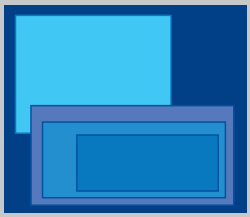


VIAL

VIGO INTERNATIONAL JOURNAL of APPLIED LINGUISTICS

Number 21-2024

Universidade Vigo

Servizo
de Publicacións

VIAL

Vigo International Journal
of Applied Linguistics

Universidade Vigo

VIAL. Vigo International Journal of Applied Linguistics.

VIAL is an international and peer-reviewed journal which is indexed and abstracted in CINDOC (Centro de Información y Documentación Científica), EBSCO, ERIH PLUS, JCR, Web of Knowledge (Arts and Humanities Citation Index, Social Sciences Citation Index), Linguistics Abstracts, Linguistics and Language Behavior Abstracts, Latindex, MLA Directory of Periodicals, PsychoInfo, Scimago Journal Rank Scopus, Ulrich's Database

Webpage: www.webs.uvigo.es/vialjournal/

Printed in Spain - Impreso en España
I.S.S.N.: 1697-0381 - Electronic ISSN: 2660-504X - Depósito Legal: VG-935-2003
Imprime e maqueta: Tórculo Comunicación Gráfica, S.A. Deseño da portada: Elena Gómez.

VIAL

Vigo International Journal of Applied Linguistics

Number 21 - 2024

Contents:

Modeling the contribution of anxiety, enjoyment, and classroomenvironment to boredom among students of English as a Foreign Language

Sami Alanazi
Department of Languages and Translation
Northern Border University, Arar, Saudi Arabia
sami.alenezi@nbu.edu.sa

Elias Bensalem
Department of Languages and Translation
Northern Border University, Arar, Saudi Arabia
elias.bensalem@nbu.edu.sa

Abstract

Over the past few years, there has been a noticeable increase in scholarly attention about the role of emotions in second-language acquisition. There is a consensus among researchers that emotions play a crucial part in learning a new language. The emotional experiences of language learners, such as anxiety, enjoyment, and motivation, have been studied. Strategies to promote positive emotions and decrease negative emotions have been identified. Motivated by a recent academic interest in boredom which is one of the most intense emotions, this study examined the effects of foreign language enjoyment (FLE), foreign language classroom anxiety (FLCA), and classroom environment (CE) on learners' foreign language learning boredom (FLLB). A sample of 481 college students enrolled in English classes in Saudi Arabia participated in the study. There were significant correlations of FLCA, FLE, and FLLB with CE. Structural equation modeling (SEM) analysis revealed that students' FLCA, FLE, and CE were significant predictors of their FLLB, with FLCA being the strongest predictor, followed by FLE and CE. The individual predictive effects of enjoyment, anxiety, and classroom environment outweighed the interactive effects between classroom environment and anxiety, classroom environment and enjoyment, and anxiety and enjoyment. Pedagogical implications are discussed based on the results.

Keywords: anxiety, boredom, classroom environment, enjoyment.

Resumen

En los últimos años ha habido un aumento de la atención académica en el estudio del papel de las emociones en la adquisición de un segundo idioma. Existe un consenso entre los investigadores de que las emociones juegan un papel crucial en el aprendizaje de un nuevo idioma. Se han estudiado las experiencias emocionales de los estudiantes de lenguas, como la ansiedad, el disfrute y la motivación, y se han identificado estrategias para promover las emociones positivas y disminuir las negativas. Motivado por un interés académico reciente en el aburrimiento, este estudio investigó los efectos del disfrute del idioma extranjero (FLE), la ansiedad en el aula de idiomas extranjeros (FLCA) y el ambiente del aula (CE) en el aburrimiento del aprendizaje de lenguas extranjeras de los estudiantes (FLLB). Una muestra de 481 estudiantes universitarios matriculados en clases de inglés participaron en el estudio. Los análisis de correlación de Pearson revelaron que había correlaciones significativas entre los tres constructos (FLCA, FLE y FLLB) y CE. Los hallazgos del análisis de modelos de ecuaciones estructurales (SEM) revelaron que FLCA, FLE y CE de los estudiantes fueron predictores significativos de su FLLB, siendo FLCA el predictor más fuerte, seguido de FLE y CE. Los efectos predictivos independientes de la ansiedad, el disfrute y el entorno del aula superaron los efectos interactivos entre el entorno del aula y la ansiedad, el entorno del aula y el disfrute y la ansiedad y el disfrute. Se discuten las implicaciones pedagógicas a partir de los resultados.

Palabras clave: ansiedad, aburrimiento, ambiente del aula, disfrute, ecuaciones estructurales.

1. Introduction

In the past, there was a lack of scholarly focus on investigating the role of emotions in the process of acquiring a foreign language (Dewaele, 2019; Prior, 2019), with the exception of anxiety (Dewaele & Li, 2018; Li, 2022). This is because emotions were viewed as less important than cognitive factors in language learning (Richards, 2022). Researchers now agree that learners undergo a diverse array of emotions when acquiring a new language, just as they do in other areas of their education (Kruk, 2019; Teimouri, 2018). This shift in view is grounded in the broaden-and-build theory (BBT), which distinguishes between the effects of positive and negative emotions in fostering learners' development (Fredrickson, 2003). The impact of these emotional experiences can be significant, affecting both the learning process and the overall performance of students (e.g., Li & Dewaele, 2021; Li & Wei, 2022; Shao, Pekrun, Marsh, & Loderer, 2020). In light of the seminal research conducted by Dewaele and MacIntyre (2014), which pioneered the examination of both positive and negative

emotions, a subsequent proliferation of empirical studies on emotions in the field of second language acquisition (SLA) has occurred. This body of research includes positive emotions, particularly enjoyment, joy, and hope (Dewaele, Botes, & Greiff, 2023; Huang, 2022; Johnson, 2019; MacIntyre, Mercer, Gregersen, & Hay, 2022; Zhao, 2023), along with negative emotions such as anxiety, anger, distress, sadness, and shame (Teimouri, Goetze, & Plonsky, 2019). However, other emotions commonly experienced by learners, such as foreign language learning boredom (FLLB), have received less attention in the literature and remain underexplored (Li, 2022; Li & Dewaele, 2020; Li & Wei, 2022; Zhao & Wang, 2023). Some second language (L2) educators may underestimate the importance of boredom in the classroom and view it as a minor issue caused by learners' lack of interest in learning the target language (Pawlak, Kruk, Zawodniak, & Pasikowski, 2020).

Nevertheless, FLLB is an emotion that merits greater consideration in the field of SLA because it has been extensively examined in the domain of educational psychology and has been found to have various effects on learning, as demonstrated in research (Li & Han, 2022; Putwain & Pescod, 2018). Increasing our understanding of boredom will help develop awareness of the factors that may affect students' emotional well-being and their language development (Li & Lu, 2022).

The examination of the interplay between positive and negative emotions has received significant scholarly interest within the field of SLA in recent years (e.g., Li, Dewaele, Pawlak, & Kruk, 2022; Li & Wei, 2022; Zhao & Wang, 2023). MacIntyre and Gregersen (2012) suggest that L2 learners experience a simultaneous presence of both negative and positive emotions. Investigating the coexistence of many emotions in L2 learning is deemed necessary, as proposed by Li and Wei (2022). The focus has been primarily on examining the relationship between FLE and FLA. Nevertheless, it is crucial to broaden the existing research by demonstrating a correlation between boredom and enjoyment (Zhao & Wang, 2023). Previous scholarly research has provided evidence to substantiate the credibility of this proposition. These investigations have indicated that the correlation between enjoyment and boredom in the context of SLA is comparable to, or even more significant than, the association between anxiety and enjoyment (Dewaele & Li, 2021; Li et al., 2022). Scholars argue that students' emotions within the educational setting play a significant role in shaping the overall CE (Harvey, Bimler, Evans, Kirkland, & Pechtel, 2012).

Several empirical studies have documented a positive association between CE and FLE as well as a negative link between CE and FLCA among Iranian students (Khajavy, MacIntyre, & Barabadi, 2018). Li, Huang, and Li (2021) conducted a study in a Chinese setting and discovered that FLCA as well as FLE were significantly influenced by CE. The conceptual assumptions and findings reported by previous studies provide

solid proof for the correlations between the classroom environment and emotions experienced by learners (Li & Dewaele, 2021). Within the same vein, Li et al. (2022) argue that FLE, FLCA, CE, and FLLB are interconnected within a complicated feedback loop. Each variable influences the others and is, in turn, influenced by them as the lesson unfolds.

