

Extensive Reading Outcomes for Adult Emergent Bilinguals in a Community-Based ESL Program

Dung Nguyen
University of North Texas
USA

Ricardo González-Carriedo¹
University of North Texas
USA

Monica Ng
University of North Texas
USA

Abstract

This study investigated the effects of an Extensive Reading (ER) program on adult Emergent Bilinguals enrolled in a community-based ESL program in the United States. Using a quasi-experimental design, we examined pre- and post-test changes in reading, language, vocabulary, spelling, and language mechanics for a treatment group ($n = 53$) and a “business-as-usual” group ($n = 43$). Results from Wilcoxon signed-rank and paired-samples t-tests revealed significant gains for the treatment group in reading, language, language mechanics, and spelling, with moderate to large effect sizes. Vocabulary did not show significant improvement. These findings provide evidence that ER can support literacy-related language development among adult ESL learners outside formal academic settings. The study contributes to the ER literature by focusing on non-college adult ESL learners and highlights the value of ER for community-based programs.

Keywords: Extensive reading (ER), emergent bilinguals, adult ESL learners, non-college-attending adults, community-based ESL programs, literacy instruction, second language acquisition (SLA), mobile library, quasi-experimental design, Wilcoxon signed-rank test, paired samples t-test, Shapiro–Wilk test, adult basic education (ABE)

Adult English as a Second Language (ESL) learners in the United States represent a diverse and growing population with varied linguistic, cultural, educational, and socioeconomic backgrounds (National Center for Education Statistics [NCES], 2020). Whereas historically ESL learners have for the most part lived in the American southwest, today they are also present in communities

¹ Corresponding Author

across the United States which traditionally have been more linguistically homogeneous (Shiffman, 2019). These individuals, many of whom are immigrants or refugees, seek to acquire English language proficiency to improve their access to employment, education, civic participation, and overall integration into American society (Bigelow & Vinogradov, 2011; Grover et al., 2014). However, only a fraction of these adult individuals participate in formal educational programs to learn English, despite the substantial population with limited literacy and language proficiency needs (National Center for Education Statistics, 2023).

Given the diverse linguistic and educational backgrounds of adult ESL learners in the United States, there is a critical need to provide them with meaningful, research-based literacy programs that are tailored to their specific needs. Learning English as a second language in this context stands in contrast to learning English as a foreign language (EFL), i.e., English is learned in environments where it is not the dominant language. Effective adult ESL literacy instruction must go beyond basic decoding and comprehension to include culturally relevant content, critical thinking skills, and opportunities for authentic language use in real-world contexts (Vinogradov & Bigelow, 2010).

Research in the field of Teaching English for Speakers of Other Languages (TESOL) has emphasized the importance of integrating oral language development with reading and writing instruction, particularly for learners with limited formal schooling in their first language (August & Shanahan, 2006). Additionally, programs must be flexible and accessible, accounting for the complex lives of adult learners who often juggle employment, family, and other responsibilities. In consequence, instruction in adult ESL programs must be grounded in evidence-based practices and honor learners' cultural and linguistic resources. Programs with these characteristics are more likely to foster greater engagement, retention, and long-term success in both academic and social spheres (Condelli & Wrigley, 2004).

This article focuses on one particular program known as Extensive Reading (ER). ER is an instructional approach that aims to engage language learners in reading large quantities of books that are within their level of language proficiency (Webb & Chang, 2015). Through reading and follow-up activities, ER seeks to enhance the four language domains: speaking, listening, reading, and writing. The central point of this approach is to encourage learners to read on a voluntary basis by guiding them to choose high-interest story books at their reading level. Frequent reading, along with follow-up activities, can also provide an impetus for positive changes in reading habits and attitudes toward the target language.

To date, numerous meta-analyses have determined the effectiveness of ER in ESL and EFL contexts and across age groups in primary, secondary, and higher education (Anggia & Habók, 2025; Bolhan & Ismail, 2024; Jeon & Day, 2016; Krashen, 2007a; Nakanishi, 2015). While “the benefits of ER have been consistently validated across contexts and populations” (Hardy & Rodrigo, 2025, p. 14), relatively few studies have examined the effects of ER among immigrant populations in the United States (Setiyadi, 2020), adult learners outside traditional K–16 educational environments. This gap is especially notable in the case of non-college-attending adults participating in community-based literacy programs. The present study addresses this gap by examining the potential impact of ER on English language acquisition among adult ESL learners in the United States, who are not enrolled in higher education.

The purpose of this study was to investigate how participation in extensive reading (ER) activities might be associated with changes in English proficiency skills among adult ESL learners at a community-based program in the southwest of the United States. Consequently, the research questions were formulated as follows:

1. To what extent does participation in an Extensive Reading (ER) program contribute to the development of English language skills among adult learners not enrolled in higher education?
2. Which language proficiency component—reading, language, language mechanics, vocabulary, or spelling—shows the greatest improvement as a result of the implementation of the Extensive Reading program?

The following section will provide a synthesized account of the existing academic literature connected to ER.

Literature Review

The impact of ER on second language learning has been extensively examined in the context of EFL education. These studies have focused on different aspects of second language acquisition, such as the use of prepositions (Song & Sardegna, 2014), reading comprehension (Chang & Millett, 2015), motivation for speaking (Rahmany et al., 2013), writing (Lee & Hsu, 2009), the use of critical thinking skills (Husna, 2019), and attitudes toward reading (Mikami, 2017).

Research on ER, therefore, is ample. In fact, the Extensive Reading Foundation has compiled a bibliography of academic literature containing over 400 books, journal articles, and dissertations related to the use of ER in foreign language education (Jacobs et al., n.d.). The list comprises studies on the affective component of ER (Burrows, 2013; de Burgh-Hirabe & Feryok, 2013), grammar acquisition (Elley, 1991; Mason, 2006), teaching of languages other than English (Rodrigo et al., 2014; Tabata-Sandom, 2013), second language learning (Arnold, 2009; Poulshock, 2010) and reading ability (Bamford & Day, 1997), among other areas of study.

One of the most notable research studies in the field is a meta-analysis of 34 studies directly related to ER (Nakanishi, 2015). After a comprehensive analysis of these studies, the researcher found that “participants with extensive reading instruction yielded better outcomes than ones with other instruction” in terms of second language acquisition (p. 25). A second important finding was that the positive effects of ER increase with older participants, leading the author to affirm that “[e]xtensive reading might be more beneficial for late learners, who tend to learn language explicitly drawing on their analytical skills, than for early learners, who presumably still enjoy their ability to learn language in an implicit manner” (p. 26). It is important to note, however, that the majority of studies analyzed by Nakanishi focused on EFL contexts. In fact, only one was aimed at studying ER with ESL students. In the case of this study, the participants were college-level international students (Zimmerman, 1997).

Another characteristic of studies on ER is that most of them have focused on adult learners in the context of institutions of higher education. This was the case with a study where the positive effects of ER on vocabulary acquisition were shown on intermediate-level ESL students at a community college in California (Pitts et al., 1989). Fewer are the studies with a focus on adults not attending college. One such study examined a group of Vietnamese government officials who traveled to Singapore for a two-month intensive English course (Renandya et al., 1999). The authors found that “ER can indeed be beneficially implemented with second language learners beyond the age of university undergraduates, even in programs of relatively short duration, such as the two months of the EIC program [the name of the English course the participants received]” (p. 54).

In some cases, the effects of ER have been studied in mixed populations of first (L1) and second (L2) language learners. For example, Rodrigo and colleagues showed that using ER with adults who have difficulty reading is not only possible but may also yield positive results. The authors reported that the participants in the study, most of whom had never read an entire book, read an average of 17 books in the 14 weeks that the program lasted (Rodrigo et al., 2007). In another study, the authors reported an increased confidence in reading after the implementation of ER as well as important gains in fluency and expressive vocabulary. However, these gains were not significant in the areas of “receptive vocabulary, word identification, comprehension, or decoding” (Greenberg et al., 2006, p. 94). Using L2 learners exclusively, Cho and Krashen (1994) concluded that reading appropriate texts “can result in substantial vocabulary acquisition” (p. 667). In this case, the participants were not enrolled in an institution of higher education.

