Reading and Writing (2019) 32:819–838
https://doi.org/10.1007/s11145-018-9886-x

Linguistic and non-linguistic outcomes of a reading-while-listening program for young learners of English

Elsa Tragant Mestres[1] · **Àngels Llanes Baró**[2] · **Àngels Pinyana Garriga**[3]

Published online: 26 July 2018

Abstract
Reading-while-listening may be especially well suited for young language learners because of the multimodality provided in many graded readers aimed at this age group (ie., the presence of oral and written text and illustrations). This study compares a group of students who were exposed to 18 sessions of reading-while-listening with a group exposed to the same number of sessions through reading-only, and a control group. Linguistic outcomes show that students in the two intervention groups obtained higher vocabulary gains than those in the control group but did not present superior scores in reading or listening comprehension or reading fluency. Non-linguistic outcomes showed a clear preference on the part of the students for the reading-while-listening mode of input. The study concludes that the lack of differences in comprehension and fluency gains may be due to the fact that graded readers for children are too short; the input they offer is too limited to make a difference in areas other than attitudes and vocabulary learning.

Keywords Reading while listening · Extensive reading · Vocabulary learning · Young learners · Assisted reading · Graded readers

✉ Elsa Tragant Mestres
tragant@ub.edu

Àngels Llanes Baró
allanes@dal.udl.cat

Àngels Pinyana Garriga
mangels.pinyana@uvic.cat

[1] Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain

[2] Universitat de Lleida, Plaça de Víctor Siurana 1, 25003 Lleida, Spain

[3] Universitat Vic, Sagrada Família 7, 08500 Vic, Spain

Introduction

Reading-while-listening (RWL), which consists of reading while simultaneously listening to an oral rendition of the text, is an instructional practice that has been used to different extents in the context of first and second language education. In the domain of literacy instruction, RWL has been widely used and researched both at school and in the home (i.e., Koskinen et al., 2000; Rasinski & Hoffman, 2003). However, in contrast to first language (L1) research, RWL has not received as much attention in the domain of second or foreign language (L2) acquisition (SLA).

This form of bimodal input (participants are exposed to the same text through two modalities: written text and auditory) has been mostly used with English as a Foreign Language (EFL) in the context of extensive reading programs. The aim of these schemes is to offer a rich source of comprehensive input in order to compensate for the lack of (quality) input in contexts in which access to the L2 is difficult. A variation of RWL is RWL to a text repeatedly (often referred to as assisted repeated reading [RR]), a procedure that involves reading the same text/s several times in order to promote reading fluency and comprehension (Chang, 2012). RR has been extensively used with young learners with no reading disabilities in the L1 acquisition domain, but little research has been carried out in L2 acquisition (Chang and Millett, 2013).

RWL is especially well suited to implementation in EFL classrooms, either as an element of a course or as a complementary activity outside the classroom. It can be carried out with children who are not experienced readers as well as with older more mature readers as an additional source of input in contexts where exposure to the L2 is restricted to the textbook and printables. The benefits of this instructional practice are likely to go beyond the development of L2 reading skills, but we do not know if there are any differential effects between young and older learners. In the following two sections we review the research conducted to date with these two age groups.

Literature review

RWL in adults and teenagers

The scarce research on the effects of RWL programs with adults and teenagers has mostly focused on the impact of these programs on vocabulary (Webb & Chang, 2012), fluency (Chang, 2012) and comprehension development (Beglar, Hunt, & Kite, 2012), as well as on participants' perceptions of this type of practice (Lightbown, 1992). One of the domains that has been shown to benefit the most from RWL is vocabulary learning, since students are able to consolidate their previous knowledge of vocabulary and learn new words in context. Early studies of incidental vocabulary learning through RWL of single texts already showed that

the audio support promoted vocabulary learning (Horst, Cobb & Meara, 1998) and that it did so to a larger extent than listening-only (LO) (Brown, Waring, & Donkaewbua, 2008). Nevertheless, the reported gains in receptive knowledge in Brown et al.'s university students were lower (16%) than those reported in later studies such as Webb and Chang (2012) on assisted RR, where adolescent learners were reported to have vocabulary gains ranging from 24 to 29%. These differences may be explained by a number of features in the two reading programs. First, while the students in Brown et al. read the texts once and were not allowed to interact, ask questions, or use dictionaries, Webb and Chang's students read the texts a minimum of two times, had access to dictionaries, and were given the opportunity to report and discuss the content of what they read. Differences could also be due to the fact that Webb and Chang used an instrument that was sensitive to partial gains in vocabulary knowledge (a modified version of Paribakht and Wesche's Vocabulary Knowledge Scale, 1993) while Brown et al. used a meaning translation test.

The characteristics of the RWL program also seem to have made a difference in Han and Chen's case study (2010) of a heritage speaker of Chinese at senior college. Their subject, who engaged in reading and listening to authentic texts for a total of 40 h, experienced higher vocabulary gains than the students in Brown et al. (2008). In that case, accuracy rates of incidental words were reported to range from 45 to 55%. As in Webb and Chang's study (2012), the program involved assisted RR and the learner had the chance of talking about the text and asking questions. In addition, she also practised reading the text orally and received feedback from the researcher on a regular basis. These instructional strategies probably contributed to explaining the superior outcomes.