A few studies, however, have investigated the combined impact of FLCA, FLE, and CE on the newly explored concept of FLLB. This type of investigation using mediation analysis can contribute to a deeper understanding of the pathways by which these variables affect FLLB and their implications for language pedagogy. Furthermore, the majority of studies pertaining to boredom in the area of SLA were conducted in China and Poland. It is important to study FLLB in various contexts, including that of the Arabian Peninsula, since different settings might yield different outcomes. This investigation addressed the following two research questions:

Research question (RQ) 1: What are the interrelationships between anxiety, enjoyment, the classroom environment, and boredom?

Research question (RQ) 2: To what extent do anxiety, enjoyment, and the classroom environment contribute to boredom among English Language Learners?

2. Literature review

2.1. Foreign Language Enjoyment and Foreign Language Classroom Anxiety

The predominant focus of research in the domain of second language (L2) acquisition for over forty years has been on the examination of negative emotions, with a specific emphasis on anxiety. The trend toward using positive psychology in SLA has motivated scholars to investigate the significance of positive emotions, particularly enjoyment, in the enhancement of L2 proficiency among learners. The broad-and-build hypothesis proposed by Fredrickson (2003) serves as the fundamental principle underlying the positive psychology movement in the field of language learning. This theoretical framework places significant emphasis on the advantageous impact of pleasant emotions on the development of second language (L2) skills among those engaged in language acquisition. The concept of foreign language enjoyment was developed by Dewaele and MacIntyre (2014). Dewaele and Li (2021) define FLE as "a positive affective state that combines challenge, happiness, interest, fun, sense of pride, and sense of meaning" (p. 5). The experience of enjoyment is particularly prominent in situations when individuals have a certain degree of independence and encounter unforeseen or demanding elements (Dewaele & MacIntyre, 2014).

Two dimensions underneath the construct of FLE have been identified in the literature: private and social. The private dimension is characterized by feelings of happiness, interest, and pride. The social dimension is characterized by shared legends among learners, meaningful engagements with peers and instructors, and the creation of an atmosphere that fosters a sense of learning (Dewaele & MacIntyre, 2016).

FLE can be attributed to a variety of factors, including learners' internal factors such as language background, gender, and age (Dewaele & Dewaele, 2018), as well as teacher-related aspects such as personality and the extent to which teachers employ the target language (Dewaele et al., 2019; Li et al., 2021). Additionally, the classroom environment also plays a role in FLE (Wei, Gao, & Wang, 2019; Li et al., 2021).

The FLCA construct, as originally conceptualized by Horwitz, Horwitz, and Cope (1986), encompasses a collection of self-perceptions, emotions, and convictions that emerge in response to the distinctive attributes of the language acquisition process inside a formal educational environment. It is regarded as a "distinct complex" associated with the experience of learning a language in a classroom setting (p. 128). According to Horwitz (2017), FLCA occurs when learners experience distress due to their incapacity to express their true selves and establish genuine connections with others as a result of the constraints imposed by the new language (p. 41). In fact, the process of introducing oneself to others using a newly acquired language that is not yet fully mastered intrinsically induces anxiety in certain individuals (Horwitz, 2017). Anxiety is characterized by the presence of worries pertaining to communication apprehension, test anxiety, and fear of negative evaluation (Horwitz, 2017). The investigation of FLCA is driven by a scholarly interest in gaining a deeper comprehension of this phenomena, with the ultimate goal of enhancing pedagogical approaches by mitigating its adverse impact (Li & Dewaele, 2021).

Recent research in the field of positive psychology has undertaken investigations on the correlation between anxiety and enjoyment, along with several other emotions. Dewaele and MacIntyre (2014) were among the pioneering researchers who investigated the association between FLE and FLCA. Their study involved a sample of 1,746 foreign language learners with varied linguistic origins. The findings of their research revealed a moderate negative association between FLE and FLCA, implying that these constructs are distinct and relatively autonomous rather than being in direct opposition to one other. Higher levels of FLE and lower levels of FLCA were found to be connected with factors such as increased language competency, older age, Western cultural background, and higher degrees of multilingualism. The study found that there were little differences between genders in both FLE and FLCA. However, it was observed that female participants tended to report somewhat elevated levels of both emotions compared to their male counterparts. The participants identified

the social dimension of FLE as being linked to positive interactions with peers and teachers, which were characterized by humor, praise, encouragement, trust, and respect. Additionally, classroom activities that allowed for greater autonomy were also associated with higher levels of FLE.

In the Saudi context, Dewaele and Alfawzan (2018) found that there were interactions between the individuals' experiences of FLCA and FLE in their language lessons. A similar pattern of relationship between anxiety and enjoyment corresponds to the one uncovered in Bensalem's (2021) study with Saudi EFL students.

In a recent meta-analysis of 56 studies from different educational settings, Botes, Dewaele, and Greiff (2022) found a moderately negative correlation between FLE and FLCA ($r = -.31$, $k = 46$, $N = 20{,}946$) (p. 3). The authors concluded that the association between FLE and FLCA is most likely circular, as the alleviation of anxiety can lead to more enjoyment in the FL classroom, which in turn leads to an even lower level of anxiety, as discussed in previous research (Botes, Dewaele, & Greiff, 2020).

2.2. Classroom environment

The classroom environment (CE) refers to the collective ambiance, atmosphere, or affective tone that is present inside the confines of the classroom setting (Dorman, Fisher, & Waldrip, 2006). The impact of CE on students' cognitive, emotional, and behavioral dimensions within the educational setting has been highlighted by Li et al. (2021). In their study, Peng and Woodrow (2010) identified three dimensions of classroom environment, namely teacher support, student cohesion, and task orientation. The aforementioned elements are exemplified by favorable features of classroom environment, such as student cohesion, teacher assistance, student engagement, goal-oriented behavior, collaborative efforts, and equitable treatment (Li, Dewaele, & Hu, 2023).

Prior research in the field of general education has investigated the correlations between CE and students' emotional states, as well as their levels of involvement and motivation. Research has shown that a positive CE is associated with positive emotions, heightened motivation, and active participation in the learning process. Conversely, a negative CE is correlated with negative emotions, decreased motivation, and a lack of interest in learning activities (Reyes, Brackett, Rivers, White, & Salovey, 2012).

One of the early studies that investigated the potential association between CE and the emotional experiences of students among 1,528 Iranian secondary school students (Khajavy et al., 2018). The results showed a connection between a positive CE and FLE. Additionally, the researchers found a negative correlation between a positive

CE and FLCA. Li et al. (2021) completed a study within the context of English as a Foreign Language (EFL) in China. The findings of their research indicated that both FLCA and FLE were influenced by CE. This implies that the presence of a supportive classroom atmosphere and the cultivation of emotional intelligence have the potential to enhance FLE and reduce FLCA in EFL students. In their study, Li, Dewaele, Pawlak, and Kruk (2022) employed a large sample of Chinese college students to examine the interconnections among CE, FLE, FLCA, FLLB, and the willingness to engage in English. The findings revealed statistically significant associations between communicative anxiety, the emotional states of students, and their willingness to communicate in the target language. The results of this study offer compelling support for a significant association between the classroom environment and the range of emotions, both positive and negative, that students experience.

2.3. Foreign Language Learning Boredom

Boredom refers to "a state of disengagement" (Kruk & Zawodniak, 2020, p. 16). According to Li, Dewaele, and Hu (2021), boredom in the context of L2 acquisition can be defined as "a negative emotion with an extremely low degree of activation/arousal that arises from ongoing activities...[that] are typically over- or under-challenging and/or of little significance, relevance, or meaning to the learners" (p. 12).

Boredom is a multidimensional construct (Pekrun Goetz, Frenzel, Barchfeld, & Perry, 2010). The experience of boredom encompasses various dimensions, including affective aspects characterized by an unpleasant feeling, cognitive aspects involving a perception of time passing slowly, motivational aspects encompassing a desire to change or escape from the source of boredom or engage in alternative activities, expressive aspects involving postural, facial, and vocal expressions, and physiological aspects characterized by decreased arousal levels. Pekrun's (2006) three-dimensional taxonomy can be utilized to articulate the several dimensions of boredom, which include valence, activation, and objective focus. Valence pertains to the degree of positive or negative affectivity connected with an emotional experience. Activation refers to the degree of physiological or cognitive activation or deactivation that accompanies the experience of an emotion. The concept of objective focus pertains to the orientation of emotion towards either activity or outcome. In other words, boredom arises as a result of a continuous activity or is triggered by its past or future results (Wang, 2023).

