocabulary in their English science class.

The ER program included 39 science graded readers, 21 of which were read during the first part of the school year (October–February) and 18 during the second part (February–May). For the sake of simplicity, this paper refers to term 1 and term 2, respectively, for these two blocks, even though these terms did not correspond to the actual academic terms. The students devoted one class session (50 min) to each book, except for three sessions in term 1 that included two short books. The books were chosen from the following collections: Macmillan Science Readers, Macmillan Children's Discover, Oxford Read and Discover, and Benchmark Education. Some titles included in the collections were *Volcanoes*, *Life in the Forest*, *Animals at Night*, *The Power of Storms*, etc. The books were between 15–31 pages long, and their difficulty increased as the school year advanced to adapt to learners' proficiency. Considering the importance of choosing reading material that is appropriate for the learners [3], in term 1, the selected graded readers were aimed at grades 3–4 and had on average 908 words; while in term 2, they were 1615 words on average and were aimed at grades 4–5. During the first term, the duration of the audio in the RWL condition was 12 min, whereas in the second term the average was 19 min. The researchers together with the teachers decided which book(s) should be read in each session and all students read the same book(s).

2.3. Reading Procedure

In order to adapt to the learners' needs and preferences, the procedure was slightly different in term 1 and 2. In term 1, at the beginning of each session, the teachers distributed the material, which included the books, a dictionary, and a workbook, which contained different post-reading activities related to the books. Additionally, the learners in the RWL condition received headphones, and MP3 players. When the students had all the material, they first browsed through the assigned book and then started reading/reading-while-listening independently. After a first reading, they were asked to write eight words they would like to remember and their corresponding L1 translation, either by using a dictionary, asking the teacher, a classmate, etc. After this, the students were asked to read the book independently again, unless the book was very long (audio 20 min or longer), in which case they only re-read some parts. After re-reading, the students were instructed to write a minimum of three questions, either true/false or multiple-choice, based on the content of the book. At the conclusion of the reading session, all materials were collected.

In order to learn about students' attitudes towards the ER program, at the end of term 1, a series of interviews were conducted [35]. Despite the overall positive attitudes towards the reading sessions, some aspects of the program received less favorable feedback. Most notably, students expressed dissatisfaction with the limited time allotted for writing questions about the content of the books after the second reading. Furthermore, the students displayed limited interest in reading the books twice.

Bearing in mind the students' feedback, and considering the classroom-based nature of our research project, adjustments were deemed necessary to encourage further engagement in ER during the second term. Instead of requiring students to write their own questions about the book's content, we provided a set of wh- and true/false comprehension questions for each book read in term 2. This new approach allowed students to further work on the content of each graded reader without having to read the entire book again, which was a requirement in term 1.

2.4. Vocabulary Tests

Two different vocabulary tests were designed in order to examine students' learning of science vocabulary through extensive RO/RWL throughout the school year. The first test contained target words that appeared in the books read during term 1, while the second test contained target words from the books read during term 2. Both tests were L2-L1 matching vocabulary tests, which measured students' meaning recognition and were similar to the tests used by Webb and Chang [20]. Meaning recognition was chosen as the target vocabulary component, because it develops earlier and it was considered that learners' vocabulary gains would be better captured through this type of test [8,18].

Each test included 50 items in total, all of which were nouns, distributed over 10 blocks, each containing five target words and five L1-matching translations (provided in both Catalan and Spanish when the word was not the same in the two languages), plus one distractor. The distractor was semantically related to one of the target words. A special effort was made to avoid cognate words. Five different versions of each test, including the same 10 blocks but in a different order, were created in order to prevent cheating. See Figure 1 for an example.

1. tar	carretera ____
2. fence	tanca/valla ____
3. tide	arrel/raíz ____
4. root	quitrà/alquitrán ____
5. road	marea ____
	branca/rama ____

Figure 1. Excerpt from the vocabulary test.