More recent research by Chang and associates has further substantiated the idea that the vocabulary gains obtained from RWL or assisted RR to graded readers are high. This is so in the case of adolescents and university students in learning both single words (Webb & Chang, 2015) and collocations (Webb, Newton, & Chang, 2013). An additional finding in Webb and Chang's more recent study (2015) is the role of proficiency in explaining incidental vocabulary learning, with higher level learners having significantly larger relative gains than lower-level participants. These results led the authors to conclude that prior vocabulary learning may have a large impact in explaining the amount of vocabulary learning that is made through extensive reading.

In sum, L2 research seems to indicate that RWL can have positive effects on incidental vocabulary learning, especially when it is accompanied by RR and certain instructional strategies (i.e., access to dictionaries, chances to talk about the books). It also indicates that the greater their proficiency, the greater the benefit learners are likely to derive from the oral rendering of a written text.

Two more domains that might also benefit from RWL are comprehension and reading fluency, but this is an issue that has not been analysed in depth. Most of the very few studies of the subject carried out to date have been led by Taguchi and Chang together with their respective associates. Chang and Millet (2014), for example, compared L2 listening fluency (defined as the automatic processing of aural input with a reasonable degree of comprehension) of three groups of participants:

Reading-only (RO), Reading-while-listening (RWL) and Listening-only (LO). 113 EFL university students were distributed in one of the three groups and it was found that after the 13-week long intervention, students in the RWL group presented the greatest gains. In two previous studies (Chang, 2011, 2012) of the effect of RWL on listening comprehension, the conflicting results were also explained in terms of the differences in the quantity of the input students were exposed to during the treatment: between 28 and 39 audio graded readers in Chang (2011), in which study students significantly improved their listening comprehension, versus 15 audio graded readers in Chang (2012) in which little improvement was found.

Nor have studies of the effects of RWL on reading comprehension yielded consistent results. In two early studies (Taguchi & Gorsuch, 2002; Taguchi, Takayasu-Maass, & Gorsuch, 2004) no significant differences were found in reading comprehension gains between an assisted RR group and a comparison group (a control group in the first study and an extensive reading group in the second). These two studies involving EFL college students were followed by a third similar study (Gorsuch & Taguchi, 2008), where comprehension was measured more accurately. This time the assisted RR group produced significantly higher levels of reading comprehension than the control group in the post-tests after an intervention of 11 weeks. The authors attribute these results to the Automaticity Theory according to which, as readers increase their word recognition skills, they can devote more attentional resources to comprehension. The benefits of exposure to simultaneous reading and listening for reading comprehension were also confirmed in Chang and Millett (2015), a study involving secondary school learners where the comparison was with a reading-only (RO) group. However, the gains were described as just 'acceptable' by the authors and lower than those in other studies (i.e., Beglar et al., 2012). It seems, then, that more visible differences in comprehension are obtained when comparisons are made with control groups (involving no reading or listening), whereas the differences are less prominent or consistent when the comparisons involve different types of input-based practices (such as RO, LO or extensive reading).

Reading comprehension has often been studied in conjunction with reading fluency in some of the abovementioned studies as well as in Chang (2012). In three of these studies (Chang, 2012; Gorsuch & Taguchi, 2008; Taguchi et al., 2004), the RWL intervention involved assisted RR (a variation of RWL which consists of RWL to a text repeatedly), while in Chang and Millett's study (2015) it did not. On the one hand, results from these studies seem to indicate that assisted RR is not as effective in developing reading fluency as another instructional practice called Timed Reading (TR), in which the reading is done under some degree of time pressure. In fact, in Chang (2012) the increased rate of the assisted RR group was approximately half that of the TR group, possibly because students in the TR group were aware that the reading task goal was to reduce reading speed, a pressure that the students in the RR group did not feel. On the other hand, another conclusion of these studies is that the effects of combining reading with listening seem to be mediated by the amount of input. This would explain why the improvement in reading fluency by the RWL group in Chang and Millett (2015) was twice that of the RO group after exposure to a considerable amount of input (a total of 115,412 words in 26 weeks); in contrast, the improvement in the assisted RR group was not superior to the control groups

in Gorsuch and Taguchi (2008) or in Taguchi et al. (2004), where treatment was shorter than Chang and Millet's treatment (11 and 17 weeks respectively) and the amount of input lower (16,963 words in Taguchi et al.).

Finally, some of the studies on RWL and assisted RR have also reported students' perceptions of these reading practices (Brown et al., 2008; Chang, 2009; Taguchi et al., 2004). Unlike the results for vocabulary, listening and reading comprehension and reading fluency, which may be sensitive to the features of the instructional program, the design of the study or the students' level of proficiency, students' opinions of a simultaneous rendering of oral and written text are always positive.

RWL in children

Reading-while-listening should be especially beneficial for children because of the nature of the reading materials addressed to this age group, which include both verbal (the written text) and visual input (i.e., the illustrations in graded readers and story books are a source of visual input). According to Paivio's (1986) dual-coding theory, which claims that the simultaneous processing of these two different input modes leads to higher learning gains, young language learners are expected to benefit from the multimodality provided in the reading materials and the simultaneous activation of the verbal and the imagery systems these materials trigger. Furthermore, RWL programs can also be beneficial for children with a low proficiency level in the L2, who tend to break the text into small incoherent parts (sometimes word by word); RWL may prevent this from happening.