Boredom has long been a neglected area of research. Early studies examining FLLB among L2 learners were conducted in the Polish context (e.g., Kruk & Zawodniak, 2018, 2020; Pawlak, Kruk, & Zawodniak, 2020; Pawlak, Kruk, Zawodniak, et al., 2020; Pawlak, Zawodniak, & Kruk, 2020). Then FLLB has garnered significant attention in recent years (e.g., Derakhshan, Kruk, Mehdizadeh, & Pawlak, 2021; Dewaele,

Albakistani, & Ahmed, 2022; Dewaele, Botes, & Greiff, 2023; Dewaele, Botes, & Meftah, 2023; Kruk, Pawlak, Shirvan, & Soleimanzadeh, 2023; Li, 2021, 22; Wang & Li, 2022). The identification of boredom as an adverse emotion that is frequently experienced by language learners (Zhao & Wang, 2023) has paved the way for studies that examined the connection between boredom and other constructs, namely FLCA and FLE. Previous research has identified a noteworthy observation that the correlation between FLE and FLLB is either equivalent to or more significant than the connection between FLE and FLCA in the field of L2 acquisition (Dewaele & Li, 2021; Li et al., 2022).

In a study with the participation of 880 college level students in China enrolled in online English courses, Wang, & Li (2022) discovered a negative medium to high negative correlation between FLLB and FLCA and with a medium positive correlation between FLLB and FLE, suggesting that students who reported higher levels of enjoyment tended to display lower levels of anxiety and boredom, whereas those who reported lower levels of enjoyment tended to exhibit higher levels of anxiety and boredom. Similar type of correlation was reported between FLLB and FLE in the Chinese face-to-face context (Li, 2022).

Other studies have examined the FLLB and its associations with various emotions in underexplored contexts. For example, in a unique study that involved a group of 332 FL learners from different countries, Dewaele, Botes, and Greiff (2023) examined the relationships between FLLB, FLCA, and FLE. The findings revealed that students with greater levels of FLE had considerably lower levels of FLCA and FLLB than students with lower levels of FLE. In addition, learners who had higher levels of FLCA experienced greater levels of boredom. The positive relationships between FLLB, FLE, and FLCA confirm the results reported by Dewaele, Botes, and Meftah (2023), as well as Dewaele and Meftah (2023), who found a significant positive correlation between FLCA and FLLB among EFL learners in Morocco.

Finally, one of the rare studies that examined the correlation of anxiety, enjoyment, boredom, CE, and willingness to communicate with a large sample of 2,268 university-level students (Li et al., 2022) revealed that FLLB was negatively correlated with FLE while FLLB was positively correlated with FLCA, corroborating the findings reported by Li (2022) and Li and Wei (2022). In addition, Li et al. (2022) reported that students who were pleased with their classroom environment experienced higher FLE, as reported by previous studies (Dewaele, 2019; Dewaele & Dewaele, 2018). One major finding is that FLCA is outweighed by FLLB, indicating that boredom is more prone to negatively impact the entire classroom climate in comparison to anxiety.

3. Method

3.1. Participants and setting

A total of 481 EFL students at various campuses of a public university in Saudi Arabia participated in the study. Females made up the majority of the sample (80.26%), with males accounting for the remaining 19.8%. The majors of the students were English (61.5%) and medicine (38.5%). Students of medicine took courses in English that were designed and taught by the same teaching staff as the English major courses. The main reason for involving medical students was to increase the study sample size. Other reasons were the convenience and availability of the participants, as one of the researchers had direct access to medical students. The ages of the students varied between 18 and 27. The average age was 21.76 (SD = 2.56). Most of the participants (35.8%) were in their first year of college, and 33.4% were in their second year. The rest of the students included in the study were in their third year (18.7%) or fourth year (12.1%) of their Bachelor program.

3.2. Instruments

The data were gathered using a composite questionnaire (see Appendix). The first section contains background information about the participants (age, gender, major, and year of study). The second section includes scales measuring the main constructs: FLCA, FLE, and FLLB, as well as the CE. All items in the survey were in both English and Arabic to make sure that all items were clear to all students. Bensalem (2021) validated the Arabic versions of FLCA and FLE. Two professors of translation whose native language is Arabic translated the FLLB and CE scales.

3.2.1. Foreign Language Enjoyment

The researchers measured FLE using an adapted version of Botes, Dewaele, and Greiff's (2021) Short Form of the Foreign Language Enjoyment Scale (S-FLES), with each item containing the term English class. This scale is based on Dewaele and MacIntyre's (2014) original 21-item scale, including nine items that assess positive emotions in second-language learning. The S-FLES takes into account three aspects: instructor appreciation, personal enjoyment, and social enjoyment. The scale was scored on a 5-point Likert scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree". Cronbach's alpha internal consistency coefficient of the scale was 0.92.

3.2.2. Foreign Language Classroom Anxiety

The assessment of FLCA was conducted using the 8-item Short-Form Foreign Language Anxiety Scale (S-FLCAS), a measurement tool developed and validated by Botes et al. (2021). The initial version of the FLCAS consisted of a total of 33 components. The assessment of FLCA was conducted using the 8-item Short-Form Foreign Language Anxiety Scale (S-FLCAS), a measurement tool developed and validated by Botes et al. (2021). The initial version of the FLCAS consisted of a total of 33 components. The participants used a five-point Likert scale to indicate their level of agreement or disagreement with the survey items, with 1 representing "strongly disagree" and 5 representing "strongly agree". The Scale of Foreign Language Classroom Anxiety (S-FLCAS) comprises two items that are reverse-coded: "I do not experience apprehension regarding errors made in English class" and "I possess a sense of assurance when engaging in oral communication in English class". This implies that those who express great agreement with these assertions will obtain a lower score on that particular item, hence indicating a lower level of anxiety. The present study demonstrated acceptable internal consistency, as evidenced by the scale's Cronbach's alpha coefficient of .71.

3.2.3. Foreign Language Learning Boredom

The measurement of FLLB was conducted using a set of eight items derived from the Achievement Emotions Questionnaire developed by Pekrun, Goetz, Frenzel, Barchfeld, and Perry (2011). The questionnaire items were modified to align with the specific context of EFL learning. The participants were instructed to express their degree of concurrence with the assertions using a five-point scale encompassing the options of "strongly disagree" to "strongly agree". The findings demonstrated strong internal consistency for the scale, as shown by a Cronbach's alpha coefficient of .92.

3.2.4. Classroom environment

The researchers employed a modified version of the classroom environment portion from the "What Is Happening In This Class?" (WIHIC) questionnaire developed by Peng and Woodrow (2010) to evaluate the construct of classroom environment. The scale comprises 10 items that have been adapted for application in an EFL context. The survey items encompass three distinct domains: task orientation (e.g., "The tasks implemented in this course are of practical value"), student cohesion (e.g., "I effectively collaborate with my peers in this class"), and teacher support (e.g., "The English instructor demonstrates patience in their teaching approach"). According to the research conducted by Peng and Woodrow (2010), the alpha reliability coefficient for the three sub-factors was found to be .80. The current investigation yielded a coefficient alpha value of .90.

3.3. *Procedure*

The data collection was conducted during the fall semester of 2022, subsequent to obtaining approval from the Institutional Review Board (IRB) of Northern Border University. The survey was built using Google Forms. A link to the survey, which includes participants' background information and the four scales (FLCA, FLE, FLLB, and CE), was sent to students via email. The participants were informed about the study's goal and objectives, as well as the confidentiality of their responses and their ability to withdraw from the study at any time. They freely completed the questionnaire, which required around 20 minutes to finish.

3.4. *Data analysis*

To evaluate the correlations between FLCA, FLE, CE, and FLLB, Pearson correlation analysis was carried out in SPSS 25. Hair, Sarstedt, Pieper, and Ringle (2012) proposed using confirmatory factor analysis (CFA) to assess the validity of the scales. The hypothesized model was then evaluated using goodness-of-fit indices utilizing structural equation modeling (SEM) with maximum likelihood. To assess model fit, four indices were used: normed chi-square (i.e., 2/df), CFI, TLI, and RMSEA. The Hu and Bentler (1999) criteria were used to determine whether the model had a satisfactory fit. An RMSEA value of less than 0.06 (0.06 to 0.08 indicates a moderate or adequate fit), as well as CFI and NNFI (or TLI) values more than or equal to 0.95 (0.90 indicates a moderate or adequate fit), are examples. Furthermore, a normed chi-square of 2/df 2 is regarded as a good match when the sample size is greater than 200, although a normed chi-square of 2/df 5 is deemed acceptable (Bentler, 1990).

4. Results

The data were first screened for normality and outliers. The data's normality was evaluated using skewness and kurtosis values that were between -2 and +2, showing normality (Kunnan, 1998). Furthermore, the reliability of each scale was calculated using the scales' coefficient alphas. The results were greater than 0.70, indicating a satisfactory level of internal consistency. Table 1 shows descriptive statistics and normality.