Both vocabulary tests had the same format and number of items, but each contained different target words. Appendix A provides a comprehensive list of all the words featured in each test, along with their frequency, as determined by Lextutor [37]. Additionally, we calculated the frequency of the target words in the graded readers, their glossaries, and in the science English coursebook. Although many of the words on the tests were likely to be unknown to most learners, the tests were designed so that children would recognize at least some of the words, thereby avoiding frustration throughout the 50-item vocabulary tests [20]. The suitability of the target words was confirmed by the teachers before the tests were administered. According to Cronbach alpha, both Test 1 (0.909) and Test 2 (0.913) achieved good reliability scores.

2.5. Data Collection Procedure

For vocabulary Test 1, which comprised the vocabulary included in the science graded readers read during the first term, data were collected over three testing times, pre-test (end of September), post-test (February), delayed post-test (June). In contrast, vocabulary test 2 assessed the vocabulary in the books read during the second term and was administered only twice with the pre-test conducted in February and the post-test in June. As June marked the end of the school year, further testing could not be carried out.

The tests were administered in written form in class, where researchers carefully went through the instructions and guided the students through the sample item, which included a block of three words familiar to the students (*house, garden, dog*). A time limit of 20 min was allotted for each test. As explained before, the data included in this study are part of a larger research project [35,36]. Therefore, apart from taking the vocabulary tests, the students took other tests in the same session, which lasted 50 min. The vocabulary tests were always administered first.

2.6. Analyses

The scoring system for the vocabulary tests awarded students one point for accurately matching each L2 target word with its corresponding L1 pair, and zero points for incorrect matches. The results of the vocabulary tests were analyzed in terms of relative gains, following previous research [20]. This approach considers students' initial vocabulary knowledge, which varied both within and between groups, as expected in a real classroom setting. The formula that was used for each participant's scores was the following: $[(\text{post-test score} - \text{pre-test score}) / (\text{total number of items} - \text{pre-test score})] \times 100$. This formula produces scores that indicate the percentage of words learned by students at the end of each term, relative to the words they were capable of learning based on their pre-test scores. Relative gains were also computed to examine long-term learning of the words included in term 1, but instead of using the post-test scores in the above formula, the delayed post-test scores were used.

Statistical analysis was conducted using SPSS version 27 [38]. Since all the data were normally distributed, according to the Kolmogorov–Smirnov test, two one-way ANOVAs were performed with the relative vocabulary gains experienced by the learners in the three conditions (RO, RWL, and control). The first ANOVA analyzed the relative gains with respect to the vocabulary that appeared in term 1 books. This analysis included immediate and long-term gains. The second ANOVA focused on the immediate gains related to the vocabulary included in term 2 books.

3. Results

3.1. Vocabulary Learning in Term 1

Table 1 shows the descriptive statistics obtained from Test 1, which includes the scores of the pre-test, post-test, and delayed post-test, as well as the immediate and delayed gains. The scores on the pre-test were very similar across conditions. As expected, the students knew some of the vocabulary included in the test. On the post-test, all the learners demonstrated an increase in vocabulary knowledge, but the gains between the two testing times were more obvious for the two ER groups. It can also be observed that the learning gains were maintained through time (four months later), as evidenced by the delayed post-test scores.

Table 1. Descriptive statistics Test 1: means and *SD* in parentheses.

Group	Pre-Test/50	Post-Test/50	Delayed Post-Test/50	Immediate Gains (%)	Long-Term Gains (%)
RO (<i>n</i> = 22)	27.31 (10.5)	37.09 (10.8)	35.94 (10.3)	46.7% (27.5)	44.9 (24.7)
RWL (<i>n</i> = 47)	28.97 (9.4)	39.00 (9.9)	38.04 (9.5)	54.1% (27.0)	50.4% (26.2)
Control (<i>n</i> = 25)	28.96 (9.3)	33.80 (9.7)	32.80 (10.5)	20.9% (36.0)	12.6% (37.9)

In the delayed post-test there were some students who were absent due to other school-related projects and the sample was: RO = 10, RWL = 41, control = 10.