However, research on the implementation of RWL programs with children is scarce. The work carried out so far includes some of the book flood studies reported by Elley (1991), a couple of literacy programs in the US (Blum et al., 1995) and a few studies in the field of multimedia learning (Huang, 2006; Nayak & Sylva, 2013). The publications that are most clearly comparable to the RWL studies with adults and teenagers cited in the preceding section are two comprehension-based programs implemented in Canada (Lightbown, 1992; Lightbown, Halter, White, & Horst, 2002) and in Spain (Tragant, Muñoz, & Spada, 2016). In both programs, young school-aged learners of English read and listened to texts of their choice (fiction and non-fiction) on a regular basis (daily in the Canadian study and twice a week in the Spanish one), and their performance was compared to that of students following regular teacher-led instruction. The children in the two reading programs read the texts (mostly storybooks and graded readers) quite independently and they spent as much time on English as their comparison groups.

The Canadian study evaluated a large-scale program in the 1980s that went on for several years, in which the participants were young Francophone learners of English. The experimental program involved children listening to a wide variety of English material while following the written text during daily 30-min periods. There was no oral practice or interaction during these periods and the teacher only provided organizational and technical support to students. The regular program involved children engaged in a variety of teacher-led listening and speaking activities like choral repetition, memorizing and practicing short dialogues and singing songs. In the

third year of the program evaluation, the authors found that on most of the measures participants in the experimental group performed as well as participants in the control group, and considerably better on the measures of receptive vocabulary. These findings were to a certain extent corroborated by the small-scale year-long study in Spain. After comparing several measures of general proficiency (dictation, listening and reading comprehension, written production and sentence imitation), the authors found that for the most part participants in the intervention group showed comparable but not superior levels of L2 development when compared to the group receiving teacher-led instruction only. In both programs, however, participants in the RWL programs showed more positive attitudes towards learning English than the comparison groups, despite having had less teacher-led instruction time.

Introduction to the study

The present work is a follow-up study of Tragant et al. (2016) with four distinguishing elements. In this study the presence of a RO group in addition to the RWL and control groups will help to identify any differential effect that RWL may have on L2 reading comprehension, fluency, vocabulary, listening comprehension and students' perceptions on the treatment. Secondly, all students read the same graded readers, which will allow us to evaluate the learning of vocabulary with a test based on the texts that students read in the intervention instead of a standardized vocabulary test. In the third place, besides measuring reading comprehension, reading fluency will also be evaluated, something that has been done in RWL research with older learners but not with children. Finally, in this study students read non-fiction (in Tragant, Muñoz and Spada three-fourths of the class library were fiction titles), which will allow us to see what perceptions students have of graded readers of this type.

The following research questions are addressed in the present study; the first three deal with linguistic outcomes and the fourth with non-linguistic ones.

1. To what extent does the reading program influence L2 reading and listening comprehension? Is there any effect of mode of input (RWL vs. RO)?
2. To what extent do primary school children in the program learn vocabulary semi-incidentally? Are there any differences between RO and RWL groups?
3. To what extent does the program influence L2 reading fluency and eye movements? Is there any effect of mode of input (RWL vs. RO) or L1 reading fluency?
4. What are students' perceptions of the task, and how engaged are they? Are there any differences between the RO and the RWL groups?

Methodology

Participants

The study took place in a school located outside the city centre of Barcelona which attracts families with a mixed socio-educational background, with 70% of the

mothers holding a university degree. The school was active in promoting English (which was the third language for all participants in the study) instruction as well as extensive reading in Catalan/Spanish during primary education.

The students were distributed across four intact Grade 5 classes (three classes participating as intervention groups and one as a comparison group). The four classes had seven periods of English exposure a week. In the intervention groups, students in two of the classes spent two of these periods (60 min each) engaged in reading/listening (the RWL intervention group). A third class spent the same amount of time on reading-only (the RO intervention group). The remaining class periods were devoted to regular teacher-led lessons (three periods of English instruction and two periods of science in English). In the comparison group, students were exclusively exposed to teacher-led lessons (five periods of English instruction and two periods of Science in English). All students received three additional periods of science instruction in Catalan.

There were 24–25 students (aged 10–11) in each of the four classes and the number of boys and girls in each class was fairly even (54% males, 46% females). All students were Catalan/Spanish bilinguals but not all of them spoke both languages at home. 42.7% of the children spoke Catalan at home, 32.3% spoken Spanish and 20.8% spoke both Catalan and Spanish. There were four students who spoke English at home together with Catalan and/or Spanish. Almost all of them (except for 1–3 in each class) reported reading books in their leisure time. In fact, children at this school used to have a book of their choice in their backpacks that they read whenever they had some spare time in class. The length of the books most students were reading during the intervention ranged from 100 to 570 pages (mean length = 315 pages). Most students' level of proficiency in English was around A1[1] (according to the CEFR 2001) and they were familiar with graded readers in English since they had read four of them as a whole-group activity the previous year.