Although the primary outcomes associated with Extensive Reading (ER) programs are gains in vocabulary, reading fluency, and comprehension, repeated exposure to orthographic forms in meaningful contexts may also support incidental learning of spelling. Prior research suggests that while ER alone produces smaller and less consistent gains in spelling than in other domains, it can nonetheless contribute to orthographic awareness through frequent encounters with written words (Pan et al., 2015). In a similar way, ER may indirectly influence learners’ command of language mechanics—such as punctuation, capitalization, and grammatical conventions—by providing sustained exposure to authentic texts that model accurate written discourse. While empirical evidence in this area is more limited, researchers have argued that extensive engagement with well-formed texts offers learners opportunities to internalize structural features of written language (Day & Bamford, 2018a; Grabe, 2009).

Analysis of the academic literature revealed a possible gap in the existing research on ER as we could not locate many quantitative studies where the participants were non-college-attending adults in an ESL setting. The methodology section of this article describes the research design in which the authors attempt to address this gap. Furthermore, while the literature most strongly supports ER’s impact on reading, vocabulary, and language development, this study reports results across five language domains, including spelling and language mechanics, following the TABE assessment instrument used in the established, community-based English language program that was the setting for our study. While we looked at all five language domains per the TABE to align for consistency and completeness, we want to point out that the evidence gathered for spelling and mechanics is more limited and thus should be interpreted with caution.

Theoretical Framework

ER can be defined as “an approach to the teaching and learning of second language reading in which learners read large quantities of books and other materials that are well within their linguistic competence” (Day & Bamford, 2004, p. xiii). The theoretical assumption underlying the ER approach is that learners need large amounts of comprehensible input to gain proficiency in a language. The comprehensible input hypothesis (Krashen, 2007b) links the acquisition of a language to its comprehension. In the case of written language, texts need “to contain aspects of language that the acquirer has not yet acquired but is *developmentally* [emphasis in the original] ready to acquire” (Krashen, 2007b, p. 3).

ER is based on 10 principles: (1) The reading material is easy; (2) a variety of reading material on a wide range of topics must be available; (3) learners choose what they want to read; (4) learners read as much as possible; (5) the purpose of reading is usually related to pleasure, information and general understanding; (6) reading is its own reward; (7) reading speed is usually faster rather than slower; (8) reading is individual and silent; (9) teacher orient and guide their students; and (10) the teacher is a role model of a reader (Day & Bamford, 2002, pp. 137–140).

The underlying assumption behind the ER approach is that we learn to read by reading and this applies both to reading in our first language and reading in second and subsequent languages. EFL instruction based on the ER approach has been shown to help students increase their vocabulary, become better writers and enhance their oral fluency. Furthermore, ER serves as a stimulus to develop positive attitudes toward reading and learning a second language (Day & Bamford, 2018b). It is also important to note that ER is based on reading a large number of texts focusing on content meaning. This makes this approach different from intensive reading, where the focus is on the close study of fewer texts using dictionaries and grammars (Ewert, 2017).

Methodology

Research Context

Our study took place at a school district in a large metropolitan area in the United States and was based on the implementation of the ER program for a period of three months. The adult literacy program of this school district offered English Language Acquisition (ELA), Adult Basic Education (ABE), and Adult Secondary Education (ASE). Our study contributed to the ELA program, which was designed for non-native English-speaking adults who wanted to learn and improve their skills in speaking, reading, listening, and writing in English.

Prior to the implementation of the program, we offered an eight-hour professional development session for the three teachers who volunteered to facilitate our research study. The teachers were trained in how to implement the ER program and how to use an ER syllabus specifically designed for their classes. The eight-hour training focused on the theoretical foundations of ER and the implementation of the program.

Next, because the ER program requires learners to read as many books as they can, graded readers were purchased. The resulting mobile library consisted of 545 fiction and non-fiction graded readers that were designed specifically for English learners. These books came from different publishers specializing in reading materials for second language learners. Topics for the books were chosen based on the results of a survey that was administered to the students of the three teachers. These books were categorized in eight different reading levels, from *starter* to *level 7*. The idea of a mobile library stemmed from the fact that the three facilitating teachers taught their classes at different locations. This mobile library was circulated across the different locations where classes were taught so that all the students had an equal chance to access the books. Teachers were trained to guide students in the selection of books that were at or slightly beyond the students' reading competency level.

The eight classes taught by the three teachers were divided into either a “business-as-usual” group or a treatment group. The graded readers were distributed among the three teachers based on the results of the locator test administered to their students. This test will be described in the following section.

Each teacher spent 90 minutes per week implementing the ER program. During the first class, the teachers explained the purpose of the ER program to their students and described the expected learning outcomes. Since the ER program required students to read graded books from the classroom library, the teachers showed students how to choose books at their reading level by counting unknown words. This method requires students to count unknown words on a page (any page) of the book. If there are fewer than five unknown words in one page, that book is considered to be at their reading level (Day & Bamford, 1998).

Extensive Reading Program Design

Extensive reading (ER) in this study is conceptualized as an approach to the teaching and learning of second language reading in which learners read large quantities of materials that are well within their linguistic competence (Day & Bamford, 2004), guided by the ten principles articulated by Day and Bamford (2002). The ER program was designed to prioritize reading for pleasure and self-directed engagement, consistent with this approach, rather than quantitative tracking of reading behavior. Accordingly, students were not required to log reading time or the number of books completed. Although the amount of reading was not quantified—thereby limiting documentation of reading volume—the program emphasized alignment with the ER principles through learner choice, ease of materials, variety, and sustained exposure to comprehensible input. Participation was supported and monitored through structured follow-up activities (e.g., group discussions, pair-share summaries, and individual reflections), which provided opportunities for students to demonstrate engagement with their selected texts, while preserving the voluntary nature of extensive reading.

Application of the Ten Principles

To operationalize Day and Bamford's (2002) ten principles, a syllabus was developed based on activities from *Extensive Reading Activities for Teaching Language* (Bamford & Day, 2013). The syllabus outlined in-class tasks such as *Introducing Reading Material* and *Motivating and*

Supporting Reading, providing teachers with structured procedures for orienting students to extensive reading, sustaining motivation, and verifying participation while preserving the program’s voluntary nature. The complete syllabus is included in Appendix A. Table 1 summarizes how Day and Bamford’s (2002) ten principles were implemented in the program, encompassing teacher preparation, classroom practice, and independent reading components.

Table 1. *Application of Day and Bamford’s (2002) ten principles in the ER program*

Principle	Application in this program
1. The reading material is easy.	A mobile library of 545 graded readers (Starter–Level 7) allowed fluent reading without dictionary use; teachers used the “five-finger rule” to guide selection.
2. A variety of reading material on a wide range of topics must be available.	Fiction and nonfiction titles from multiple L2 publishers covered diverse genres and student-survey-identified themes.
3. Learners choose what they want to read.	Students freely selected level-appropriate books of interest, with teacher guidance as needed.
4. Learners read as much as possible.	Individual silent reading was completed at home, while 30 minutes of each two-hour class were devoted to ER activities (e.g., motivating and supporting reading, enhancing fluency and writing, and monitoring progress) to sustain engagement and reading volume.
5. The purpose of reading is usually related to pleasure, information, and general understanding.	In-class ER activities highlighted enjoyment, comprehension, and discussion of ideas, reinforcing reading for pleasure and information rather than assessment.
6. Reading is its own reward.	No grades or page targets were assigned; in-class ER activities such as sharing, reflection, and fluency practice emphasized the intrinsic value of reading.
7. Reading speed is usually faster rather than slower.	Students practiced fluent, continuous reading for meaning during home reading and fluency-oriented class activities.
8. Reading is individual and silent.	Individual silent reading occurred primarily outside class at students’ convenience.
9. Teachers orient and guide their students.	Three teachers (trained in an eight-hour professional development session) implemented the ER syllabus (see Appendix) and led 30-minute in-class ER activities to enhance fluency, motivation, and accountability while supervising students’ engagement with graded readers.
10. The teacher is a role model of a reader.	Teachers modeled positive reading habits, shared personal experiences, and applied insights from their ER training to foster a classroom reading culture.

The ER program was designed to integrate both in-class and at-home reading experiences in alignment with the core philosophy of extensive reading. Each of the three trained teachers devoted 30 minutes of every two-hour class to ER activities aimed at enhancing reading and oral fluency, writing development, motivation, and progress monitoring. Students complemented

these activities through individual silent reading at home, completed at their own pace. These practices ensured that the program was firmly grounded in Day and Bamford's (2002) framework.

Although the approach differed from the quantitative tracking recommended by Arai and Takizawa (2025), the study provided a clear rationale and transparent description of implementation. The program embodied Day and Bamford's (2002) principles by emphasizing choice, ease, variety, pleasure, and teacher modeling, supported by a structured ER syllabus that specified in-class activities for introducing and sustaining reading. Engagement was verified through meaningful classroom interactions, thereby avoiding the motivational drawbacks of mandatory reading logs and maintaining the intrinsic spirit of extensive reading.