The results of the first ANOVA comparing term 1 relative gains across conditions show that there were significant differences between the three groups in immediate and long-term gains: $F(2, 93) = 10.29, p < 0.001, \eta^2 = 0.184$, and $F(2, 69) = 7.41, p = 0.001, \eta^2 = 0.184$, respectively. Post-hoc analyses with Bonferroni adjustments for multiple comparisons applied to the immediate-gain scores suggest that the mean difference (*MD*) between the control and the RWL groups was statistically significant ($MD = -33.21, p < 0.001$); this was also the case between the control and the RO groups ($MD = -25.76; p = 0.012$). The difference between RO and RWL was not statistically significant ($MD = -7.45, p = 1.00$). Similarly, Bonferroni comparisons for long-term gains indicate that the difference between the control and RO groups was statistically significant ($MD = -32.20, p = 0.12$), as well as between the control and RWL groups ($MD = -37.75; p = 0.001$), with no differences between RO and RWL ($MD = -5.55; p = 1.00$).

3.2. Vocabulary Learning in Term 2

Test 2 pre-test scores were very similar across conditions and very similar to those of Test 1, which suggests that the level of difficulty was equivalent considering students' initial knowledge. However, the results of the post-test and the vocabulary gains were different from those reported for term 1. The students learned fewer words in term 2, and the group that made the smallest gains was the RO (see Table 2). According to the results of the ANOVA, there were no differences in relative gains among the three conditions: $F(2, 95) = 0.995, p = 0.373, \eta^2 = 0.021$.

Table 2. Descriptive statistics Test 2: means and SD in parentheses.

Group	Pre-Test/50	Post-Test/50	Immediate Gains (%)
RO ($n = 24$)	28.83 (9.6)	30.50 (8.71)	5.90% (26.77)
RWL ($n = 47$)	28.59 (10.3)	33.04 (9.6)	17.92% (35.4)
Control ($n = 25$)	27.76 (8.7)	31.48 (11.2)	14.84% (37.5)

4. Discussion

The present classroom-based study examined the contribution of RO and RWL to English science vocabulary learning through an ER program in a primary school in Spain. The ER program lasted for the whole school year and vocabulary learning was examined at the end of two terms: term 1 which included 21 graded readers that were read from October until February; and term 2, which concerned 18 books that were read from February until the end of May. The results of term 1 and term 2 will be discussed separately.

4.1. Vocabulary Learning in Term 1

The results of vocabulary Test 1 revealed that the use of science graded readers in both the RO and RWL conditions promoted gains in vocabulary knowledge among the students. These gains were significantly higher than those observed in the control group, both immediately after the intervention and in the long-term, four months later. These positive findings align with previous studies that have shown the effectiveness of fiction books in enhancing vocabulary acquisition.

The current ER program produced an average immediate and long-term relative gain in vocabulary of 50.43% and 47.64%, respectively, when combining both approaches (RO and RWL). These gains are consistent with those reported in controlled studies where vocabulary learning was incidental, such as Pellicer-Sánchez and Schmitt [18], who reported a 46% increase in meaning recognition on an immediate post-test. The present results are also similar to those of Brown et al. [8] and Waring and Tataka [17], who reported immediate vocabulary gains of over 40%. Furthermore, the present study aligns with research that has examined the implementation of ER in naturalistic classroom settings, such as Webb and Chang's [20] study on Taiwanese high school students, which reported gains of 44.06% and 36.6% on the immediate and delayed post-tests, respectively. It is worth emphasizing that the vocabulary gains achieved in the current ER program were more successfully retained than in previous studies [8,17,20]. The target students were able to recall most of the words they had learned in term 1 four months later at the end of the school year, experiencing only a slight average loss of 2.8%.

In summary, the findings from term 1 of the current study confirm previous findings regarding the positive effects of ER with fiction graded readers, and thus support ER programs including science books for the learning of scientific vocabulary.

Regarding the second research question, the results of Test 1 showed that, even though RWL promoted more immediate vocabulary gains than RO (54.1% vs. 46.7%) and also better long-term gains (50.4% vs. 44.9%), the difference did not reach statistical significance. These findings are in contrast with those reported in studies that have used fiction graded readers for the learning of collocations, such as Vu and Peters [29] or Webb and Chang [7]. According to these studies, RWL is more beneficial than RO because the audio helps learners

to segment the input in a more target-like fashion, which could facilitate the noticing and processing of L2 collocations. Similarly, in the context of repeated reading of short stories, Webb and Chang [6] also found that the RWL mode encouraged higher vocabulary gains.