The reading program

The intervention program took place between October and February,[2] after four introductory sessions to familiarize students with the materials and the procedure. The program ran over 18 reading/listening sessions that lasted 60 min each and were usually run 2 days a week. Except for three sessions (in which two short graded readers were read), students read one graded reader per session and a total of 21 graded readers in all. The books students read were from four different collections featuring science matters addressed to primary learners (Macmillan Science readers,

[1] A student at the A1 level "Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help".

[2] The intervention program continued after February till the end of the school year in June, but after February the sessions were restructured in order to increase the motivation of students in the RO group, and linguistic assessment was discontinued.

Macmillan children's books, Oxford Read and Discover, Benchmark Education). The titles (i.e., *Dangerous Weather, Recycling, Amazing minibeasts*, etc.) were broadly connected to topics learners would cover in their science classes that year. Their length ranged from 15 to 31 pages and on average they contained 909 words (12 min). The 21 books together included 14,535 words (a total of 4 h of audio track) and they contained less controlled input than the instructional materials that were used for regular instruction. With the aim of ensuring that learners would be able to read/listen semi-autonomously, the level of proficiency of the books was one or two stages lower than if the books had been used as a whole-class activity.

Every reading/listening session followed the same pattern. The first minutes of class time were devoted to the distribution of the books and the students' workbooks. In the two classes from the RWL group, a set of headphones and an MP3 were also distributed for each learner. Dictionaries were placed on desks so that they would be available to all students. Once students were ready, the teacher signalled them to start the session by briefly reviewing the vocabulary they had listed in their workbook from previous session(s). The next step would be to start with that day's book by first browsing through it. If there was a glossary, students were also expected to read it. Once they had a general idea of what the book was about, they started reading/listening to it for the first time. The two classes in the RWL group would turn on their MP3 in order to simultaneously read and listen to the book. The class in the RO group would start reading the book with no audio support. After that, students were asked to select eight words that they wanted to learn and write them down in their workbook together with a translation. They were told to use the dictionary or ask their classmates before resorting to the teacher for a translation. Then, students read/listened to the book for a second time from beginning to end or partially (depending on the length of the book). When the books were 20-min long or longer, there was no second reading/listening. After the second reading/listening, students were asked to write down a minimum of three questions about the contents of the book, choosing between true/false, multiple choice or wh-questions. These questions were later used as the basis for preparation of a class contest that took place after every 9th reading/listening session. When students had time left during a session, they were encouraged to show and/or ask their questions to the classmate sitting next to them. Finally, the last few minutes of class time were devoted to putting the materials away.

Design

A pre-post test (henceforth referred to as T1 and T2) design was followed to assess any linguistic changes and to record students' perceptions of the program. The program was also monitored during the intervention with classroom observations. The pre-tests were administered in September at the beginning of the school year (Time 1, T1) and before the start of the four training sessions for the intervention groups. The post tests were administered in February after the 18th sessions of the intervention program.

Table 1 Instruments

Instrument	Language of assessment	Time of administration
Reading comprehension	English and Catalan/Spanish	Oct. and Feb.
Dictation	English	Oct. and Feb.
Vocabulary test	English	Oct. and Feb.
L1 reading fluency	Catalan/Spanish	Oct.
L2 reading fluency	English	Oct. and Feb
Eye tracker	English	Oct. and Feb.

Instruments and procedure

Five instruments were used to assess students' linguistic outcomes: a reading comprehension test, a dictation, a vocabulary test, an L1 and an L2 reading fluency test, and an eye-tracker. The first three tests were administered in class, the fluency test was administered in the school's computer lab and the eye movement data were recorded in a quiet room. Non-linguistic outcomes were measured with a questionnaire which was administered in class. The language of assessment of each these instruments together with when they are administered is indicated in Table 1.

Reading comprehension was part of an institutional examination produced by the Catalan government. The test was based on two descriptive texts and it included 24 multiple choice items with three possible answers. Texts ranged from 200 to 275 words long. Two parallel tests were administered at T1 and T2. A dictation was used as an integrative measure of listening comprehension and it included a descriptive text that was pre-recorded into 12 segments (a total of 50 words). The text had been used in previous research (Muñoz, 2006) with late primary school students. The same dictation was used at T1 and T2. Vocabulary learning was measured with a bilingual matching test (Webb & Chang, 2015) which was created based on a selection of concrete nouns (n = 50) that appeared in the 21 graded readers students read from October to February and which students did not know at the beginning of the school year (as reported by their teachers). Students were presented with 10 blocks of five items each. The L1 meaning (in Catalan and Spanish) of the five target words and one distractor were provided in each block. The students' task consisted in matching the correct L1 word meaning with each target word.

Reading fluency was assessed through a computerized test and with eye movement data. In the computerized test, participants were asked to read two age-appropriate texts: one in their L1 (Catalan or Spanish) at T1 and one in their L2 both at T1 and T2. The L1 text was a narrative passage from a book titled 'El misterio de la Calle de las Glicinas' (Pradas, 2015). This book was chosen because it was published in both Catalan and Spanish and the translation was done by the author herself. The Catalan text contained 192 words, and the Spanish one comprised 185 words. The English texts used at T1 and T2 were a narrative passage that contained a two-line dialogue. The text used at T1 was taken from a book titled 'PB3 and the vegetables' (Cadwallader, 2010) and the one used at T2 was taken from a book titled

'PB3 and Coco the Clown' (Cadwallader, 2012), which belonged to the A1 level and comprised 220 and 208 words respectively. The English texts were below the students' level of proficiency, as previous research shows that a requirement for measuring reading rate is text suitability, according to which the text has to be well within the students' capability (Carver, 1990; Huffman, 2014; Rasinski, 2003). Their Flesh-Kincaid readability index was 91.7 for the text read at T1 and 85.1 for the text read at T2 (scores which indicate that they were appropriate). The three texts were computer delivered. They were previously piloted and a few small changes were made (for example, in the L1 text the word *peseta* was changed to *moneda*). In order to control for any task-order effects at T1, half of the participants in each group were asked to start reading the L1 text and the other half started reading the L2 text.