Participants

For this study, a convenience sample of 96 students from the school district's adult education and literacy program were recruited. These students were enrolled in one of the eight ESL classes taught by the three teachers identified by the school district to participate in the ER program. Participation in the study was voluntary, and informed consent was obtained from each student who agreed to be part of the study.

The average age of the participants was 39 years with a range from 21 to 79. Two-thirds ($n = 64$) were female. All the participants were immigrants and the average length of stay in the United States was 8.5 years. Spanish was the primary language for 79% of the students with a total of 10 languages represented. The average age when students first studied English in a formal setting was 24 years. One third of the students reported having completed high school and another third, college. Almost one in five students (18%) reported not attending school beyond middle school.

Research Design

The purpose of the study was to quantify the effects of the implementation of the ER program on the English language learning of a specific English learner population. A quasi-experimental quantitative approach was chosen to answer the research questions. All students from the school district's Adult Education and Literacy program voluntarily participated in the research. We divided the participants ($n = 96$) into two groups—a "business-as-usual" group and a treatment group. These two groups differed at baseline due to the non-random assignment of the participants (assignment was based on student enrollment in one class or another). To avoid biased and uninterpretable statistical comparisons of the treatment effects, we chose to focus on within-group comparisons. In other words, the pre-existing differences and the potential for different rates of exposure to the English language outside of the classroom led us to think that conducting a direct comparative analysis between a control and a treatment group would violate key assumptions for causal or quasi-experimental inference and risk overstating the relationship between the intervention and observed changes (Shadish et al., 2002).

To establish baseline equivalence for contextual interpretation, we examined the demographic characteristics of the "business-as-usual" and treatment groups. The results indicated no meaningful differences between the two groups. For example, the mean age of participants in the

"business-as-usual" group was 40 years, compared to 38 years in the treatment group. In terms of gender distribution, the "business-as-usual" group consisted of 39% males and 61% females, while the treatment group included 20% males and 80% females. Similarly, participants in the "business-as-usual" group had resided in the United States for an average of 8.4 years, compared to 8.5 years in the treatment group. Additional variables, including level of education, home language, and age at which English learning began, also reflected comparable patterns across groups. These results suggest that the treatment and "business-as-usual" groups were broadly similar in their demographic profiles at baseline.

The "business-as-usual" group ($n = 43$) received instruction aligned with the methods and pedagogical approaches traditionally used in ESL classrooms. Instructional delivery prioritized the development of oral language skills, with particular emphasis on listening comprehension and speaking fluency. In contrast, comparatively less instructional time was devoted to the development of reading and writing skills. Teachers applied a blended instructional model that integrated elements of the communicative language teaching (CLT) approach with features of the grammar-translation method. The communicative approach emphasized functional language use, promoting interactive activities such as dialogues and role-playing. Simultaneously, the grammar-based component provided structured practice in grammatical forms and language rules.

The treatment group ($n = 53$) received instruction aligned with the principles and methods of the ER program, which was fully integrated into the official curriculum. Instructional practices emphasized sustained reading of comprehensible texts, with students receiving explicit guidance on how to select appropriate graded readers based on their language proficiency and personal interests. The ER approach prioritized reading for general understanding and enjoyment, encouraging students to engage with large volumes of text without the pressure of detailed analysis or translation. Students were encouraged to read regularly and independently, with an emphasis on learner autonomy and the development of intrinsic motivation. Teachers facilitated the program by offering support during the selection process and incorporating various ER-related activities such as letter writing and informal discussions. Formal grammar instruction was minimized, as language acquisition was expected to occur incidentally through repeated exposure to vocabulary and grammar structures in the books that students read. The treatment and "business-as-usual" groups spent a similar amount of time reading during class sessions.

The ER intervention was implemented at the classroom level, as entire classes were designated as either treatment or "business-as-usual" groups according to teacher participation and instructional scheduling. The analyses, however, were conducted at the student level to examine individual reading outcomes. Because classroom-level clustering was not modeled, the results should be interpreted with caution, as students within the same classroom may have shared contextual or instructional characteristics that could have influenced their performance.

Instruments

The TABE Locator was administered only as a placement tool to determine the appropriate level of testing for each participant, whereas the TABE 9 & 10 served as the primary assessments of

academic skills and the basis for evaluating learning outcomes in this study. We describe both in the following paragraphs.

TABE Locator Test. Initially, a locator test that focused on language and reading skills was administered to all participants. The purpose of this locator test was to determine the level of the Test of Basic Adult Education (TABE) the participants had to take as a pre-assessment measure. Locator tests are used to assess the students’ English language proficiency and to identify the correct version of the assessment for each student. In other words, locator tests permit students to take tests at their language proficiency level. The results of the locator test were used to choose among five different levels of the TABE test. Table 2 shows these levels and the content grade level range equivalent.

Table 2. *TABE test levels and grade level equivalents*

TABE test level	Content grade level range
Level L (Literacy)	0-1.9
Level E (Easy)	2.0-3.9
Level M (Medium)	4.0-5.9
Level D (Difficult)	6.0-8.9
Level A (Advanced)	9.0-12.9

Source: DRC/CTB (n.d.)

The locator test assessed participants in the areas of reading and language (also in Mathematics computation and Applied Mathematics, although these two parts were not administered). The reading section assesses critical thinking skills and understanding of texts in everyday life, school, and work. It covers phonics, phonological awareness, word recognition, and comprehension of both informational and literary texts. The language section focuses on language mechanics, vocabulary, and spelling skills (TABE, n.d.). Students whose lack of basic literacy skills in English prevented them from taking the locator test were assigned the TABE level L test.

TABE 9 & 10. The school district in which this study was conducted employed the TABE (Test of Adult Basic Education) to assess English learners. This instrument was selected because it evaluates general academic skills rather than focusing solely on English language proficiency. Some students enrolled in the adult ESL classes subsequently transitioned to ABE or GED programs and, ultimately, to technical training courses. To maintain consistency with the district’s established assessment practices and to minimize the testing burden on participants, the research team elected to use the TABE rather than introduce an additional assessment (e.g., TABE CLAS-E).

The TABE, one of the most widely used adult tests in the United States, assesses language skills in adult basic education. According to the publisher’s technical manual,

An Item Response Theory (IRT) model was employed in the selection and scaling of TABE 9 & 10 test items to provide the greatest statistical accuracy of test results. The test items were carefully reviewed by [...] editors and educational community professionals to minimize bias and ensure content appropriateness (DRC/CTB, n.d.).

This test has undergone extensive analysis for validity and reliability and has been normed and correlated to the General Educational Development (GED) test (Piccone, 2006). The test publisher, DRC/CTB, reports that the test is aligned with the national standards from the National Council of Teachers of English (NCTE) and the International Reading Association (IRA), among other prominent organizations in the area of education (DRC/CTB, n.d.). The TABE 9 & 10 comes in two different versions, a complete battery and a survey. The survey version of the TABE was chosen because it is designed with the same objective structure as the complete battery but takes less time to complete (DRC/CTB, n.d.). The difference between one and the other is that the complete battery provides norm- and criterion-referenced information. Since this was not needed for this study, the survey option was chosen.

The test is published in two forms, 9 and 10. Administering one as a pre-test and the other as a post-test helped eliminate any potential memory effect among participants. Each version of the TABE consists of two different tests. The first includes reading, mathematics, and language. The mathematics section was not administered. The second test includes language mechanics, vocabulary, and spelling. The reading sub-test focuses on the ability to read and comprehend real-life and work-related materials. The language sub-test covers skills typically found in jobs or other real-life settings. The language mechanics sub-test places attention on conventions of written language such as editing and proofreading. The vocabulary sub-test measures knowledge of word meaning. Finally, the spelling sub-test assesses the ability to form words from letters. Appendix B describes these sub-tests and their components.

The TABE and its accompanying locator test were administered at the beginning of the program and at the end with the purpose of determining the level of improvement in reading, language, language mechanics, vocabulary, and spelling by the participants in the ER program.

Demographic survey. Each participant responded to questions related to their age, gender, length of time in the United States, level of education attained, and first language. The survey also included questions relative to the English language development of the participants, such as the age when first exposed to English in an academic setting and the length and type of English language instruction received.