On the other hand, the results of our investigation are consistent with Brown et al.'s [8] study, which found gains of 48% and 45% in the RWL and RO conditions, respectively, using three fiction graded readers. Similarly, studies focusing on academic vocabulary acquisition [9,30,32] have failed to show differences between RWL and RO. These findings support the argument that the complexity of scientific content may demand more attentional resources from learners, potentially limiting the benefits of audio input.

Additionally, when implementing RWL, it is crucial to consider whether learners' natural reading pace is aligned with the audio. As Tuzcu [9] suggests, misalignment between the two modes of input might account for cases where RWL does not provide a significant advantage. Given the complexity of the content of the books in the current study, it is plausible that the audio was too fast for some participants, which could have undermined the benefits of RWL.

4.2. Vocabulary Learning in Term 2

The analyses of vocabulary gains experienced during the second term were not in line with those reported during term 1, although there is one common finding, which is the lack of significant differences between the RO and RWL modes, even if the scores in the RWL mode were descriptively higher. Considering this result, it is important to conduct more studies comparing the two modes for different types of text genres (e.g., fiction vs. non-fiction), vocabulary targets (e.g., single words vs. collocations), and learner populations (younger vs. older learners) to explain the conflicting findings in the literature regarding the benefit of audio-supported reading for vocabulary learning.

The vocabulary gains observed in term 2 were disappointing when compared to the gains in term 1. Although students had similar prior knowledge of the target words in Test 1 and Test 2, the percentage of vocabulary gains in term 2 was low (RO = 5.90%; RWL = 17.91%). These gains are more in line with studies investigating incidental vocabulary acquisition through reading alone as opposed to "reading plus" conditions [39].

Although there was a certain intentional learning component in the ER program under study, it is worth noting that the vocabulary task required students to choose their own words to focus on, but this approach may not have been the most effective to promote engagement with vocabulary. During the interviews, many students reported facing various difficulties with this task. For some, it was challenging to identify unknown words in the graded readers or to find some words in the dictionary, while others reported choosing words they already knew to avoid the extra effort of looking up the meaning of new words. We can speculate that these challenges became more frustrating towards the end of the program than at the beginning, which could partly explain the lower vocabulary gains in term 2. Another possible explanation for the higher gains in term 1 may be connected to the benefits of repeated reading [6]. As explained in Section 2.3, the students were instructed to read the whole graded reader twice during term 1 but not during term 2.

Additionally, it should be highlighted that, in order to adapt to the learners' proficiency development throughout the year, the graded readers in term 2 were slightly more difficult than those used in the first part of the year. This difference in complexity may have also influenced the results.

Furthermore, the length of the ER program may have played a role in the lack of clear benefits observed in the second half of the school year. Learners may have lost over time their initial motivation for doing something different from their regular English classes, especially in the RO group, which obtained the lowest vocabulary gains in term 2. As noted in Tragant and Vallbona [35], only 31.8% of the students in that group reported an interest in continuing with the ER sessions the following year, compared to 62.5% in the RWL group. It is worth mentioning that the availability of MP3 players for the RWL group

probably served as an additional motivational factor for the young learners. The higher motivation to do RWL as opposed to RO has also been observed in other studies [8,27].

Another aspect that could account for the results is related to the timing of Test 2, which was administered a few weeks before the end of the school year. By this point, students may have been less motivated to take tests, particularly as they had already completed exams in their other courses.

In summary, the ER program in term 2 did not yield results as positive as in term 1, with lower overall vocabulary gains (11.9% gains vs. 50.4% in term 1) that were not significantly different from the control group. Several factors could have contributed to the ER treatment being less supportive of vocabulary learning, such as the employment of more demanding graded readers, the potential decrease in students' motivation during the program's final weeks, or the timing of Test 2 at the end of the school year.