The test was conducted in the computer lab. Participants were asked to read the texts silently and at their normal pace. Participants first selected the text (Catalan/Spanish or English) and when they clicked on the 'Start' button, the text appeared and the chronometer started (although the chronometer was not visible to students and they were not told that their reading speed was being assessed). When they finished reading the text they clicked on the 'Finish' button and their reading speed score was obtained, which was automatically calculated through the formula words read per minute (number of words read * 60 s/number of seconds needed to read the text).

Finally, an eye-tracker was used to record eye movement data while students were reading silently. Two chapters from the same non-fiction graded reader (*Super Structures* (Undrill, 2015), a book that was not part of the reading program) were selected as stimuli for T1 (chapters 1–3: 354 words) and T2 (chapters 4–6: 343 words) and each chapter was presented on six different screens, which students could change upon pressing the return button to continue reading. A remote desktop eye-tracker (Tobii T120) was used to record the data on a one-on-one basis. Tobii T120 has a sampling rate of 120 Hz, which is considered adequate for the examination of fixations to larger regions of interest (Conklin and Pellicer-Sánchez, 2016). It has a typical accuracy of 0.5° and 0.2° resolution. Before starting the recording of participants' eye movements, the eye tracker was calibrated using a 5-point calibration grid. The stimuli were displayed on a 24″ screen using Tobii Pro Studio (version 3.4.2). This experimental session was conducted individually on the school premises.

Non-linguistic outcomes were measured through a questionnaire. Ten closed questions (most of them four-level Likert items) were used to record information on students' attitudes towards the reading sessions and their level of engagement. This questionnaire was administered to students in the RWL and RO groups. For a more thorough analysis of non-linguistic outcomes based on a combination of questionnaire and interview data see Tragant and Vallbona (2018).

Analysis

In order to examine linguistic gains, analyses were conducted without students who had learning difficulties (one student) or who studied extracurricular English (eight

students). Students who spoke English at home with some regularity (four students) were also excluded. In addition, if a student did not complete a test properly, he or she was excluded from the corresponding analysis. The final sample for the RWL group ranged from 37–40 students: 15–20 for the RO group and 14–20 for the control group, depending on the test. Analyses with eye movement data were conducted with a subsample of students (n = 35), after excluding six students due to the poor quality of their recordings.

The maximum score for the reading comprehension test was 24 (1 per item). The maximum score for the dictation was 50 points (one per word) and the exact-word scoring method was used. The maximum score for the vocabulary test was 50 (one per item). The score in the fluency test was produced automatically in words per minute. The measure used to analyse eye movement data was 'average fixation duration' (ms), which is a score of the average length of the pauses (fixations) made while reading. Before calculating the average fixation duration for each group, it was calculated for each page and averaged for each participant.

Because data were normally distributed in all the tests, parametric statistics were used for the analyses. A mixed ANOVA was used for the reading comprehension, dictation, vocabulary tests as well as the eye movement data, with time as the within-subject variable and condition as the between-subject factor. A repeated measures analysis of covariance was used to analyse reading fluency, with L1 fluency as a covariate.

When examining students' perceptions, no participants were excluded from analyses. Questionnaires were administered to whole classes (48 students from the RWL group and 24 from the RO group) and they were analysed descriptively.

Results

Linguistic outcomes

The means and standard deviations of the scores of all the tests and measures by group are presented in Table 2. Analysis of the reading comprehension test showed a statistically significant main effect for time [$F(1, 73) = 105.41$; $p = .00$, partial eta2 = .59] but no main effect for condition [$F(2, 73) = 1.61$; $p = .07$, partial eta2 = .01] and no interaction effect time*condition either [$F(2, 73) = 1.78$; $p = .18$, partial eta2 = .05]. In order words, the students in the three groups made significant progress in reading comprehension by T2 and the effect size was large but there were no differences in the effect of reading comprehension between the RWL, RO and control groups.

Similarly, analysis of the dictation showed a statistically significant main effect for time [$F(1, 69) = 111.66$; $p = .00$, partial eta2 = .62] but no main effect for condition [$F(2, 69) = 1.59$; $p = .21$, partial eta2 = .04] and no interaction effect time*condition [$F(2, 69) = 2.26$; $p = .11$, partial eta2 = .06]. In other words, students made significant progress in listening comprehension by T2 and the effect size was large, but there were no differences between the RWL, RO and control groups.