Table 1: Descriptive statistics (means and SDs) and normality (N = 481)

	Mean	SD	Skewness	Kurtosis
1. Anxiety	3.38	0.58	.06	-.19
2. Enjoyment	3.76	0.77	-.98	1.15
3. ClassroomEnvironment	4.05	0.85	-.24	-.59
4. Boredom	2.62	.98	.21	-.65

**Indicates significance at the 0.01 level.

Following that, the scales were investigated for their construct validity using confirmatory factory analysis (CFA). The goodness of fit indices fell within the acceptable range, as seen in Table 2.

Table 2: Measurement Model of the Latent Variables

	$\chi 2$	DF	$\chi 2/df$	CFI	TLI	RMSEA	Cronbach's α
Anxiety	37.37	19	1.96	.99	.99	.04	.71
Enjoyment	199.21	25	7.97	.94	.92	.02	.92
Classroom Environment	442.84	31	14.28	.96	.94	.03	.90
Boredom	118.83	19	6.25	.96	.94	.02	.92

The Pearson correlation analysis results, as shown in Table 3, demonstrated a significant and negative association between FLLB and FLE, as well as between FLLB and CE, with moderate effect sizes (r = -.527, p.001; r = -.468, p < .001). FLLB was found to have a moderately significant relationship with FLCA, with a medium effect size (r =.521, p < .001). Furthermore, FLCA was revealed to be negatively associated with both CE (r = -.231, p < .001) and FLE (r = -.170, p < .001). FLE and CE, on the other hand, had a significant positive correlation with a large effect size (r =.622, p < .001).

Table 3: The relationships between FLCA, FLE, CE, and FLLB

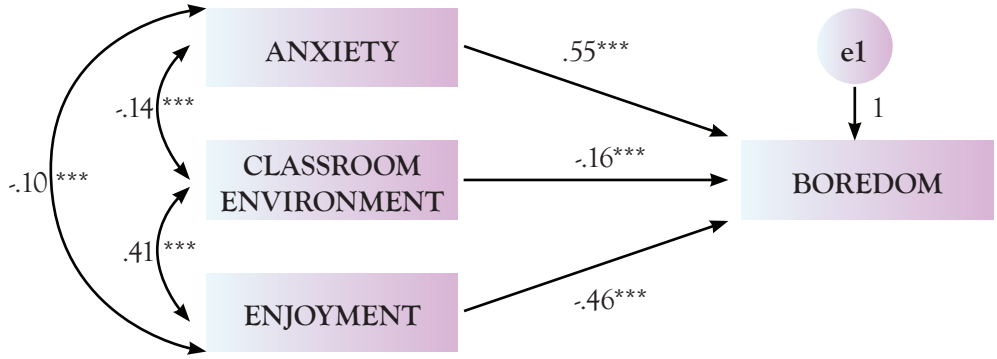
Correlations				
	1	**2**	**3**	**4**
1. Anxiety	1.0			
2. Enjoyment	-.170**	1.0		
3. ClassroomEnvironment	-.231**	.622**	1.0	
4. Boredom	.521**	-.527**	-.468**	1.0

**Indicates significance at the 0.01 level.

AMOS 26 was then used to test the hypothesized model. Variance-covariance matrices were employed as input, and the maximum likelihood technique was used. All of the coefficients were significant ($p < .05$), and the fit indices were within acceptable limits (see Figure 1). FLCA had the biggest effect on FLLB (β = .55, R2 =. 16, f2 =.19) with a medium effect size, according to SEM data. Additionally, the SEM findings revealed that FLE predicted FLLB negatively and significantly (β =. 46, R2 =.10, f2 =.11) with a small effect size. In addition, FLE predicted FLLB negatively and significantly CE (β = .16, R2 = .07, f2 =.07) bigger effect size.

The associations among the predictors were also explored, and it was found that FLCA had a significant and negative relationship with CE ($r = .14$, $p < .001$) and FLE ($r = .10$, $p < .005$). Moreover, a significant data-driven correlational path was identified between CE and FLE ($r = 0.41$, $p < 0.001$), as well as between FLE and FLCA ($r = 0.10$, $p < 0.001$). Overall, the SEM analysis suggests that FLCA, CE, and FLE, are significant predictors of FLLB in EFL students, with FLCA being the strongest predictor, followed by FLE and CE. Notably, the independent predictive effects of FLCA, FLE, and CE were more substantial than the interactive effects between general CE and FL, CE and FLE, and FLCA and FLE.

Figure 1: The Final Model of Anxiety, Enjoyment, Classroom Environment, and Boredom



***p < .001

5. Discussion

The first RQ examined the associations of FLCA, FLE, and CE with FLLB. The results indicated a positive association between FLLB and FLCA, suggesting that there is a tendency for students who report greater boredom to also experience higher levels of anxiety. Similar results were reported by Li et al. (2021) and Wang and Li (2022). A poor classroom environment can increase the probability of students' concurrent experiences of anxiety and boredom (Li et al., 2022). This explains the negative correlation between boredom and the current study participants' perceptions of the classroom environment. This finding offers additional support for the significant contribution made by teachers in creating a cooperative learning environment and designing engaging tasks, which can facilitate positive emotions and impede negative emotions in the process of SLA (Khajavy et al., 2018; Li, 2022; Li et al., 2023).

In addition, the study revealed a significant negative relationship between FLLB and FLE, indicating that students who do not enjoy learning their foreign language are more likely to get bored. This result is not surprising, confirming previous studies (Dewaele, Albakistani, et al., 2022; Dewaele, Botes, & Greiff, 2023; Dewaele, Botes, & Meftah, 2023; Li et al., 2022; Wang & Li, 2022; Zhao & Wang, 2023). The positive correlation between boredom and enjoyment provides additional support for the validation of the "undo hypothesis" proposed by the BBT (Fredrickson, 2004, p. 1371). Specifically, it suggests that the presence of positive emotions, such as enjoyment has the potential to counteract the negative impacts of boredom-induced negative emotions. This process is thought to occur by expanding an individual's range of cognitive and behavioral responses. However, it is important to acknowledge that positive emotions may not necessarily completely eradicate bad emotions, despite their significance in mitigating negative emotional experiences (Li & Han, 2022).

The second RQ investigated the influence of FLCA, FLE, and CE on Saudi EFL students' FLLB. The SEM analysis revealed that FLCA positively and significantly predicted FLLB with the greatest impact. Conversely, FLE and CE negatively and significantly predicted FLLB. These findings suggest that FLCA and FLLB are closely related, with heightened FLCA being associated with higher levels of FLLB. This could be due to personal tendencies, such as being more prone to boredom, as described in Pawlak et al.'s (2020) study, or the influence of personality on FLCA, as noted by Horwitz et al. (1986). Personality traits "refer to consistent patterns in the way individuals behave, feel, and think" (Pervin & Cervone, 2010, p. 228). The taxonomy of personality traits, which is called the Big Five, Conscientiousness, Extraversion, Agreeableness and Neuroticism (2010, p. 228). According to a study conducted by Dewaele and Al-Saraj (2015), it was found that Saudi EFL students who displayed

higher levels of emotional stability and extraversion tended to have lower levels of anxiety. The results reported by the current study demonstrate the complex interplay between emotions and variables that contribute to the experience of FLLB.

The construct of enjoyment was found to be the second-biggest negative predictor of FLLB. This outcome supports claims that positive emotions such as enjoyment can potentially play a preventive or protective role in mitigating unpleasant emotions endured by L2 learners (Li & Wei, 2022; MacIntyre, 2017) and foster behaviors such as play, creativity, curiosity, and exploration, which are widely regarded as advantageous to learning (Boudreau, MacIntyre, & Dewaele, 2018). Furthermore, learners who derive enjoyment from their learning experience tend to exhibit increased interest and engagement in the L2 (MacIntyre & Gregersen, 2012). Consequently, the level of FLLB will likely decrease when teachers prioritize the optimization of their curriculum planning and enhance the delivery of their lessons to ensure a more engaging and captivating experience for students since FLLB is associated with a sense of disengagement and a lack of interest in the target language. The findings can be explained by Fredrickson's (2003) BBT, which posits that encountering positive emotions may foster the cultivation of individual resources that can boost one's state of well-being and resilience in the long run. When students enjoy the English language learning process, they may exhibit a greater propensity to engage in classroom activities and are more prone to experiencing positive affective states, such as curiosity and excitement. These positive emotions, in turn, may broaden their perspectives and build personal resources, leading to lower levels of FLLB. Moreover, it has been observed that the presence of positive emotions, such as enjoyment, can potentially mitigate the adverse consequences of anxiety, which in turn is positively associated with boredom. The experience of positive emotions can potentially enhance learners' ability to effectively manage the stressors and difficulties that arise in the process of acquiring a foreign language.