4.3. Limitations and Future Research

The present study has limitations that stem primarily from its classroom-based approach and its emphasis on ecological validity. One limitation is that the divergent outcomes observed between term 1 and term 2 could be attributed to one or more factors that are difficult to disentangle. To differentiate these potentially influential factors, further controlled studies are necessary. First, future research should consider including the same reading procedure as well as the same type of post-reading activities for reading comprehension and vocabulary learning throughout the program, ensuring that students are consistently exposed to effective and engaging exercises. Secondly, it is important to ensure that all books are of comparable difficulty, in order to prevent discrepancies in learning outcomes related to this factor. Thirdly, learners' motivation could be thoroughly and systematically examined throughout the program and statistical analyses could be performed to assess more rigorously the role of motivation in students' vocabulary gains throughout the ER program.

Furthermore, future research could aim to control for exposure to the target words both within and outside of the ER program. The vocabulary-focused activity used in this study may have resulted in varying degrees of engagement with the target words among learners, which were not considered in the analyses. Regarding exposure to the target words outside ER, aiming for complete control is probably an unrealistic goal in a long-term ER program. Previous studies that have attempted to achieve such control used non-words and were more limited in their scope, typically including only a few graded readers. It is crucial to emphasize, however, that ER is not expected to be the sole source of L2 vocabulary learning [2].

Another avenue for research would be to explore how the number of encounters with the target words in the graded readers affect vocabulary gains, as examined in previous studies [20]. A further line of inquiry could involve investigating how text and audio alignment affects vocabulary learning by analyzing learners' eye movements while reading in the RWL mode.

Despite the limitations, the findings of this exploratory classroom-based study hold ecological validity, allowing for more straightforward pedagogical implications to be drawn compared to studies with more controlled learning conditions. The study investigated an authentic ER program, which was implemented based on input from teachers and students and was aimed at serving their interests, rather than being purely research-focused.

It should be mentioned, however, that certain aspects of the current reading program did not adhere to Day and Bamford's recommendations for ER [5]. For instance, the program did not provide a range of literary genres, but rather focused on scientific vocabulary learning, which was a primary objective of the study. Additionally, learners were not allowed to choose their own books in order to facilitate the investigation of vocabulary acquisition. However, some scholars have argued that these factors may not be essential components of ER programs, as defined in the literature [10].

5. Conclusions and Pedagogical Implications

The present study provides valuable insights into the effectiveness of an ER program that incorporates science graded readers for improving vocabulary learning among primary school students. Although most previous research on ER has focused on older learners, this study highlights the potential benefits of non-fiction ER for primary school students. However, further research is needed to replicate these findings in different contexts to obtain generalizable results.

The current findings have important pedagogical implications, particularly for schools that offer science instruction in English as a foreign language, as in CLIL contexts. Using ER programs that include science graded readers can facilitate not only the acquisition of new vocabulary but also the learning of scientific content. In fact, when the target students were asked to describe what they had learned during the ER program, most of them referenced information related to the content of the graded readers, rather than language-related aspects [35].

A second recommendation would be for teachers to include audiobooks. Although the vocabulary gains between RWL and RO were not statistically significant in the present study, RWL consistently showed higher gains. Additionally, it was found that the students in the RWL group had significantly more positive attitudes towards ER than their peers in the RO group [35]. Maintaining motivation is crucial, and teachers should regularly assess their students' attitudes towards ER to determine the optimal duration of the program. Additionally, it is important to keep learners engaged while they are being tested by incorporating additional motivational strategies that encourage them to perform their best during tests. Doing so would make it easier to gauge more accurately the learning outcomes after ER.

Another pedagogical recommendation in light of the findings from the present study, and also considering previous findings reported in the literature [20,22], is to include stimulating vocabulary-focused post-reading activities to maximize the degree of students' engagement with the vocabulary that appears in the books.

To conclude, the present study reflects the complexity of classroom-based research and underscores the importance of collaboration among L2 acquisition researchers, practitioners, and L2 learners in generating insights with practical implications for language education. Such collaboration has the potential to bridge the gap between research and teaching practice, ultimately leading to more effective support for L2 learning.