Table 2 Means and standard deviations in pre- and post-tests

	RWL		RO		Control group	
	T1 M (SD)	T2 M (SD)	T1 M (SD)	T2 M (SD)	T1 M (SD)	T2 M (SD)
Reading comprehension (max. 24)	15.43 (3.83)	18.88 (3.94)	14.94 (4.15)	18.83 (3.33)	14.22 (4.22)	19.44 (3.38)
Dictation (max. 50)	23.95 (10.87)	31.74 (10.89)	28.47 (9.06)	35.17 (8.12)	20.65 (9.96)	31.55 (8.79)
Vocabulary (max. 50)	28.52 (7.9)	39.21 (8.4)	26.50 (8.7)	38.25 (10.6)	28.73 (8.9)	32.89 (10.1
L2 reading fluency	92.34 (36.48)	124.57 (42.06)	95 (37.46)	113.25 (19.98)	73.64 (33.15)	89.67 (28.26)
Average fix. duration (ms)	242.51 (38.56)	225.83 (37.24)	250.71 (36.91)	224.70 (42.08)	222.53 (46.51)	200.27 (24.54)

In contrast to the results for dictation and reading comprehension, the results of the vocabulary test showed a main effect of time [F(1, 78) = 115.9; p = .000, partial eta2 = .60] and no main effect of condition [F(2, 78) = .862; p = .43; partial eta2 = .02], but a significant interaction between time and condition [F(2, 78) = 6.98; p = .01, partial eta2 = .15]. The relative gains from T1 to T2 for the RWL and RO groups were 21.38% and 19.5% respectively and 8.3% for the control group. ANOVAs and post-hoc tests with these gain scores for each of the groups suggest that there were significantly higher gains in the RWL and the RO groups than the control group [F(2, 78) = 6.98, p = .002] and no significant differences between the two intervention groups. See Serrano, Andriá and Pellicer-Sánchez (2016) for a detailed analysis of the vocabulary learning during the intervention.

With regard to L2 fluency, and given that previous research has shown that it may be significantly correlated with L1 fluency (Durgunoglu, Mir, & Ariño-Martí, 1993; Nassaji, 2014), the correlation between the two in the pre-test was checked and was found to be significant (r = .569**, p = .000, n = 65). Therefore, a Repeated Measures Analysis of Covariance was conducted with L2 fluency at T2 as the dependent variable, condition (RWL, RO or control) as the independent variable and L1 fluency as the covariate to see whether there were significant differences between groups controlling for L1 fluency. The results are in line with those found for reading comprehension and the dictation in so far as no main effect for condition was found [F(2, 57) = 1.962; p = .150, partial eta2 = .064]. However, a main effect for L1 fluency was found [F(1, 57) = 17.509; p = .000, partial eta2 = .235], indicating that it significantly predicted fluency in the L2 and explained 23% of the variance.

Finally, analysis of the average fixation duration showed a statistically significant main effect for time [F(1, 32) = 11.79; p = .00, partial eta2 = .27] but no main effect for condition [F(2, 32) = 1.92; p = .16, partial eta2 = .12] and no interaction effect time*condition [F(2, 32) = .20; p = .82, partial eta2 = .01]. In other words, on average students' fixations were shorter by T2, but there were no differences between the RWL, RO and control groups.

In sum, when the linguistic outcomes after participating in the reading/listening program were compared with those in the control group, no significant differences were found in comprehension (reading and listening), L2 reading fluency or eye movements. However, students in the RWL and the RO groups did obtain significantly higher scores than the control group for vocabulary learning and the two intervention groups (RWL and RO) obtained similar scores.

Non-linguistic outcomes

According to the questionnaires, students' level of engagement in the RWL group was high, in terms of both use of class time and attention during the reading/listening process. Hardly any students reported low levels of engagement (see Table 3). Their predisposition towards the post reading/listening activities was less homogeneous, especially with regard to the writing of the questions: 17% of the students reported not being motivated, though 33% reporting being highly motivated.

Table 3 Engagement during reading/listening sessions

	RWL (n = 48)					RO (n = 24)			
	Not much (%)	A little (%)	Quite (%)	A lot (%)	N/A	Not much (%)	A little (%)	Quite (%)	A lot (%)
Do you make efficient use of time?	0	19	50	31	–	12	33	37	17
Do you read/listen attentively?	4	17	33	46	–	4	42	50	4
Are you motivated to write the questions?	17	29	21	33	–	33	29	13	25
Are you motivated to write the vocabulary list?	17	11	40	30	–	29	33	25	13
Do you like reading/ listening to the books in English?	8	6	23	63	4%	54	13	25	4

As in the RWL group, few students in the RO group said they had not made good use of time or had not paid attention while reading. Overall, however, students in this group reported lower levels of engagement during the reading process and during the post-reading activities (see Table 3). For example, only 4% of the students in the RO group said they read/listened to the books with a lot of attention (the proportion in the RWL group was 46%) and only 13% said they wrote the vocabulary list with 'a lot' of motivation (the proportion in the RWL group was 30%). Observations also showed that some students in the RO group tended to spend less time reading than those in the RWL group, in which the pace of the reading was marked by that of the audio support.

Differences between the two groups became much more evident when students were asked about how much they liked reading/listening to the books (see Table 3). While 63% of the students in the RWL group said they liked it 'a lot', this figure fell to only 4% in the RO group.