Control of Type I Error Across Multiple Tests

Because the study examined pre–post change across multiple linguistic domains, each statistical test addressed a distinct construct with its own theoretical rationale (e.g., reading comprehension, vocabulary, language mechanics). The analyses were therefore treated as domain-specific rather than as components of a single family of tests evaluating one overarching hypothesis. Consistent with the exploratory nature of the study, no familywise p -value adjustment (e.g., Holm–Bonferroni) was applied. However, because multiple tests were conducted on the same sample, readers are cautioned that the probability of Type I error may be inflated. Future confirmatory studies employing randomized designs may use familywise correction procedures or multivariate analytic approaches (e.g., MANOVA or multilevel modeling) to evaluate domain-level effects more rigorously.

Data Analysis

The results of the Locator Tests and TABE Tests from both the "business-as-usual" and treatment groups were used for analysis. These assessments were administered by instructors prior to the implementation of the ER program and again during the final month of the program. Our inferential testing focused on within-group pre–post change on two assessments—the Locator and the TABE—for each instructional condition: the standard English-language curriculum in the "business-as-usual" group and the same curriculum augmented with an Extensive Reading (ER) component in the treatment group.

To ensure appropriate model choice, we first evaluated the distributional assumptions of each pre- and post-test dataset using the Shapiro–Wilk test (Shapiro & Wilk, 1965), with $p < .05$ indicating a significant deviation from normality and examined Q–Q plots for visual confirmation. Based on these results, we used the Wilcoxon signed-rank test for all datasets and additionally reported paired-samples t -tests where the normality assumption was satisfied. The Holm–Bonferroni adjustment, or any familywise correction, is not necessary in this case because we are not making multiple independent inferences that would require controlling the overall chance of a false positive. In this study, the two paired t -tests do not represent independent hypotheses; rather, they serve as alternative analyses of the same two pre–post comparisons already tested by the Wilcoxon procedure. These two paired t -tests are simply robustness checks of the Wilcoxon results and do not create a new family of hypotheses.

Data analyses were initially conducted using an AI-assisted statistical tool to streamline processing and interpretation. To ensure the validity and reliability of these analyses, all results were subsequently cross-checked using the statistical software package SPSS (version 31). The outputs from both approaches were consistent, thereby confirming the accuracy and trustworthiness of the reported findings.

Table 3 presents the Shapiro–Wilk test statistics and corresponding p -values for each language measure in the "business-as-usual" group. Based on these results, datasets with $p < .05$ were considered to deviate significantly from normality. The Locator Reading Test was normally distributed in both pre- and post-tests; therefore, a paired-samples t -test is appropriate for this comparison. The Locator Language Test showed normality in the pre-test but not in the post-test, while all other measures (Reading, Language, Vocabulary, Language Mechanics, and Spelling) violated the normality assumption in at least one testing instance. Visual inspection of Q–Q plots confirmed substantial skewness and distributional irregularities, especially in Reading, Language, and Vocabulary, providing further justification for using non-parametric methods such as the Wilcoxon signed-rank test for these datasets. In conclusion, parametric tests are suitable only for the Locator Reading Test, while non-parametric tests are recommended for all other measures in the "business-as-usual" group due to violations of normality assumption.

Table 3. *The Shapiro–Wilk results for the “business-as-usual” group*

Measure	Pre-Test		Post-Test	
	<i>W</i>	<i>p</i>	<i>W</i>	<i>p</i>
Locator Reading Test	.97	.673	.95	.309
Locator Language Test	.92	.091	.89	.015
Reading	.59	< .001	.85	.027
Language	.60	< .001	.66	< .001
Vocabulary	.64	< .001	.57	< .001
Language Mechanics	.66	< .001	.66	< .001
Spelling	.86	< .001	.91	.053

Table 4 presents the Shapiro–Wilk test statistics and corresponding *p*-values for each language measure in the Treatment Group. Normality was evaluated based on these values, with $p < .05$ indicating a significant deviation from normality. The Locator Language Test scores were normally distributed in both pre- and post-tests; therefore, a paired-samples *t*-test is appropriate for analyzing this comparison. All other datasets (Locator Reading Test, Reading, Language, Vocabulary, Language Mechanics, and Spelling) showed significant or marginal deviation from normality in at least one testing instance. Q–Q plots further confirmed skewness and outliers in datasets such as Reading, Language, and Vocabulary, reinforcing the decision to apply non-parametric methods such as the Wilcoxon signed-rank test to those comparisons. In conclusion, non-parametric tests are recommended for all measures in the treatment group except the Locator Language Test, which satisfies the assumptions for parametric testing.

Table 4. *The Shapiro–Wilk results for the treatment group*

Measure	Pre-Test		Post-Test	
	<i>W</i>	<i>p</i>	<i>W</i>	<i>p</i>
Locator Reading Test	.89	.001	.94	.047
Locator Language Test	.96	.158	.95	.103
Reading	.87	.002	.91	.019
Language	.86	.002	.90	.016
Vocabulary	.89	.017	.92	.089
Language Mechanics	.87	.007	.92	.068
Spelling	.86	.006	.91	.054

Differences in language proficiency gains following the implementation of the program were primarily assessed using the Wilcoxon signed-rank test (Field, 2018; Sheskin, 2020), as most datasets did not satisfy the normality assumption required for parametric testing. This non-parametric method was applied consistently across the five language domains—reading, language, language mechanics, vocabulary, and spelling—to establish baseline equivalence for contextual interpretation of results. However, for datasets that did meet the normality assumption, including the Locator Reading Test ("business-as-usual" group) and the Locator Language Test (Treatment Group), paired-samples *t*-tests were also conducted.

This dual-testing approach allowed for greater statistical power where appropriate and served as a robustness check. In cases where both the Wilcoxon and *t*-tests yielded consistent results (e.g., both significant or both non-significant), the findings were considered more reliable. Reporting both test results also reflects a transparent and rigorous analytical process, ensuring that conclusions are both methodologically sound and reproducible. Due to variations in participants' attendance, sample sizes varied by language area. Only participants who completed both pre- and post-tests were included to maintain data integrity.

After these tests were performed, the effect size was calculated to determine whether an intervention was meaningfully impactful, which skills or domains benefited most, and how the magnitude of change compared across tests. Cohen's (1988) conventional thresholds were used to interpret the effect size. Table 5 shows the conventional thresholds for interpreting effect sizes (*r*).

Table 5. *Cohen's (1988) conventional thresholds*

Effect Size	Magnitude	Interpretation
0.10 – 0.29	Small	A real effect, but modest in practical terms
0.30 – 0.49	Medium	A moderate effect that's likely noticeable and meaningful
≥ 0.50	Large	A strong effect with clear practical significance

Results

The Effect of the ER Program on Different Language Areas

A series of Wilcoxon signed-rank tests were conducted separately for the treatment and "business-as-usual" groups across the five components of the TABE assessment —Reading, Language, Language Mechanics, Vocabulary, and Spelling— to examine within-group pre-post changes across multiple skill domains. The results are presented in Tables 6 and 7.

Table 6. *Wilcoxon signed-rank test results of "business-as-usual" group*

Test Type	<i>n</i>	Pre-Test		Post-Test		<i>W</i>	<i>z</i>	<i>p</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Locator Reading Test	22	5.77	3.12	6.00	2.96	84	-0.46	.322
Locator Language Test	22	3.09	2.20	3.55	2.20	63	-0.98	.164
Reading	13	16.38	17.66	20.2	20.94	26	-1.06	.145
Language	13	14.31	18.30	13.92	18.78	34	-0.43	.333
Vocabulary	12	11.75	14.30	8.00	12.66	20	-1.16	.124
Language Mechanics	12	12.75	13.90	7.00	13.20	10	-2.09	.018
Spelling	12	12.33	14.11	2.75	4.35	8	-2.47	.007

Table 7. *Wilcoxon signed-rank test results of treatment group*

Test Type	n	Pre-Test		Post-Test		W	z	p
		M	SD	M	SD			
Locator Reading Test	31	8.86	2.4	8.11	2.78	174	-1.47	.071
Locator Language Test	31	5.31	3.19	6.72	2.91	92	-3.06	.001
Reading	28	34.54	27.35	41.11	27.1	62	-2.7	.003
Language	27	31.96	27.98	38.3	29.36	90	-2.18	.014
Vocabulary	22	22.82	17.81	26.73	19.58	82	-0.9	.185
Language Mechanics	22	21.73	17.31	26.86	19.46	60	-2.16	.015
Spelling	20	19.1	15.33	27.25	20.68	34	-2.47	.007

Baseline results indicated that the treatment group began the study with higher proficiency than the "business-as-usual" group on several measures. For example, mean pre-test scores were higher on the Locator Reading Test ($M = 8.86$ vs. 5.77) and the Locator Language Test ($M = 5.31$ vs. 3.09). These disparities, stemming from non-random assignment, limit the validity of direct between-group comparisons. However, because the analyses focus on change scores within each group, the presence of different starting levels does not invalidate the tests of growth. Each participant serves as their own "business-as-usual", and while Matthew effects are theoretically relevant, the design isolates whether each group shows significant gains relative to its own baseline rather than whether one group outperforms the other.