Funding: This research was funded by the Spanish Science and Innovation Ministry grants FFI2013-40952-P and PID2019-110536GB-I00.

Institutional Review Board Statement: The study adhered to the ethical principles proposed by the Research Ethics Committee at the University of Barcelona (2020-2022) and respected the privacy and confidentiality of the participants involved.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are unavailable due to privacy restrictions.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Test 1					Test 2				
Words	Freq. Band	Freq. Readers	Freq. Glossary	Freq. Courseb.	Words	Freq. Band	Freq. Readers	Freq. Glossary	Freq. Courseb.
bark	2	4	2	0	arrow	4	3	0	0
bat	2	14	0	1	lane	2	5	0	0
beetle	6	27	0	1	cattle	3	3	0	0
blood	1	8	0	0	stream	2	6	1	0
bone	1	12	3	2	glass	1	32	2	0
brain	2	8	0	0	goods	3	5	0	0
bud	5	5	0	0	race	1	2	0	0
burrow	6	4	1	0	law	1	8	0	0
claw	5	3	1	2	candle	4	2	0	0
crust	4	15	1	5	saddle	5	3	0	0
eagle	4	8	0	1	tar	7	10	1	0
feather	2	15	2	2	fence	2	3	0	0
field	1	3	0	0	tide	2	6	0	1
flea	7	12	0	0	root	2	10	0	1
flock	4	7	0	0	road	1	15	2	0
flood	2	2	0	0	shelter	2	2	1	0
fox	2	11	2	0	shark	5	58	0	2
fur	2	33	3	2	squid	8	10	0	1
hill	1	5	0	2	heat	1	10	0	6
hole	1	14	3	0	cliff	2	15	0	3
kite	6	14	0	0	steam	2	5	1	0
land	1	26	1	19	nostril	6	3	1	0
moth	6	10	0	0	gold	1	4	0	0
mud	2	13	0	2	reef	5	17	1	0
nest	2	26	2	0	layer	3	14	0	14
paw	6	3	0	0	hut	4	5	0	0
poison	2	9	2	0	shield	3	3	0	0
pond	4	4	0	1	path	2	13	0	0
seal	2	8	2	0	army	2	7	0	0
seed	2	25	2	1	earthquake	5	12	1	4
silk	3	12	2	0	shell	2	7	1	4
skin	1	13	1	4	eel	8	3	0	0
sloth	11	6	1	0	pipe	2	10	0	1
slug	6	6	0	0	patern	2	5	1	0
snail	6	11	0	1	health	1	9	0	0
soil	2	15	3	2	stall	4	2	0	0
stem	3	12	0	1	crop	3	2	1	0
swarm	6	6	0	0	fin	6	2	0	2

Test 1					Test 2				
Words	Freq. Band	Freq. Readers	Freq. Glossary	Freq. Courseb.	Words	Freq. Band	Freq. Readers	Freq. Glossary	Freq. Courseb.
tadpole	10	10	2	0	lake	1	36	1	4
tail	1	12	2	10	clay	4	4	0	0
tooth	1	6	1	0	wall	1	14	0	2
trunk	2	7	1	5	coal	2	2	0	2
tusk	10	12	1	0	concrete	3	6	0	0
walrus	12	2	1	0	jellyfish		9	0	1
wasp	6	13	0	0	twin	2	5	0	0
wave	1	7	0	3	octopus	9	6	0	1
web	4	10	1	0	limestone		19	0	0
wheel	1	3	0	0	cabbage	6	3	0	0
wing	2	48	1	6	danger	1	12	3	0
wolf	2	6	1	0	rope	2	7	0	0
Mean	3.66	11.3	0.9	1.46	Mean	3.12	9.08	0.36	0.98
Median	2	10	1	0	Median	2	6	0	0