Students were also asked about the graded readers and what they had learned from them. In general, their answers were shaped by the input modality they were exposed to, as can be seen from Table 4. While the students in the RWL group evaluated the books quite positively, the evaluation was quite divided in the RO group with 25% of the students saying they 'quite liked' them but 29% of them saying they 'did not like them much'. The difference between the two groups was also noticeable regarding the amount of English they felt they learned, with only 4% in the RWL group but 21% in the RO group feeling they had not learnt much. Students' perceptions about how much science they had learned was also somewhat higher in the RWL group, even though answers in this group were divided quite evenly, with similar proportions of students saying they 'learned a little', 'quite learned' and 'learned a lot'.

In sum, the examination of non-linguistic outcomes shows that students who participated in the RWL reported higher levels of engagement during the sessions and higher levels of satisfaction with the program, the reading materials, and the amount of learning.

Discussion

In the present study, young learners were exposed to 21 graded readers on science, and two different modes of input (RWO and RO) were compared. We aimed to report students' perceptions of the experience and also to observe language learning through the use of five instruments.

Students' perceptions of the RWL and the RO sessions were markedly different after their experience with the intervention from October to February. RWL was more popular among students than RO, in agreement with previous research by Tragant et al. (2016) who also found that students in the RWL group reported a very positive experience. There are several possible reasons for the popularity of RWL with young language learners: the appeal of technological devices, the privacy that the headphone confers and/or the preference for a dual mode of input (especially for students who are not fond of reading). The lower levels

Table 4 Students' perceptions of the books and their learning

	RWL (n = 48)					RO (n = 24)				
	N/A	Not much (%)	A little (%)	Quite (%)	A lot (%)	N/A (%)	Not much (%)	A little (%)	Quite (%)	A lot (%)
Did you like the books?	–	6	19	37	38	17	29	17	25	12
Are the sessions helping you learn about natural science?	2%	10	31	27	29	8	17	25	37	13
Are the sessions helping you learn English?	2%	4	17	56	21	4	21	46	4	25

of popularity of the RO program could in part be due to an awareness that they did not have access to the devices the RWL group were using. It could also be the case that the pattern of the sessions, including repeated reading and a set of post-reading activities (which was the same for the two intervention groups for the sake of comparison) was better suited to RWL than to RO. It would therefore be interesting to conduct further research with a different design in which the same group of young learners undergo a number of sessions with RO and then followed by a number of sessions with RWL.

In spite of the higher perceptions in the RWL group, students on the two intervention programs (RWL and RO) learned similar amounts of vocabulary in the books they had read. The relative gains of the two groups were similar to those reported among adolescent students on Webb and Chang's (2015) RWL program. While the relative gain in receptive vocabulary of this study was 21.38%, the gains reported by Webb and Chang ranged from 24 to 29%. The slightly higher percentage obtained in that study could be attributed to the age of the learners and their higher level of general cognitive maturity for learning vocabulary in a semi-incidental manner.

Age, together with the characteristics of the intervention, may also be an explanatory factor for the lack of significant differences between the two intervention groups and the control group in the rest of the linguistic outcomes reported in this study. The intervention took up only a relatively small fraction of the time the students spent learning English at school (two out of seven school periods a week) and this may have been insufficient to make a difference in the development of receptive skills in the three groups under comparison. With regard to age, the young learners on this program read/listened to a similar amount of graded readers (specifically, 21 titles) to that recorded in a comparable RWL program (20 titles) by Chang and Millett (2015), aimed at older students. However, and contrary to our findings, those authors found superior gains in comprehension and L2 fluency in the RWL group than in the control group. The fact that the graded readers used in this study were addressed to primary school children meant that they were much shorter (book length averaging 909 words per book, vs. 5770 words per book in the graded readers used in Chang and Millett). In view of this notable difference in book length, it is possible that the amount of input to which our students were exposed was insufficient to make a difference in how well they understood a text or how fast they could read it. Authors such as Beglar and Hunt (2014) have also pointed out that the benefits of extensive reading programs will only become visible if students read abundantly; in fact, they recommend 200,000 words/year in adults. The length of the books (ranging from 15 to 31 pages) read by our 10–11 year-old learners also seems to underestimate their reading capacity, if we take into account that these students were reading much longer books in their L1 for pleasure (see "Participants" section). Given that graded readers for primary school learners (no matter the publisher) are all similar in length, it seems reasonable to think that much longer readers for second language learners should be used if the aim of this type activity is to develop extensive reading.

Conclusion and limitations

In this study we compared a RWL group, a RO group and a control group over a period of 4 months. The results showed that primary school students' perceptions of RWL were much more positive than those reported for RO, even though the two groups were shown to learn similar amounts of vocabulary. No differences were found between the two intervention groups and the control group in other linguistic measures. This was probably because of the small amount of time devoted each week to the intervention, and also because graded readers for children may be too short and offer insufficient input for the development of reading and comprehension skills.

This study, however, has its limitations. One such limitation is the fact that only receptive measures were employed. It would be interesting to see if the participants in the different treatment groups would experience the same L2 development if productive measures were used. Finally, the study did not aim to assess how much scientific content students' actually learned from reading/listening to the books; this would have been a valuable complement to the results obtained.