The classroom environment emerged as a significant predictor of boredom. This result corroborates the results reported by Li et al. (2022). Thus, it seems that the events taking place during a lesson, such as the teacher's behavior and the type of interaction between students, as well as the implementation of tasks, are related to students' experiences of FLLB (Dewaele & MacIntyre, 2014, 2019). The instructor plays a pivotal role in establishing a conducive learning environment in the classroom. Moskowitz and Dewaele (2019) suggest that the passion displayed by instructors might exert a favorable influence on students, resulting in an enhancement of enjoyment. When students share positive emotional experiences, it can reduce negative emotions such as fear and FLLB while promoting positive emotions (Li & Dewaele, 2021; Sadoughi & Hejazi, 2021), establishing a positive and engaging CE, and encouraging

participation in class activities (Sadoughi & Hejazi, 2021). These findings are consistent with educational psychology research, which shows a link between positive emotions and increased engagement on the one hand and negative emotions and disengagement on the other (Oga-Baldwin, 2019; Philp & Duchesne, 2016).

6. Conclusion

Structural equation modeling revealed that FLCA, FLE, and CE all had an impact on EFL students' FLLB. The current study extended our comprehension of the connections between FLLB and two major emotions (FLCA and FLE), as well as the CE, in the understudied context of Saudi Arabia. FLCA was the strongest positive predictor, while FLE was the strongest negative predictor of students' FLLB. The current study provides additional quantitative evidence that FLE also plays a role in reducing the effects of negative emotions. Positive emotions may help to counteract or reduce the negative impact of negative emotions (Li & Wei, 2022). The CE has been identified as a potential predictor of student FLLB, emphasizing the importance of teachers in creating an environment that promotes enjoyment, reduces FLCA, and prevents FLLB. As a result, teachers play an important role in shaping the learning environment to promote student engagement and success. Such a classroom environment will result not only in linguistic progress but also in fostering the emotional well-being of learners. Learners experience different types of emotions while learning a second language. Future research could investigate how other emotions such as shame, resilience, and grit affect the relationship between FLLB and students. By exploring these potential mediators, researchers have the opportunity to enhance their comprehension of the intricate elements that contribute to student FLLB.

The findings of this study provide implications for language practitioners. Undoubtedly, the establishment of a stimulating and interactive classroom setting, characterized by the development of strategies aimed at mitigating anxiety and preventing boredom (Li & Wei, 2022), is crucial for the enhancement of language proficiency (Li et al., 2021). In order to help students reduce the lingering effects of FLLB, which are related to FLCA, a lack of FLE, and the CE, language teachers should consider adopting a variety of strategies that make learning more interactive, interesting, and meaningful. This could be achieved by creating a positive learning environment that fosters a sense of community, which is crucial to keeping students engaged. Teachers can encourage students to feel comfortable sharing their ideas and opinions by creating a supportive and nonjudgmental atmosphere. Providing regular feedback and praise also helps students stay motivated and engaged. In this regard, Li and Dewaele (2021) asserted that the use of humor and sensible praise from teachers, in conjunction with language challenges that are neither extremely

difficult nor overly easy, can increase the level of enjoyment experienced by learners and help anxious students overcome the paralyzing effects of anxiety. Finally, teachers need to continuously adapt and modify their lessons to suit students' interests and needs. Personalizing the learning experience makes it more relevant to learners, which in turn makes it more interesting and engaging. By making the learning experience relevant, teachers can help students overcome FLLB and stay motivated throughout the language learning journey.

The present study has a number of limitations. First, the researchers utilized only quantitative data, which limits the extent and variety of observations that can be drawn regarding the connections among CE, FLE, FLCE, and FLLB. A more comprehensive understanding of the connections and underlying factors between these variables could be gained by conducting in-depth research that involves interviews or classroom observations. Second, this study did not examine other possible factors, such as learner-internal variables related to age, gender, and self-perceived proficiency in English, that may contribute to boredom beyond the quality of the CE and the two emotions it triggers. Third, the boredom scale that was used did not include items that reflect the different dimensions of boredom. In order to enhance the accuracy of capturing learners' different emotional experiences, it is recommended that future research use more comprehensive scales that encompass both in-class and out-of-class activities (Li & Wei, 2022). This approach is supported by scholars such as Kruk (2022) and Kruk & Pawlak (2022). In addition, the study sample included participants from two different majors, namely English and medicine, with different levels of proficiency. Even though all courses that were offered to medical students were designed and taught by faculty members from the English program, the study outcomes might have been affected. Participants from various majors may have varying levels of motivation and interest, which could lead to disparities in findings. Finally, the outcomes of this study are context-specific, focusing on university students in Saudi Arabia. In order to enhance comprehension of the extent to which the findings may be applied to various educational settings, it would be beneficial to investigate the applicability of the results to alternative educational contexts, diverse language learning environments, and varying educational levels, such as primary or secondary schools. Conducting further research in these areas would yield a more comprehensive understanding of the relationships between CE, emotions, and FLLB.

Acknowledgement

The authors wish to acknowledge the approval and the support of this research study by the grant no. EAAA 2022-11-1681 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

7. References

Bensalem, E. (2021). Classroom enjoyment and anxiety among Saudi undergraduate EFL students: Does gender matter? *Vigo International Journal of Applied Linguistics, 18*, 9–34. https://doi.org/10.35869/vial.v0i18.3363

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. http://dx.doi.org/10.1037/0033-2909.107.2.238

Botes, E., Dewaele, J.-M., & Greiff, S. (2020). The foreign language classroom anxiety scale and academic achievement: An overview of the prevailing literature and a meta-analysis. *Journal for the Psychology of Language Learning, 2*(1), 26–56. https://doi.org/10.52598/jpll/2/1/3

Botes, E., Dewaele, J.-M. & Greiff, S. (2021) The development of a short-form foreign language enjoyment scale. *The Modern Language Journal, 105*(4), 858-876. https://doi.org/10.1111/modl.12741

Botes, E., Dewaele, J.-M., & Greiff, S. (2022). Taking stock: A meta-analysis of the effects of foreign language enjoyment. S*tudies in Second Language Learning and Teaching, 12*(2), 205-232. https://doi.org/10.14746/ssllt.2022.12.2.3

Boudreau, C., MacIntyre, P. D., & Dewaele, J.-M. (2018). Enjoyment and anxiety in second language communication: An idiodynamic approach. *Studies in Second Language Learning and Teaching, 8*(1), 149–170. https://doi.org/10.14746/ssllt.2018.8.1.7

Derakhshan, A., Kruk, M., Mehdizadeh, M., & Pawlak, M. (2021). Boredom in online classes in the Iranian EFL context: Sources and solutions. *System, 101*, 102556. https://doi.org/10.1016/j.system.2021.102556

Dewaele, J.-M. (2019). The Effect of classroom emotions, attitudes toward English, and teacher behavior on willingness to communicate among English foreign language learners. *Journal of Language and Social Psychology, 38*(4), 523–535. https://doi.org/10.1177/0261927x19864996

Dewaele, J.-M., Albakistani, A., & Ahmed, I. K. (2022). Levels of foreign language enjoyment, anxiety and boredom in emergency remote teaching and in in-person classes. *The Language Learning Journal,* 1–14. https://doi.org/10.1080/09571736.2022.2110607

Dewaele, J.-M., & Alfawzan, M. (2018). Does the effect of enjoyment outweigh that of anxiety in foreign language performance? *Studies in Second Language Learning and Teaching, 8*(1), 21–45. https://doi.org/10.14746/ssllt.2018.8.1.2

Dewaele, J.-M., & Al-Saraj, T. M. (2015). Foreign Language Classroom Anxiety of Arab learners of English: The effect of personality, linguistic and sociobiographical variables. *Studies in Second Language Learning and Teaching, 5*(2), 205–228. https://doi.

org/10.14746/ssllt.2015.5.2.2

Dewaele, J.-M., Botes, E. & Greiff, S. (2023). Sources and effects of foreign language enjoyment, anxiety and boredom: A structural equation modelling approach. *Studies in Second Language Acquisition, 45*(2), 461 – 479. https://doi.org/10.1017/S0272263122000328

Dewaele, J.-M., Botes, E., & Meftah, R. (2023). A three-body problem: The effects of foreign language anxiety, enjoyment, and boredom on academic achievement. *Annual Review of Applied Linguistics,* 1–16. https://doi.org/10.1017/s0267190523000016

Dewaele, J.-M., & Dewaele, L. (2018). Learner-internal and learner-external predictors of Willingness to Communicate in the FL Classroom. *Journal of the European Second Language Association, 2*(1), 24. https://doi.org/10.22599/jesla.37

Dewaele, J.-M., & Li, C. (2018). Editorial. *Studies in Second Language Learning and Teaching, 8*(1), 15–19. https://doi.org/10.14746/ssllt.2018.8.1.1

Dewaele, J.-M., & Li, C. (2021). Teacher enthusiasm and students' social-behavioral learning engagement: The mediating role of student enjoyment and boredom in Chinese EFL classes. *Language Teaching Research, 25*(6), 922–945. https://doi.org/10.1177/13621688211014538

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., & MacIntyre, P. D. (2016). Foreign language enjoyment and foreign language classroom anxiety: The right and left feet of the language learner. In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 215–234). Bristol: Multilingual Matters.