For the treatment group, statistically significant improvements were observed across several language components following the intervention. Scores on the Locator Language Test showed significant gains ($z = -3.06$, $p = .001$), suggesting enhanced overall language proficiency, particularly in usage, sentence formation, paragraph development, and conventions of written English. Similarly, participants exhibited a significant improvement in Reading ($z = -2.70$, $p = .003$), a component that measures the ability to interpret graphic information, understand words in context, recall details, construct meaning, and evaluate or extend meaning. This result indicates that the intervention may have supported the development of both literal and inferential reading comprehension skills.

Significant progress was also noted in the Language component ($z = -2.18$, $p = .014$), reflecting increased accuracy in grammatical usage, sentence structure, and paragraph cohesion. Improvements in Language Mechanics ($z = -2.16$, $p = .015$) indicate better command of writing conventions such as capitalization and punctuation—skills critical for written communication. Furthermore, the treatment group showed statistically significant gains in Spelling ($z = -2.47$, $p = .007$), suggesting improved phonemic awareness and ability to encode words using standard spelling patterns, including mastery of vowels, consonants, and structural units of words. These results collectively point to the effectiveness of the intervention in strengthening foundational and applied language skills, with moderate to large practical significance.

However, no statistically significant differences were observed in the Locator Reading Test ($z = -1.47$, $p = .071$) or the Vocabulary component ($z = -0.90$, $p = .185$). The Vocabulary subtest focused on understanding word meanings, identifying multiple-meaning words, and using words

in context. The lack of significant changes in this area may suggest that vocabulary development requires more targeted or extended instructional input.

In contrast, the "business-as-usual" group did not demonstrate any statistically significant improvement across the five language components. Instead, significant declines were observed in Spelling ($z = -2.47, p = .007$) and Language Mechanics ($z = -2.09, p = .018$), indicating a deterioration in orthographic and writing convention skills over time. This decline may reflect a lack of exposure to explicit instruction or reinforcement in these areas. The absence of gains across the remaining components further supports the conclusion that the positive outcomes observed in the treatment group were attributable to the intervention rather than to maturation or external factors.

A paired samples *t*-test was conducted to compare the mean scores of the Locator Reading Test ("business-as-usual" group) and the Locator Language Test (Treatment Group), as both datasets met the normality assumption.

Table 8. Paired samples *t*-test results ("business-as-usual" group – Locator Reading Test)

Measure	Pre-Test		Post-Test		<i>t</i> (21)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Locator Reading Test	5.77	3.12	6.00	2.96	0.34	.734

The paired samples *t*-test showed no statistically significant difference between pre-test ($M = 5.77, SD = 3.12$) and post-test ($M = 6.00, SD = 2.96$) scores on the Locator Reading Test for the "business-as-usual" group. This suggests that the slight improvement observed was not statistically meaningful. The result is consistent with the outcome of the Wilcoxon signed-rank test, reinforcing the conclusion.

Table 9. Paired samples *t*-test results (treatment group – Locator Language Test)

Measure	Pre-Test		Post-Test		<i>t</i> (35)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Locator Language Test	5.31	3.19	6.72	2.91	3.50	.0013

The paired samples *t*-test revealed a statistically significant increase in scores from pre-test ($M = 5.31, SD = 3.19$) to post-test ($M = 6.72, SD = 2.91$) on the Locator Language Test for the treatment group. This finding indicates that participants demonstrated meaningful gains in language proficiency. The result is consistent with the outcome of the Wilcoxon signed-rank test, further supporting the conclusion.

Language Areas that Benefitted the Most

Multiple Wilcoxon signed-rank tests were conducted to assess changes in performance across seven language-related domains for both the treatment and "business-as-usual" groups. The number of paired observations (*n*), test statistics, significance levels, and effect sizes (*r*) are reported below and summarized in Table 10: Locator Reading Test, Locator Language Test,

Reading, Language, Vocabulary, Language Mechanics, and Spelling. Effect sizes (r) were calculated to assess the magnitude of change, using Cohen’s (1988) conventional thresholds.

Treatment Group. Statistically significant improvements were observed in the treatment group for multiple components. The Locator Language Test demonstrated a large effect size ($r = .55$), suggesting a strong impact of the intervention on overall language proficiency. Similarly, significant gains were found in Reading ($r = .51$) and Spelling ($r = .55$), both classified as large effects, indicating that the intervention substantially enhanced students’ abilities in reading comprehension and orthographic accuracy.

Language Mechanics ($r = .46$) and Language ($r = .42$) yielded medium effect sizes, reflecting moderate, yet meaningful improvements in grammatical conventions and syntactic structures. A small effect was noted for Vocabulary ($r = .19$), indicating some positive impact, although not statistically significant. The Locator Reading Test also showed a small effect ($r = .26$), suggesting modest practical gains in general reading comprehension, albeit below the threshold for statistical significance.

These findings suggest that the intervention was particularly effective in improving complex, integrated language skills such as reading comprehension, grammatical accuracy, and spelling.

Table 10. *Effect sizes and interpretations for Wilcoxon signed-rank tests*

Test Type	r	Magnitude
"Business-as-Usual" Group		
Locator Reading Test	0.1	Very Small
Locator Language Test	0.21	Small
Reading	0.29	Small
Language	0.12	Small
Vocabulary	0.33	Medium
Language Mechanics	0.6	Large
Spelling	0.71	Large
Treatment Group		
Locator Reading Test	0.26	Small
Locator Language Test	0.55	Large
Reading	0.51	Large
Language	0.42	Medium
Vocabulary	0.19	Small
Language Mechanics	0.46	Medium
Spelling	0.55	Large

“Business-as-usual” Group. In contrast, the “business-as-usual” group showed limited improvements and notable declines. The Spelling component demonstrated a large effect size ($r = .71$), but in the direction of performance decline, indicating a strong negative change. A similar pattern was observed in Language Mechanics ($r = .60$), also reflecting a large, adverse effect.

These results suggest that the absence of intervention may have contributed to regression in specific foundational skills.

Small effects were observed in Locator Language Test ($r = .21$), Reading ($r = .29$), and Language ($r = .12$), all indicating modest gains that were not statistically significant. Interestingly, Vocabulary showed a medium effect size ($r = .33$), suggesting a moderate change, though interpretation is limited by non-significance. The “business-as-usual” group’s outcomes highlight a concerning trend in which lack of targeted instruction not only failed to produce meaningful gains but was associated with significant deterioration in critical language subskills.

Discussion

The purpose of this study was to explore the potential of a language learning approach known as ER to support adult non-college ESL learners to improve their English language skills. Under this approach, learners read large amounts of texts that are within their language proficiency level with a focus on enjoyment.

A review of the academic literature revealed many studies had been conducted in ESL and EFL contexts and across age groups in formal educational settings. However, no studies found looked at the impact or effectiveness of ER among adult ESL learners in community-based programs. Consequently, our study focused on adult ESL learners outside of college-level programs. This study offers evidence that ER is effective for adult learners enrolled in community-based ESL programs. In this section we will interpret and explain the significance of the findings in relation to the research questions, existing academic literature, and the theoretical framework.

The study acknowledges the baseline proficiency gap, with the treatment group demonstrating higher pre-test scores on several measures. To avoid misinterpretation, the results should not be taken as evidence of treatment superiority but rather as evidence of within-group growth over time. The analyses were explicitly designed to evaluate gains relative to each group’s own baseline, not to make direct between-group comparisons. Any implications of comparative advantage must therefore be interpreted with caution. Future research should employ random assignment or pre-test-adjusted models, such as ANCOVA, to more rigorously examine group differences and address the potential for Matthew effects.

Research Question 1: To what extent does participation in an Extensive Reading (ER) program contribute to the development of English language skills among adult learners not enrolled in higher education?