Despite its limitations, the present study has contributed to shed light on the effects of two reading programs (RWL and RO) on L2 development with young learners. While participants in the two intervention groups experienced comparable vocabulary gains, those in the RWL group derived considerably more enjoyment from the program.

Funding Funding was provided by MINECO (Grant Nos. FFI2013-40952-P; FFI2016-80576-P), and AGAUR, Generalitat de Catalunya (Grant No. 2014SGR1089).

References

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language, 26*(1), 29–48.

Beglar, D., Hunt, A., & Kite, D. (2012). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning, 62*(3), 665–703.

Blum, I. H., Koskinen, P. S., Tennant, N., Parker, E. M., Straub, M., & Curry, C. (1995). Using audiotaped books to extend classroom literacy instruction into the homes of second-language learners. *Journal of Reading Behavior, 27,* 535–563.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136–163.

Cadwallader, J. (2010). *PB3 and the vegetables*. Recanati: ELI.

Cadwallader, J. (2012). *PB3 and Coco the clown*. Recanati: ELI.

Carver, R. P. (1990). *Reading rate: A review of research and theory*. New York: Academic.

Chang, A. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*(4), 652–663.

Chang, A. (2011). The effect of reading while listening to audiobooks: Listening fluency and vocabulary gain. *Asian Journal of English Language Teaching, 19,* 43–64.

Chang, A. (2012). Improving reading rate activities for EFL students: Timed reading and repeated oral reading. *Reading in a Foreign Language, 24*(1), 56–83.

Chang, A., & Millett, S. (2013). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language Journal, 25*(2), 126–148.

Chang, A., & Millett, S. (2014). The effect on extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal, 68*(1), 31–40.

Chang, A., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System, 52,* 91–102.

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*(3), 453–467.

Council, European. (2001). *Common European framework of reference for languages: Learning, teaching and assessment*. Cambridge: Cambridge University Press.

Durgunoglu, A. Y., Mir, M., & Ariño-Martí, S. (1993). The relationship between bilingual children's reading and writing in their two languages. In S. Randsdell & M. L. Barbier (Eds.), *Psycholinguistic approaches to understanding second language writing* (pp. 81–100). Dordrecht: Kluwer.

Elley, W. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 41*(3), 375–411.

Gorsuch, G., & Taguchi, E. (2008). Repeated reading for developing reading fluency and reading comprehension: The case of EFL learners in Vietnam. *System, 36*(2), 253–278.

Han, Z., & Chen, Ch A. (2010). Repeated-reading-based instructional strategy and vocabulary acquisition: A case study of a heritage speaker of Chinese. *Reading in a Foreign Language, 22*(2), 242–262.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11,* 207–223.

Huang, H.-L. (2006). The effects of storytelling on EFL young learners' reading comprehension and word recall. *English Teaching & Learning, 30*(3), 51–74.

Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language, 26*(2), 17–33.

Koskinen, P. S., Blum, I. H., Bisson, S. A., Phillips, S. M., Creamer, T. S., & Baker, T. K. (2000). Book access, shared reading, and audio models: The effects of supporting the literacy learning of linguistically diverse students in school and at home. *Journal of Educational Psychology, 92*(1), 23–36.

Lightbown, P. M. (1992). Can they do it themselves? A comprehension-based ESL course for young children. In R. Courchene, J. St John, C. Therrien, & J. Glidden (Eds.), *Comprehension-based second language teaching* (pp. 353–370). Ottawa: University of Ottawa Press.

Lightbown, P. M., Halter, R. H., White, J., & Horst, M. (2002). Comprehension-based learning: The limits of "do it yourself". *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 58*(3), 427–464.

Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.

Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching, 47,* 1–37.

Nayak, G., & Sylva, K. (2013). The effects of a guided reading intervention on reading comprehension: A study on young Chinese learners of English in Hong Kong. *The Language Learning Journal, 41*(1), 85–103.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9–29.

Serrano R., Andriá M., & Pellicer-Sánchez, A. (2016). Reading-while-listening in primary school: Linguistic and non-linguistic outcomes. In *Paper presented at the 'Multimodal input in second language learning' Symposium*, Barcelona, Spain.

Pradas, N. (2015). *El misterio de la calle de las Glicinas*. Barcelona: Bambú Editorial.

Rasinski, T. V. (2003). *The fluent reader: Oral reading strategies for building word recognition, fluency, and comprehension*. New York: Scholastic Inc.

Rasinski, T., & Hoffman, J. (2003). Oral reading in the school literacy curriculum. *Reading Research Quarterly, 38,* 510–548.

Taguchi, E., & Gorsuch, G. (2002). Transfer effects of repeated EFL reading on reading new passages: A preliminary investigation. *Reading in a Foreign Language, 14*(1), 43–65.

Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language, 16*(2), 70–96.

Tragant, E., Muñoz, C., & Spada, N. (2016). Maximizing young learners' input: An intervention program. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 72*(2), 234–257.

Tragant, E., & Vallbona, A. (2018). Reading while listening to learn: Young EFL learners' perceptions. *ELT Journal*. https://doi.org/10.1093/elt/ccy009.

Undrill, F. (2015). *Super structures (Oxford read and discover)*. Oxford: OUP.

Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Journal/La Revue canadienne des langues vivantes, 68*(3), 267–290.

Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667–686.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91–120.