Dewaele, J.-M., & MacIntyre, P. D. (2019). The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy: A collection of instructed second language acquisition studies* (pp. 263-286). London: Routledge.

Dewaele, J.-M., & Meftah, R. (2023). The trajectory of English foreign learners' emotions and motivation from the start to the end of their learning journey: A Pseudo-longitudinal investigation. *Journal of the European Second Language Association, 7*(1), 16–30. https://doi.org/10.22599/jesla.94

Dorman, J. P., Fisher, D. L., & Waldrip, B. G. (2006). Learning environments, students' perceptions of assessment, academic efficacy and attitudes to science: A LISREL analysis. In D. Fisher, & M. S. Khine (Eds.), *Contemporary approaches to research on learning environments* (pp. 1 - 28). World Scientific Publishing.

Fredrickson, B. L. (2003). The value of positive emotions: The emerging science of positive psychology is coming to understand why it's good to feel good. *American Scientist, 91*, 330-335.

Hair, J. F., Sarstedt, M., Pieper, T. M., & Ringle, C. M. (2012). The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. *Long Range Planning, 45*(5–6), 320–340. https://doi.org/10.1016/j.lrp.2012.09.008

Harvey, S. T., Bimler, D., Evans, I. M., Kirkland, J., & Pechtel, P. (2012). Mapping the classroom emotional environment. *Teaching and Teacher Education, 28*(4), 628–640. https://doi.org/10.1016/j.tate.2012.01.005

Horwitz, E. K. (2017). On the Misreading of Horwitz, Horwitz and Cope (1986) and the Need to Balance Anxiety Research and the Experiences of Anxious Language Learners. *New Insights into Language Anxiety,* 31–48. https://doi.org/10.21832/9781783097722-004

Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125-132. http://dx.doi.org/10.2307/327317

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55. http://dx.doi.org/10.1080/10705519909540118

Huang, C. (2022). Self-regulation of learning and EFL learners' hope and joy: A review of literature. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.833279

Johnson, M. K. (2019). Joy: a review of the literature and suggestions for future directions. *The Journal of Positive Psychology*, *15*(1), 5–24. https://doi.org/10.1080/17439760.2019.1685581

Khajavy, G. H., MacIntyre P. D., & Barabadi, E. (2018). Role of the emotions and classroom environment in willingness to communicate: Applying doubly latent multilevel analysis in second language acquisition research. *Studies in Second Language Acquisition, 40*, 605–624. https://doi.org/10.1017/s0272263117000304

Kruk, M. (2019). Dynamicity of perceived willingness to communicate, motivation, boredom and anxiety in Second Life: the case of two advanced learners of English. *Computer Assisted Language Learning, 35*(1–2), 190–216. https://doi.org/10.1080/09588221.2019.1677722

Kruk, M. (2022). Dynamicity of perceived willingness to communicate, motivation, boredom and anxiety in second life: The case of two advanced learners of English.

Computer Assisted Language Learning, 35, 190–216. https://doi.org/10.1080/09588221.2019.1677722

Kruk, M., Pawlak, M., Shirvan, M. E., & Soleimanzadeh, S. (2023). Revisiting boredom in practical English language classes via exploratory structural equation modeling. *Research Methods in Applied Linguistics, 2*(1), 100038. https://doi.org/10.1016/j.rmal.2022.100038

Kruk, M., Pawlak, M., & Zawodniak, J. (2021). Another look at boredom in language instruction: The role of the predictable and the unexpected. *Studies in Second Language Learning and Teaching, 11*(1), 15–40. https://doi.org/10.14746/ssllt.2021.11.1.2

Kruk, M., & Zawodniak, J. (2018). Boredom in practical English language classes: Insights from interview data. In L. Szymański, J. Zawodniak, A. Łobodziec, & M. Smoluk (Eds.), *Interdisciplinary views on the English language, literature and culture* (pp. 177–191). Zielona Góra: Uniwersytet Zielonogórski.

Kruk, M., & Zawodniak, J. (2020). A Comparative study of the experience of boredom in the L2 and L3 classroom. *English Language & Learning, 44*, 417-437. https://doi.org/10.1007/s42321-020-00056-0

Kunnan, A. J. (1998). An introduction to structural equation modelling for language assessment research. *Language Testing, 15*, 295-332. https://doi.org/10.1177/026553229801500302

Li, C. (2021). A Control–Value Theory Approach to boredom in English classes among university students in China. *The Modern Language Journal, 105*(1), 317–334. https://doi.org/10.1111/modl.12693

Li, C. (2022). Foreign language learning boredom and enjoyment: The effects of learner variables and teacher variables. *Language Teaching Research,* 136216882210903. https://doi.org/10.1177/13621688221090324

Li C., & Dewaele J.-M. (2020). The predictive effects of trait emotional intelligence and online learning achievement perceptions on foreign language class boredom among Chinese university students. *Foreign Languages and Foreign Language Teaching, 5*, 33-44. https://doi.org/10.13458/j.cnki.flatt.004711

Li, C., & Dewaele, J.-M. (2021). How classroom environment and general grit predict foreign language classroom anxiety of chinese efl students. *Journal for the Psychology of Language Learning, 3*(2), 86-98. Retrieved from https://www.jpll.org/index.php/journal/article/view/71

Li, C., Dewaele, J.-M. & Hu, Y. (2023). Foreign language learning boredom: conceptualization and measurement. *Applied Linguistics Review, 14*(2), 223-249. https://doi.org/10.1515/applirev-2020-0124

Li, C., Dewaele, J.-M., Pawlak, M., & Kruk, M. (2022). Classroom environment and willingness to communicate in English: The mediating role of emotions experienced by university students in China. *Language Teaching Research,* 136216882211116. https://doi.org/10.1177/13621688221111623

Li, C., & Han, Y. (2022). Learner-internal and learner-external factors for boredom amongst Chinese university EFL students. *Applied Linguistics Review.* https://doi.org/10.1515/applirev-2021-0159

Li, C., Huang, J., & Li, B. (2021). The predictive effects of classroom environment and trait emotional intelligence on foreign language enjoyment and anxiety. *System, 96*, 1-11. https://doi.org/10.1016/j.system.2020.102393

Li, C. & Lu, X. (2022). Foreign language learning boredom: A scoping review. *Foreign Language Education, 43*(6), 70-77.

Li, C., & Wei, L. (2022). Anxiety, enjoyment, and boredom in language learning amongst junior secondary students in rural China: How do they contribute to L2 achievement? *Studies in Second Language Acquisition, 45*(1), 93-108. https://doi.org/10.1017/s0272263122000031

MacIntyre, P. D., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive-broadening power of the imagination. *Studies in Second Language Learning and Teaching, 2*(2), 193-213. https://doi.org/10.14746/ssllt.2012.2.2.4

MacIntyre, P. D., Mercer, S., Gregersen, T., & Hay, A. (2022). The role of hope in language teachers' changing stress, coping, and well-being. *System, 109*, 102881. https://doi.org/10.1016/j.system.2022.102881

Moskowitz, S., & Dewaele, J.-M. (2019). Is teacher happiness contagious? A study of the link between perceptions of language teacher happiness and student attitudes. *Innovation in Language Learning and Teaching, 15*(2), 117–130. https://doi.org/10.1080/17501229.2019.1707205

Oga-Baldwin, W. L. Q. (2019). Acting, thinking, feeling, making, collaborating: The engagement process in foreign language learning. *System, 86*, 102128. https://doi.org/10.1016/j.system.2019.102128

Pawlak, M., Derakhshan, A., Mehdizadeh, M., & Kruk, M. (2021). Boredom in online English language classes: Mediating variables and coping strategies. *Language Teaching Research*, 136216882110649. https://doi.org/10.1177/13621688211064944

Pawlak, M., Kruk, M., & Zawodniak, J. (2020). Investigating individual trajectories in experiencing boredom in the language classroom: The case of 11 Polish students of English. *Language Teaching Research, 26*(4), 598-616. https://doi.org/10.1177/1362168820914004

Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2020). Investigating factors responsible for boredom in English classes: The case of advanced learners. *System, 91*, 102259. https://doi.org/10.1016/j.system.2020.102259

Pawlak, M., Zawodniak, J., & Kruk, M. (2020). The neglected emo¬tion of boredom in teaching English to advanced learners. *International Journal of Applied Linguistics, 30*(3), 497–509. https:// doi.org/10.1111/ijal.12302

Pekrun, R. (2006). The Control-Value Theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Peng, J.-E., & Woodrow, L. (2010). Willingness to communicate in English: A model in the Chinese EFL classroom context. *Language Learning, 60*(4), 834–876. https://doi.org/10.1111/j.1467-9922.2010.00576.x

Pervin, L. A., & Cervone, D. (2010). *Personality: Theory and research. Hoboken*, NJ: John Wiley.