Participants in the treatment group demonstrated statistically significant gains in several language domains, most notably in Locator Language Test ($z = -3.06, p = .001$), Reading ($z = -2.70, p = .003$), Language ($z = -2.18, p = .014$), Spelling ($z = -2.47, p = .007$), and Language Mechanics ($z = -2.16, p = .015$). These components correspond to increased proficiency in grammatical usage, syntactic construction, spelling accuracy, and reading comprehension. These results align with previous research suggesting that sustained exposure to meaningful, level-appropriate texts can foster language development through increased input and learner autonomy

(Day & Bamford, 2002; Krashen, 2007b). The observed gains reinforce the utility of ER as a viable instructional approach for adult ESL learners who are not engaged in formal higher education settings.

Interestingly, the Locator Reading Test did not yield statistically significant gains ($z = -1.47$, $p = .071$), although a positive trend was observed in the mean scores. This may suggest that while ER supports targeted reading subskills, broader general reading assessments may require longer periods of engagement or complementary instruction to produce measurable gains.

The Wilcoxon test results also revealed moderate to strong effects across components that showed statistical significance, indicating that the observed changes were not only statistically meaningful but also practically relevant. These findings support the argument that ER fosters literacy development among adult learners, with strength in areas such as syntactic awareness, orthographic accuracy, and decoding fluency.

Contrary to expectations, the Vocabulary component did not demonstrate statistically significant improvement ($z = -0.90$, $p = .185$), despite a modest increase in mean scores. This result diverges from studies emphasizing vocabulary acquisition as a core benefit of ER (e.g., Cho & Krashen, 1994). The discrepancy may be attributable to the relatively short duration of the program or the absence of explicit vocabulary instruction within the ER framework. Future research could explore whether supplementing ER with vocabulary-focused activities enhances lexical development more robustly.

An additional factor worth considering is the role of affective variables. Although the "business-as-usual" group did not show significant improvement in any domain, the treatment group's overall positive performance may also reflect increased learner motivation or confidence associated with the self-directed nature of ER. Although these affective factors were not formally measured, they merit further investigation given their potential contribution to language learning outcomes, particularly for adult learners operating outside traditional academic structures.

Research Question 2: Which language proficiency components—reading, language, language mechanics, vocabulary, or spelling—show the greatest improvement as a result of the implementation of the Extensive Reading program?

The findings of this study imply that the most substantial gains resulting from the ER program were observed in literacy-based domains—specifically Locator Language, Reading, and Spelling—with large effect sizes ($r = .51-.55$). These results are consistent with longstanding claims in the field of second language acquisition that ER enhances learners' literacy and overall language proficiency by providing repeated, meaningful exposure to language in context (Krashen, 1989; Day & Bamford, 2002).

Krashen's Input Hypothesis (1982, 1989) posits that language acquisition occurs most effectively when learners are exposed to comprehensible input that is slightly above their current proficiency level ($i+1$). The ER program applies this theory by offering level-appropriate texts that allow learners to engage with authentic language in low-anxiety, self-paced environments. In this study, repeated interaction with written texts strengthened students' decoding, grammatical

processing, and spelling skills—as seen in the gains in Language Mechanics ($r = .46$)—all of which are foundational to reading fluency.

Other studies have reported similar domain-specific effects. For example, Cho and Krashen (1994) and Yamashita (2008) found that ER contributed importantly to reading comprehension and fluency but had more limited effects on oral proficiency and vocabulary development unless paired with additional instructional supports. Similarly, Renandya and Jacobs (2016) argued that while ER strengthens input-based language skills, productive language domains (e.g., speaking and writing) benefit more from integrated approaches that combine reading with output-oriented activities.

The relatively limited impact of ER on vocabulary in this study—($r = .19$) classified as a small effect and not statistically significant—aligns with the understanding that vocabulary acquisition through incidental reading tends to be gradual and cumulative (Nation, 2001). This suggests that for adult ESL learners, especially those with limited exposure to English outside the classroom, ER may need to be supplemented with explicit vocabulary instruction to accelerate lexical gains (Laufer, 2005).

Interestingly, the Language component yielded a medium effect size ($r = .42$), indicating a moderate but meaningful improvement in learners' grasp of syntax, usage, and sentence structure, which are critical for both comprehension and writing accuracy. In contrast, although the Locator Reading Test and Vocabulary components showed small effects ($r = .26$ and $r = .19$, respectively), these did not reach statistical significance, suggesting more modest practical gains in general reading comprehension and incidental word learning.

Conclusion

This study examined the effects of an ER program on the English language development of adult learners not enrolled in higher education. The results indicate that ER can be a highly effective instructional approach for promoting multiple literacy-related language skills, particularly in the areas of Locator Language, Reading, Language, Spelling, and Language Mechanics. Learners in the treatment group demonstrated statistically significant improvements in these domains, with moderate to large effect sizes, while the "business-as-usual" group showed no statistically significant gains and, in fact, experienced declines in key areas such as Spelling and Language Mechanics.

These findings contribute to the growing body of research supporting the use of ER in second language acquisition, extending its applicability to underserved adult populations outside traditional academic environments. By offering accessible, self-selected reading materials within a supportive and flexible instructional framework, the ER program aligned with established theories in the field of second language acquisition, particularly the Input Hypothesis, and provided a meaningful avenue for language growth. However, not all domains benefited equally. The absence of statistically significant improvement in Vocabulary and Locator Reading suggests that ER may need to be paired with complementary strategies—such as explicit

vocabulary instruction or structured reading comprehension tasks—to address these areas more effectively.

In addition to its empirical contributions, the study offers practical implications for adult education programs seeking to implement learner-centered approaches to language instruction. The success of the mobile library model and the training provided to instructors underscore the importance of intentional program design that respects learners' needs and backgrounds. Moreover, the findings support the integration of ER into adult literacy initiatives not only as a means to promote academic growth but also to mitigate skill loss in the absence of formal education.

Future research should explore the long-term effects of ER on adult learners' academic and professional trajectories, the role of affective variables such as motivation and self-efficacy, and the integration of ER with other pedagogical approaches to support comprehensive language development. Also, future studies should investigate how demographic characteristics, such as age, gender, and educational background, may interact with instructional interventions like an ER program. Examining these interactions could provide a deeper understanding of educational outcomes among ESL learners and inform more targeted instructional strategies.

In sum, this study affirms that ER is not only feasible but impactful in community-based ESL contexts, and that it holds significant promise for fostering equitable language learning opportunities for adult Emergent Bilinguals.

Limitations

Several limitations should be acknowledged in the design, implementation, and interpretation of this study. First, the relatively small sample size—particularly in the “business-as-usual” group—may have limited statistical power and reduced the generalizability of the findings. In addition, the duration of the intervention may not have been sufficient for robust gains to emerge in domains such as vocabulary development or oral language proficiency, which typically require sustained exposure over longer instructional periods.

Second, the study did not employ random assignments to instructional conditions. As a result, preexisting differences between groups—most notably higher baseline English proficiency in the ER condition—may have influenced learning trajectories. These differences raise the possibility of Matthew effects, whereby learners with stronger initial proficiency are better positioned to benefit from instruction as well as from opportunities for language exposure beyond the classroom.

Third, although extensive reading (ER) was conceptualized and implemented in this study as an instructional approach grounded in Day and Bamford's (2002, 2004) ten principles, the amount of reading completed by participants was not quantified. Reading volume (e.g., time spent reading, number of pages or books read) was not systematically measured, which limits documentation of the extent to which learners engaged in “large quantities” of reading as described in canonical ER definitions. Consequently, the study cannot examine dose–response

relationships or verify the amount of out-of-class reading completed by individual participants. While structured in-class activities supported engagement with reading materials, these procedures cannot fully rule out incomplete reading outside class.

Fourth, the study was conducted in an ESL context in which participants were likely exposed to varying amounts of English outside the instructional setting (Nation & Waring, 2020). Because out-of-class language use was not measured, it is not possible to determine the extent to which gains observed in the ER condition were influenced by external input or practice. In addition, participation was based on convenience sampling in a voluntary, non-college adult education program. Participants may therefore have been especially motivated to improve their English proficiency and may have sought additional exposure beyond the instructional activities provided, further complicating efforts to isolate instructional effects.

Fifth, the assessment framework focused primarily on literacy-based outcomes (e.g., reading, spelling, and language mechanics). As a result, potential gains in speaking and listening were not directly assessed, yielding an incomplete picture of overall language development. In addition, although the TABE assessments are widely used and supported by publisher documentation, there is limited independent, peer-reviewed research establishing their psychometric properties.