References

- Nation, P. Is It Worth Teaching Vocabulary? *TESOL J.* **2021**, *12*, e564. [CrossRef]
- Nation, P. The Four Strands. *Innov. Lang. Learn. Teach.* **2007**, *1*, 2–13. [CrossRef]
- Nation, P.; Waring, R. *Teaching Extensive Reading in Another Language*; Routledge: New York, NY, USA, 2020.
- Nakanishi, T. A Meta-Analysis of Extensive Reading Research. *TESOL Q.* **2015**, *49*, 6–37. [CrossRef]
- Day, R.R.; Bamford, J. Top Ten Principles for Teaching Extensive Reading. *Read. Foreign Lang.* **2002**, *14*, 136–141.
- Webb, S.; Chang, A.C.-S. Second Language Vocabulary Learning through Extensive Reading with Audio support: How do frequency and distribution of occurrence affect learning? *Lang. Teach. Res.* **2015**, *19*, 667–686. [CrossRef]
- Webb, S.; Chang, A.C.-S. How Does Mode of Input Affect the Incidental Learning of Collocations? *Stud. Second Lang. Acquis.* **2022**, *44*, 35–56. [CrossRef]
- Brown, C.; Waring, R.; Donkaewbua, S. Incidental Vocabulary Acquisition from Reading, Reading-While-Listening, and Listening to Stories. *Read. Foreign Lang.* **2008**, *20*, 136–163.
- Tuzcu, A. The Effects of Input Mode in Learning Declarative and Nondeclarative Knowledge of L2 Collocations. *System* **2023**, *113*, 103006. [CrossRef]
- Waring, R.; McLean, S. Exploration of the Core and Variable Dimensions of Extensive Reading Research and Pedagogy. *Read. Foreign Lang.* **2015**, *27*, 160–167.
- Beglar, D.; Hunt, A. Pleasure Reading and Reading Rate Gains. *Read. Foreign Lang.* **2014**, *26*, 29–48.
- Beglar, D.; Hunt, A.; Kite, Y. The Effect of Pleasure Reading on Japanese University EFL Learners' Reading Rates. *Lang. Learn.* **2012**, *62*, 665–703. [CrossRef]
- Bui, T.; Macalister, J. Online Extensive Reading in an EFL Context: Investigating Reading Fluency and Perceptions. *Read. Foreign Lang.* **2021**, *33*, 1–29. Available online: <http://hdl.handle.net/10125/67391> (accessed on 1 February 2023).
- McLean, S.; Rouault, G. The Effectiveness and Efficiency of Extensive Reading at Developing Reading Rates. *System* **2017**, *70*, 92–106. [CrossRef]
- Bell, T. Extensive Reading: Speed and Comprehension. *Read. Matrix* **2001**, *1*, 1–13.
- Krashen, S. Extensive Reading in English as Foreign Language by Adolescents and Young Adults: A Meta-Analysis. *Int. J. Foreign Lang. Teach.* **2007**, *7*, 23–29.
- Waring, R.; Takaki, M. At What Rate Do Learners Learn and Retain New Vocabulary from Reading a Graded Reader? *Read. Foreign Lang.* **2003**, *15*, 130–163.
- Pellicer-Sánchez, A.; Schmitt, N. Incidental Vocabulary Acquisition from an Authentic Novel: Do Things Fall Apart? *Read. Foreign Lang.* **2010**, *22*, 31–55.
- Pigada, M.; Schmitt, N. Vocabulary Acquisition from Extensive Reading: A Case Study. *Read. Foreign Lang.* **2006**, *18*, 1–28.
- Webb, S.A.; Chang, A.C.-S. Vocabulary Learning through Assisted and Unassisted Repeated Reading. *Can. Mod. Lang. Rev.* **2012**, *68*, 267–290. [CrossRef]
- Green, C. Integrating Extensive Reading in the Task-Based Curriculum. *ELT J.* **2005**, *59*, 306–311. [CrossRef]