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics, 36*, 50–72. https://doi.org/10.1017/s0267190515000094

Piniel, K., & Albert, Á. (2018). Advanced learners' foreign language-related emotions across the four skills. *Studies in Second Language Learning and Teaching, 8*(1), 127–147. https://doi.org/10.14746/ssllt.2018.8.1.6

Prior, M. T. (2019). Elephants in the Room: An "affective turn," or just feeling our way? *The Modern Language Journal, 103*(2), 516–527. https://doi.org/10.1111/modl.12573

Putwain, D. W., & Pescod, M. (2018). Is reducing uncertain control the key to successful test anxiety intervention for secondary school students? Findings from a randomized control trial. *School Psychology Quarterly, 33*(2), 283–292. https://doi.org/10.1037/spq0000228

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*(3), 700–712. https://doi.org/10.1037/a0027268

Richards, J. C. (2022). Exploring Emotions in Language Teaching. *RELC Journal, 53*(1), 225–239. https://doi.org/10.1177/0033688220927531

Sadoughi, M., & Hejazi, S. Y. (2021). Teacher support and academic engagement among EFL learners: The role of positive academic emotions. *Studies in Educational Evaluation, 70*, 101060. https://doi.org/10.1016/j.stueduc.2021.101060

Shao, K., Pekrun, R., Marsh, H. W., & Loderer, K. (2020). Control-value appraisals, achievement emotions, and foreign language performance: A latent interaction analysis. *Learning and Instruction, 69*, 101356. https://doi.org/10.1016/j.learninstruc.2020.101356

Teimouri, Y. (2018). Differential Roles of shame and guilt in L2 learning: How bad is bad? T*he Modern Language Journal, 102*(4), 632–652. https://doi.org/10.1111/modl.12511

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis – erratum. *Studies in Second Language Acquisition, 41*(2), 489–489. http://dx.doi.org/10.1017/s0272263119000445

Wang, Y. (2023). Probing into the boredom of online instruction among Chinese English language teachers during the Covid-19 pandemic. *Current Psychology*. https://doi.org/10.1007/s12144-022-04223-3

Wang, X., & Li, Y. (2022). The predictive effects of foreign language enjoyment, anxiety, and boredom on general and domain-specific English achievement in online English classrooms. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1050226

Wei, H., Gao, K., & Wang, W. (2019) Understanding the relationship between grit and foreign language performance among middle school students: The roles of foreign language enjoyment and classroom environment. *Frontiers in Psychology, 10*, 1508. https://doi.org/10.3389/fpsyg.2019.01508

Zhao, Y. (2023). On the relationship between second language learners' grit, hope, and foreign language enjoyment. Heliyon, *9*(3), e13887. https://doi.org/10.1016/j.heliyon.2023.e13887

Zhao, X., & Wang, D. (2023). The role of enjoyment and boredom in shaping English language achievement among ethnic minority learners. J*ournal of Multilingual and Multicultural Development*, 1–13. https://doi.org/10.1080/01434632.2023.2194872

Appendix: FLA Questionnaire

To what extent do you agree with the following statements?

1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree, 0 = Not sure

Short Form of the Foreign Language Enjoyment Scale (S-FLES)
1. I enjoy my English class
2. I've learnt interesting things in my English class
3. In class, I feel proud of my accomplishments in my English class
4. The teacher is encouraging in my English class
5. The teacher is friendly in my English class
6. The teacher is supportive in my English class
7. We form a tight group in my English class
8. We have common "legends", such as running jokes in my English class
9. We laugh a lot in my English class

Short-form Foreign Language Classroom Anxiety Scale (S-FLCAS)
1. Even if I am well prepared for my English class, I feel anxious about it
2. I always feel that the other students speak English better than I do
3. I can feel my heart pounding when I'm going to be called on in my English class
4. I don't worry about making mistakes in my English class (reverse coded)
5. I feel confident when I speak in my English class (reverse coded)
6. I get nervous and confused when I am speaking in my English class
7. I start to panic when I have to speak without preparation in my English class
8. It embarrasses me to volunteer answers in my English class

Foreign Language Learning Boredom (FLLB)
1. I get bored in my English class
2. The lecture bores me in my English class
3. I think about what else I might be doing rather than sitting in this boring English class
4. I get restless because I can't wait for the English class to end
5. Studying for my English courses bores me
6. The material in my English class is so boring that I find myself daydreaming
7. I would rather put off this boring work for my English class till tomorrow
8. While studying for my English class I seem to drift off because it's so boring

Classroom Environment (CE)
1. Tasks designed in my English class are useful
2. Tasks designed in my English class are attracting
3. Activities in my English class are clearly and carefully planned
4. Class assignments are clear so everyone knows what to do in my English class
5. I work well with other class members in my English class
6. I make friends among students in my English class
7. I help other class members who are having trouble with their work in my English class
8. The teacher provides a timely response to students' concerns in my English class
9. The teacher is patient in teaching in my English class
10. The teacher asks questions that solicit viewpoints or opinions in my English class

Self-efficacy as a protective factor when translating under time pressure

Paula Cifuentes-Férez
Department of Translation and Interpreting
Universidad de Murcia, Spain
pualacf@um.es

Ana Rojo López
Department of Translation and Interpreting
Universidad de Murcia, Spain
anarojo@um.es

Laura Espín López
Department of Human and Psychobiological Anatomy
Universidad de Murcia, Spain
lespin@um.es

Abstract

This study explores the relationship between translation trainees' self-efficacy, their hormonal and subjective responses to the stress of translating under strict time limits, and how this impacts their actual translation performance. Participants completed a questionnaire on self-efficacy beliefs (Costa, Serrano, & Salvador, 2016), along with the State-and-Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988). They were also tasked with translating three literary texts from English to Spanish under varying time constraints and afterward provided feedback on their performance. Additionally, five cortisol salivary samples were collected during the session. Overall, the results indicate that self-efficacy beliefs act as a protective factor against stress, mitigating the negative effects of translating under time pressure. Interestingly, translation trainees with higher self-efficacy produced less accurate translations in terms of meaning under strict time pressure, in contrast to those with lower self-efficacy beliefs. Furthermore, elevated cortisol levels appeared to positively impact translation accuracy under the same conditions.

Keywords: time pressure, self-efficacy, anxiety, cortisol, translation quality.

Resumen

El objetivo principal de este trabajo es investigar las relaciones entre la autoeficacia de los estudiantes de traducción y las respuestas hormonales y subjetivas ante la situación de estrés que supone traducir bajo presión temporal y cómo estas pueden repercutir en su rendimiento en la traducción. Los participantes completaron un cuestionario en el que se les preguntaba por sus creencias de autoeficacia (Costa, Serrano, y Salvador, 2016) y otros dos cuestionarios autoinformados: el Inventario de Ansiedad Estado-Rasgo (Spielberger, Gorsuch, Lushene, Vagg, y Jacobs, 1983) y el Positive and Negative Affect Schedule (Watson, Clark, y Tellegen, 1988). Asimismo, se les pidió que tradujeran tres textos literarios comparables del inglés al español en diferentes condiciones de límite de tiempo. Al terminar, completaron un cuestionario sobre sus percepciones y sentimientos acerca de su desempeño. Se recogieron, además, cinco muestras salivales de cortisol en diferentes momentos de la sesión experimental. En general, nuestros resultados muestran que las creencias de autoeficacia son un factor protector contra el estrés que reduce el impacto negativo de traducir bajo presión temporal. En lo que respecta a la calidad de la traducción, los estudiantes de traducción con mayor autoeficacia parecen producir textos meta menos precisos en términos de significado en la condición con mayor presión temporal que aquellos estudiantes con creencias de autoeficacia más bajas. Asimismo, un mayor nivel de cortisol parece tener un efecto beneficioso sobre la precisión del texto meta en la misma condición.