Finally, although participating teachers received professional development prior to implementation, variation in instructional practices across classrooms may have affected implementation fidelity. Moreover, while the ER intervention was assigned at the classroom level, statistical analyses were conducted at the student level without adjustment for potential classroom clustering effects. This approach assumes independence of observations, which may not fully be held in shared instructional contexts. Future studies employing multilevel or hierarchical modeling would allow for more precise estimation of classroom- and teacher-level variance.

Taken together, these limitations underscore the need for caution in interpreting the findings. While the results document within-group growth under an ER approach defined by established principles, they do not permit causal attribution of observed gains to the ER intervention alone. Future research using randomized designs, pre-test-adjusted analytic models, systematic indicators of reading volume, direct measures of out-of-class exposure, and multilevel statistical approaches are needed to more rigorously evaluate the effects of extensive reading programs.

Declaration of AI Use in the Writing Process

During the writing of this work the authors used generative AI or AI-assisted technologies for the purpose of conducting some statistical analyses. The authors take responsibility for the content and intended meaning of this article.

References

Anggia, H., & Habók, A. (2025). The efficacy of online extensive reading among university students and the relationship between affective variables and English reading

- comprehension. *Scientific Reports*, 15, Article 8373. <https://doi.org/10.1038/s41598-025-92326-9>
- Arai, Y., & Takizawa, K. (2025). Day and Bamford's (2002) ten principles for teaching extensive reading revisited: A methodological synthesis of research practice. *Reading in a Foreign Language*, 37(1), 1–23.
- Arnold, N. (2009). Online extensive reading for advanced foreign language learners: An evaluation study. *Foreign Language Annals*, 42(2), 340–366. <https://doi.org/10.1111/j.1944-9720.2009.01024.x>
- August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth*. Lawrence Erlbaum Associates.
- Bamford, J., & Day, R. R. (1997). Extensive reading: What is it? Why bother? *The Language Teacher*, 21(5), 6–8.
- Bamford, J., & Day, R. R. (2013). *Extensive reading activities for teaching language*. Cambridge University Press.
- Bigelow, M., & Vinogradov, P. (2011). Teaching adult second language learners who are emergent readers. *Annual Review of Applied Linguistics*, 31, 120–136. <https://doi.org/10.1017/S0267190511000109>
- Bolhan, N. S., & Ismail, H. (2024). Extensive reading and solving reading difficulties among English language learners. *International Journal of Academic Research in Business and Social Sciences*, 14(12), 4559–4578. <http://doi.org/10.6007/IJARBS/v14-i12/24424>
- Burrows, L. (2013). The effects of extensive reading and reading strategies on reading self-efficacy. *Studies in Linguistics and Language Teaching*, 24, 129–156.
- Chang, A. C. S., & Millett, S. (2015). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language*, 27(1), 126–148.
- Cho, K.-S., & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, 37(8), 662–667.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Condelli, L., & Wrigley, H. S. (2004). Real-world research: Combining qualitative and quantitative research for adult ESL. In D. W. Clements & G. M. White (Eds.), *New Directions for Adult and Continuing Education* (Vol. 2004, No. 103, pp. 73–82). Jossey-Bass.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge University Press.
- Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, 14(2), 136–141.
- Day, R. R., & Bamford, J. (2004). *Extensive reading activities for teaching language*. Cambridge University Press.
- Day, R. R., & Bamford, J. (2018a). *Extensive reading in the second language classroom* (2nd ed.). Cambridge University Press.
- Day, R. R., & Bamford, J. (2018b). Extensive reading. In J. I. Liantas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–7). John Wiley & Sons. <https://doi.org/10.1002/9781118784235.eelt0472>
- de Burgh-Hirabe, R., & Feryok, A. (2013). A model of motivation for extensive reading in Japanese as a foreign language. *Reading in a Foreign Language*, 25(1), 72–93.

- DRC/CTB. (n.d.). *TABE 9 & 10 Test Overview*. Data Recognition Corporation.
<https://tabetest.com>
- Elley, W. B. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning*, 41(3), 375–411. <https://doi.org/10.1111/j.1467-1770.1991.tb00611.x>
- Ewert, D. (2017). Getting ER into the curriculum: No more excuses! *CATESOL Journal*, 29(2), 5–20. <https://files.eric.ed.gov/fulltext/EJ1164371.pdf>
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage Publications.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Greenberg, D., Rodrigo, V., Berry, A., Brinck, T., & Joseph, H. (2006). Implementation of an extensive reading program with adult learners. *Adult Basic Education*, 16(2), 81–97.
- Grover, K. S., Miller, M. T., Swearingen, B., & Wood, N. (2014). An examination of the self-directed learning practices of ESL adult language learners. *Journal of Adult Education*, 43(2), 12–19.
- Hardy, J. E., & Rodrigo, V. (2025). Incorporating extensive reading into the foreign language curriculum. *Dimensions*, 60, 10–27.
- Husna, N. (2019). Developing students' critical thinking through extensive reading. *TEFLIN Journal*, 30(2), 212–230. <https://doi.org/10.15639/teflinjournal.v30i2/212-230>
- Jacobs, G. M., Renandya, W. A., & Bamford, J. (n.d.). *Annotated bibliography of works on extensive reading in a second language*. Extensive Reading Foundation.
<https://bib.erfoundation.org>
- Jeon, E-Y., & Day, R. R. (2016). The effectiveness of ER on reading proficiency: A meta-analysis. *Reading in a Foreign Language*, 28(2), 246–265.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440–464.
<https://doi.org/10.2307/326879>
- Krashen, S. D. (2007a). Extensive reading in English as a foreign language by adolescents and young adults: A meta-analysis. *The International Journal of Foreign Language Teaching*, 3(2), 23–29.
- Krashen, S. D. (2007b). *Explorations in language acquisition and use*. Heinemann.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5(1), 223–250. <https://doi.org/10.1075/eurosla.5.11lau>
- Lee, S.-Y., & Hsu, Y.-Y. (2009). Determining the crucial characteristics of extensive reading programs: The impact of extensive reading on EFL writing. *International Journal of Foreign Language Teaching*, 5(1), 12–20.
- Mason, B. (2006). Free voluntary reading and autonomy in second language acquisition: Improving TOEFL scores from reading alone. *International Journal of Foreign Language Teaching*, 2(1), 2–5.
- Mikami, A. (2017). Students' attitudes toward extensive reading in the Japanese EFL context. *TESOL Journal*, 8(2), 471–488. <https://doi.org/10.1002/tesj.283>
- Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly*, 49(1), 6–37. <https://doi.org/10.1002/tesq.157>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

- Nation, I. S. P. & Waring, R. (2020). *Teaching extensive reading in another language*. Routledge.
- National Center for Education Statistics. (2020). *Adult English language learners in the United States*. U.S. Department of Education. <https://nces.ed.gov>
- Pan, S. C., Rubin, B. R., & Rickard, T. C. (2015). Does testing with feedback improve adult spelling skills relative to copying and reading? *Journal of Experimental Psychology: Applied*, 21(4), 356–369. <https://doi.org/10.1037/xap0000062>
- Piccone, J. E. (2006). Administering the Test of Adult Basic Education at intake: A biased marker of offender ability. *Journal of Correctional Education*, 57(3), 239–248.
- Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*, 5(2), 271–275.
- Poulshock, J. (2010). Extensive graded reading in the liberal arts and sciences. *Reading in a Foreign Language*, 22(2), 304–322.
- Rahmany, R., Zarei, A. A., & Gilak, S. (2013). The effect of extensive reading on Iranian EFL learners' motivation for speaking. *Journal of Language Teaching and Research*, 4(6), 1238–1246. <https://doi.org/10.4304/jltr.4.6.1238-1246>
- Renandya, W. A., Rajan, B. R. S., & Jacobs, G. M. (1999). Extensive reading with adult learners of English as a second language. *RELC Journal*, 30(1), 39–61. <https://doi.org/10.1177/003368829903000103>
- Renandya, W. A., & Jacobs, G. M. (2016). Extensive reading and listening in the L2 classroom. In W. A. Renandya & H. P. Widodo (Eds.), *English language teaching today: Linking theory and practice* (pp. 97–110). Springer. https://doi.org/10.1007/978-3-319-38834-2_8
- Rodrigo, V., Greenberg, D., & Segal, D. (2014). Changes in reading habits by low literate adults through extensive reading. *Reading in a Foreign Language*, 16(1), 20–34.
- Rodrigo, V., Greenberg, D., Burke, V., Hall, R., Berry, A., Brinck, T., Joseph, H., & Oby, M. (2007). Implementing an extensive reading program and library for adult literacy learners. *Reading in a Foreign Language*, 19(2), 106–119.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Setiyadi, A. G. (2020). *Teaching English as a foreign language*. Graha Ilmu.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>
- Sheskin, D. J. (2020). *Handbook of parametric and nonparametric statistical procedures* (6th ed.). CRC Press.
- Shiffman, C. D. (2019). Supporting immigrant families and rural schools: The boundary-spanning possibilities of an adult ESL program. *Educational Administration Quarterly*, 55(4), 537–570. <https://doi.org/10.1177/0013161X18809344>
- Song, J., & Sardegna, V. G. (2014). EFL learners' incidental vocabulary acquisition from extensive reading. *Reading in a Foreign Language*, 26(2), 182–205.
- Tabata-Sandom, M. (2013). The reader-text-writer interaction: L2 Japanese learners' responses toward graded readers. *Reading in a Foreign Language*, 25(2), 264–282.
- TABE (n.d.). *TABE suite of assessments*. <https://tabetest.com>
- Uvin, J., Kallenbach, S., Crowe, A., González, C., Singh, N., Vanek, J., & Webber, A. A. (2020). *The transformation of adult ESOL learning: A practice and policy brief*. *World Education*. <https://edtech.worlded.org/wp-content/uploads/2020/12/esol-ex-summary.pdf>