22. Boutorwick, T.J.; Macalister, J.; Elgort, I. Two Approaches to Extensive Reading and Their Effects on L2 Vocabulary Development. *Read. Foreign Lang.* **2019**, *31*, 150–172.
23. Peters, E.; Hulstijn, J.H.; Sercu, L.; Lutjeharms, M. Learning L2 German Vocabulary through Reading: The Effect of Three Enhancement Techniques Compared. *Lang. Learn.* **2009**, *59*, 113–151. [[CrossRef](#)]
24. Sonbul, S.; Schmitt, N. Direct Teaching of Vocabulary after Reading: Is It Worth the Effort? *ELT J.* **2010**, *64*, 253–260. [[CrossRef](#)]
25. Tragant, E.; Pellicer-Sánchez, A. Young EFL Learners' Processing of Multimodal Input: Examining Learners' Eye Movements. *System* **2019**, *80*, 212–223. [[CrossRef](#)]
26. Pellicer, A.; Tragant, E.; Conklin, K.; Rodgers, M.; Serrano, R.; Llanes, À. Young Learners' Processing of Multimodal Input and Its Impact on Reading Comprehension: An Eye-Tracking Study. *Stud. Second Lang. Acquis.* **2020**, *42*, 577–598. [[CrossRef](#)]
27. Serrano, R.; Pellicer, A. Young L2 Learners' Online Processing of Information in a Graded Reader during Reading-Only and Reading-While-Listening Conditions: A Study of Eye-Movements. *Appl. Linguist. Rev.* **2022**, *13*, 49–70. [[CrossRef](#)]
28. Montero Perez, M. Second or Foreign Language Learning through Watching Audio-Visual Input and the Role of On-Screen Text. *Lang. Teach.* **2022**, *55*, 163–192. [[CrossRef](#)]
29. Vu, D.V.; Peters, E. Incidental Learning of Collocations from Meaningful Input: A Longitudinal Study into Three Reading Modes and Factors that Affect Learning. *Stud. Second Lang. Acquis.* **2021**, *43*, 1–23. [[CrossRef](#)]
30. Dang, T.N.Y.; Lu, C.; Webb, S. Incidental Learning of Collocations in an Academic Lecture through Different Input Modes. *Lang. Learn.* **2022**, *72*, 728–764. [[CrossRef](#)]
31. Conklin, K.; Alotaibi, S.; Pellicer-Sánchez, A.; Vilkaite-Lozdiene, L. What Eye-Tracking Tells Us about Reading-Only and Reading-While-Listening in a First and Second Language. *Second Lang. Res.* **2020**, *36*, 257–276. [[CrossRef](#)]
32. Dang, T.N.Y.; Lu, C.; Webb, S. Open Access Academic Lectures as Sources of Incidental Vocabulary Learning: Examining the Role of Input Mode, Frequency, Type of Vocabulary, and Elaboration. *Appl. Linguist.* **2022**, amac044. [[CrossRef](#)]
33. Lightbown, P.M. Can They Do It Themselves? A Comprehension-Based ESL Course for Young Children. In *Comprehension-Based Second Language Teaching*; Courchene, R., St John, J., Therrien, C., Glidden, J., Eds.; University of Ottawa Press: Ottawa, ON, Canada, 1992; pp. 353–370.
34. Tragant, E.; Muñoz, C.; Spada, N. Maximizing Young Learners' Input: An Intervention Program. *Can. Mod. Lang. Rev.* **2016**, *72*, 234–257. [[CrossRef](#)]
35. Tragant, E.; Vallbona, A. Reading while Listening to Learn: Young EFL learners' Perceptions. *ELT J.* **2018**, *72*, 395–404. [[CrossRef](#)]
36. Tragant, E.; Llanes, À.; Pinyana, À. Linguistic and Non-Linguistic Outcomes of a Reading-while-listening Program for Young Learners of English. *Read. Writ.* **2019**, *32*, 819–838. [[CrossRef](#)]
37. Cobb, T.; Lextutor. Vocabprofile [Computer Program]. Available online: <https://www.lex tutor.ca/vp/comp/> (accessed on 17 March 2023).
38. IBM Corp. *IBM SPSS Statistics for Windows, Version 27.0*; IBM Corp: Armonk, NY, USA, 2020.
39. Laufer, B. Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *Can. Mod. Lang. Rev.* **2003**, *59*, 567–588. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.