Vinogradov, P., & Bigelow, M. (2010). *Using oral language skills to build on the emerging literacy of adult English learners (CAELA Network Brief)*. Center for Applied Linguistics. <https://www.cal.org/caelanetwork/pdfs/using-oral-language-skills.pdf>

Webb, S., & Chang, A. C. S. (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. <https://doi.org/10.1177/1362168814559800>

Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, 36(4), 661–672. <https://doi.org/10.1016/j.system.2008.04.003>

Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 5–19). Cambridge University Press.

Appendix A

Syllabus for the Extensive Reading Program

Teachers will devote 90 minutes of class time each week to Extensive Reading (ER) activities to help students build fluency, confidence, and enjoyment in reading. They will select and adapt the suggested activities in the table below to match their students’ interests, proficiency levels, and classroom goals, ensuring that every session supports meaningful and engaging reading experiences.

CLASS MEETINGS	ACTIVITIES
Week 1	<p><u>Getting started</u></p> <ul style="list-style-type: none"> - Reading Record Form by Ken Schmidt (you are encouraged to do this at the beginning of the program) - Reading Notebook by Midori T.Jwano (you are encouraged to do this at the beginning of the program) - Vocabulary Journal by Richard R. Day (you are encouraged to do this at the beginning of the program) - Exploring Personal Reading Histories by Anne Burns - Exploring Reading by Andy Barfield - Getting Acquainted by Maureen Snow Andrade - What Do You Think? By Marc Helgesen
Week 2	<p>Spring Break - No Class</p>
Week 3	<p><u>Introducing Reading Material</u></p> <ul style="list-style-type: none"> - Identify the Books by Yamanaka Junko - Blurb and Title Match by Ken Schmidt -Genres and Titles by Ken Schmidt - Little and Often by Philip Prowse - Find Your Level by Ken Schmidt - Issue Logs by Janet Watson

CLASS MEETINGS	ACTIVITIES
Week 4	<p><u>Motivating and Supporting Reading; Developing and Consolidating Vocabulary</u></p> <ul style="list-style-type: none"> - Reading Partners by Andy Barfield (with modification: you can ask each pair to take turns reading two books because we don't have more than one copy of each title) - Reading with Children by Paul Nation (books with CD would be appropriate for this activity) - Interactive Storybook Reading by Karen Irene Burrell (interactive story books should work here) - "I Read a Good Book the Other Day!" by Ken Schmidt - Reading Marathon by Ken Kanatani - Stimulating Ads by Yamanaka Junko - One-Sentence Check by Richard R. Day - One a Day by Anthony Silva - Vocabulary Discussion Groups by Richard R. Day
Week 5	<p><u>Developing and Consolidating Vocabulary; Having Fun</u></p> <ul style="list-style-type: none"> - Individualized Vocabulary Tests by Richard R. Day - Identifying Idioms by Richard Day - What Next? By Patrick Fulmer - Anybody You Know? By Patrick Fulmer - Guess Who? By Midori T. Ivano
Week 6	<p><u>Oral Reading Reports; Developing and Consolidating Vocabulary</u></p> <ul style="list-style-type: none"> - Instant Book Report by Marc Helgesen - The 4/3/2 Technique by Willy A. Renandya - Book Review Round Robin by Ken Schmidt - I Know What Comes Next! by Jeremy Taylor - Collecting Collocations by Richard R. Day
Week 7	<p><u>Drama and Role Play; Oral Reading Reports</u></p> <ul style="list-style-type: none"> - Act It Out by Magda Kalinowska - Minidramas by Eva F. K. Lai - Here's the News by Averil Coxhead - The Story and Me by Victoria Rodrigo - Favorite Books by Ken Schmidt - Draw a Picture by Marc Helgesen
Week 8	<p><u>Drama and Role Play; Oral Reading Reports</u></p> <ul style="list-style-type: none"> - Where's the Drama? by Patrick Fulmer - Dramatic Conversations by Richard R. Day - From Character to Drama by Carol Foye - Storytelling Sticks by Marc Helgesen - Too Many Questions by Marc Helgesen - Mind's Eye by Marc Helgesen - Poster Presentations by Andy Barfield

CLASS MEETINGS	ACTIVITIES
Week 9	<p><u>Developing awareness in Reading; Evaluating Reading; Having Fun</u></p> <ul style="list-style-type: none"> - Favorite Quotation by Jana Harper Makaafi - Cultural Clues by Patrick Fulmer - My Favorite Passage by Richard R. Day - One Minute by Midori T. Iwano - Predicting Content from Title by Patrick Fulmer - Shared Dictation by Midori T. Iwano - Musical Chairs by Midori T. Iwano - Describe and Present by Nguyen Thai An
Week 10	<p><u>Written Reading Reports; Writing Creatively</u></p> <ul style="list-style-type: none"> - Quick Book Report Forms by Ken Schmidt - Book Report Checklist by Magda Kalinowska - The Best of Books, the Worst of Books by Simon Evans - Character by the Letter by Donna Prather - Gifts by Patricia Reiss - The Movie Version by Thomas S.C. Farrell - Poetry with Character by Richard R. Day
Week 11	<p><u>Oral Reading Reports; Written Reading Reports; Writing Creatively</u></p> <ul style="list-style-type: none"> - Weekly Reading Review by Averil Coxhead - Book Clubs by Averil Coxhead - Getting Personal by Andy Barfield - One-Sentence Summaries by Jana Harper Makaafi - Picture It by Marc Helgesen - A Touch of Haiku by Tachee Choi - Once Upon a Time by Claire Hitosugi
Week 12	<p><u>Writing Creatively; Increasing Reading Rate</u></p> <ul style="list-style-type: none"> - A Different Ending by Eva F. K. Lai - Sentence Detective by Jeremy Taylor - My Own Story by Anne Burns - Tell Me a Story by Janet Watson - Timed Repeated Readings by Jana Harper Makaafi - Paced Reading by Jana Harper Makaafi

Appendix B

The Components of TABE Tests

Reading	Language	Language Mechanics	Vocabulary	Spelling
Interpret graphic information	Usage	Writing conventions	Word Meaning	Vowels
Words in context	Phrase/sentence formation		Multi-meaning words	Consonants
Recall information	Paragraph development		Words in context	Structural unit
Construct meaning	Capitalization			
Evaluate/Extend meaning	Punctuation			

About the Authors

Dung T. Nguyen, Ph.D., is the Principal of Huong Dao Vietnamese Language School and a Ford Foundation Scholar. She has led teacher development programs in Vietnam, secured international funding, and presented at major conferences. Her research interests include translanguaging, English as a second language, extensive reading, and artificial intelligence. Email: dungnguyenhpu@gmail.com

Ricardo González-Carriedo, Ph.D., (corresponding author) is an associate professor of bilingual education at the University of North Texas. He worked at the University of Paris X-Nanterre (France) for two years and was a secondary teacher in North Carolina and Arizona for eleven years. His research interests include language ideologies in teacher education and literacy development among emergent bilinguals. Email: Ricardo.Gonzalez@unt.edu

Monica Ng, MA., is a professional technical writer. She was a graduate student, pursuing MA degrees in Linguistics and Technical Communication at the time of the study. Email: ngmoni@gmail.com