Palabras clave: presión temporal, autoeficacia, ansiedad, cortisol, calidad de la traducción.

1. Introduction

The interest in studying what takes place in the translators and interpreters' minds began around 1960s, after the Cognitive Revolution in psychology which started to focus on learning, perception, memory and thinking from an experimental point of view (Miller, 2003). Since then, Cognitive Translation and Interpreting Studies (CTIS, henceforth) has witnessed a dramatic growth in the number of publications in edited volumes and scientific periodicals (Xiao & Muñoz, 2020).

Research in CTIS has focused on a wide range of issues, and many new avenues have emerged. Two decades ago, some scholars began to point out that certain differences found in the performance of translators might be due to personality traits (Jääskeläinen, 2000, p. 73) and that a trait such as tolerance of uncertainty, which is part of any cognitive decision-making process, deserves to be taken into account in translator training (Tirkkonen-Condit, 2000, p.141). Moreover, Bandura's (1977) concept of self-efficacy is also of great interest for CTIS (cf. Bolaños, 2012) as the

belief in one's own ability to succeed in a particular situation is a mediating variable in processes such as decision-making. Especially during the last few years, further research on translators' personality profiles (e. g., Abihssira, 2019; Bolaños-Medina, 2014; Bontempo & Napier, 2014; Hubscher-Davidson, 2016; Lehka-Paul, 2020), as well as on affective dimensions and emotions (e.g., Cifuentes-Férez & Fenollar-Cortes, 2017; Cifuentes-Férez & Meseguer, 2018; Lehr, 2014; Rojo López & Cifuentes-Férez, 2021) is breaking into this arena. Empirical evidence is also available for the impact of affective and personality variables of translators and interpreters upon translation and interpreting performance (cf. Lehka-Paul, 2020). However, much research is still necessary to gain insight into what role these traits play in the process (e.g., at different stages of the translation process, including reading the source text) and the product of translation (e.g., quality of the target texts, effects of personality on text types, etc.).

In the present paper we are mainly concerned with translator's perception of self-efficacy since it has been linked with numerous benefits, such as resilience to stress and improved performance (e.g., Hitches et al., 2023; Travis et al., 2020; Udayar et al., 2020). We aim at exploring the associations between translation trainees' perception of self-efficacy and their subjective and hormonal responses to the stressful situation of translating under strict time constraints. Moreover, we are also interested in investigating the potential impact of these variables on translation trainees' performance in terms of quality of the translated text. To this purpose, section 2 focuses on time pressure and reviews relevant work on time pressure within CTIS. Section 3 discusses the concept of self-efficacy, its sources and summarizes relevant literature on the effects of self-efficacy within CTIS. Section 4 introduces the study and section 5 summarizes the main conclusions and some avenues for future research.

2. Time pressure in translation

Time and deadlines are extremely important for translators (cf. Gouadec, 2017, p.5). When faced with tight deadlines or stringent time constraints for a prolonged period of time, translators run the risk of suffering from acute stress, which might affect their psychological and physical wellbeing in addition to their performance. Investigating the impact of time pressure on translators' psychological/emotional and physiological responses might be of interest in order to gain a better understanding of the translation process as well as to raise translators' awareness of its impact on their emotional and physical wellbeing. The term time pressure is used in the present paper to refer to the psychological reaction that individuals have when they believe the amount of available time is less than the amount they perceive they need to perform a translation task (cf. Kleiner, 2914; Ordóñez, Benson, & Pittarello, 2015), whereas time constraint is used when individuals face a time limit imposed externally when performing a translation task.

Most of the studies on the effects of time pressure which have used physiological and self-reported measures have focused on interpreting (e.g., Korpal, 2016; Korpal & Jankoviak, 2021; Rojo López, Foulquié-Rubio, Espín López, & Martínez Sánchez, 2021) whereas scarce research attention has been paid on written translation. Over the last two decades, translation research, mainly through key-logging and eye-tracking instruments, has investigated its role on the different translation stages (Jensen 1999, 2000), translators' fixations on source and target texts (Sharmin et al., 2008), and written translation product quality (De Rooze, 2003; Ghobadi et al., 2017; Kourouni, 2012), results up to date being far from conclusive.

In most of the studies on time pressure in translation, time pressure is generally induced by objectively constraining the time frame for a translation task, but other supplementary subjective time-pressure manipulation strategies are available, such as giving instructions about the time frame before starting the task or visualizing the elapse of time (Weng & Zheng, 2020). Thus, it is crucial to consider how time pressure is induced and how it can be measured in the studies.

Bayer-Hohenwarter (2009) was, to our knowledge, the first to explore the methods used in the experimental studies on time pressure in translation. She distinguishes three time-pressure measurement approaches: subjective definition/rating, which include the use of self-report measures such as the State-Trait Anxiety Inventory (STAI) or retrospective questionnaires; pragmatic objectivation (e.g., signs of stress on the face and in body language, that is, signs of the stressful state of the participant); and physiological objectivation, that is, the use of biomarkers, such as stress hormones, blood analysis, etc., to measure the biological reactions to stress induced by time pressure. Despite the fact that according to Bayer-Hohenwarter (2009), physiological markers might be the most reliable for studying time pressure due to their objective character, she argues in favour of combining self-report measures with physiological measures in the study of time pressure since psychological and emotional factors help the researcher to reach conclusive results.

Jensen (1999) examined by means of key-logging the translation process of professionals, non-professionals and young translators, who have to translate four texts with 10, 15, 20 and 30-minute time constraints. The only significant effect of time pressure was observed when more time was available; in this case, two or more rounds of trying to solve problems were found where previous problem-solving time had already been devoted (Jensen & Jakobsen, 2000: 9). Much in the same line, De Rooze (2003) used keystroke logging but instead of focusing on coping tactics during the translation process he paid special attention to stress caused by time pressure and how it affected translation performance. In his research, participants were asked to perform a 10-minute heating task, then asked to translate a text in 15 minutes, and

after that, another text in 10 minutes. De Rooze, unlike Jensen (1999), started with the less stressful translation and then moved onto the text to be translated under stringent time pressure. He found that (a) when asked to translate more than 200 words within 10 minutes, translation quality lowers or decreases more than 15%; (b) there was a tendency to make mistakes in the target text just after making one; and (c) 25% of the participants produced higher quality texts under time pressure. In line with these findings, the results of the study by Ghobadi, Madadi, and Najafian (2017) showed that time pressure had a significant impact on both the quality and quantity of the translation task carried out by the participants, namely, participants in the time pressure group produced more translated materials but their quality lagged behind those of the texts by participants in the control group (no time pressure). However, Kourouni (2012) did not find any statistically significant difference in the overall translation quality for the 30-minute, 20-minute and 10-minute tasks.

Other researchers have used eye-tracking to provide an insight of the effects of time pressure. Sharmin, Špakov, Räihä, and Jakobsen (2008) included text complexity as another variable (using Flesh-Kincaid reading scores) and participants were given 6, 5 and 4 minutes to translate each text. Their findings reveal that time pressure was found to affect fixations, more concretely, in the source text, suggesting that translators can adapt their reading-for-comprehension to variable time constraints, whereas it is more difficult for them to adapt their reading-and-monitoring of the target text (p. 126).

On the whole, these studies provide evidence for the higher occurrence of errors under stringent time conditions, but it is also noted that having enough time does not always guarantee less errors (e.g., Lorenzo, 2002; Künzli, 2007) and that some translators seem to work much better under time pressure (e.g., De Rooze, 2003; Khalzanova, 2008), pointing out at the intervening role of individual differences and personality traits. Recently, Rojo López, Cifuentes-Férez, and Espín López (2021) found that two personality traits seem to play an important role on both the translation process and product: self-esteem as a protective factor against stress and trait anxiety as a predictor of higher accuracy. Despite the fact that self-esteem was found to be a protective factor against stress produced by translating under time pressure, it seems to have a negative effect on target texts since translation trainees with higher self-esteem produced less accurate translations under extreme time pressure in the attempt to translate more words, as also noted by Ghobadi, Madadi, and Najfian, 2017. Moreover, Rojo López, Cifuentes-Férez, and Espín López (2021) noted a decreasing pattern for cortisol levels which pointed to the effect of trainees' attentional response to the translation task (cf. Rojo López & Naranjo Sánchez, 2021; Rojo López, Ramos Caro, & Espín López, 2021).

Drawing from Rojo López, Cifuentes-Férez, and Espín López (2021), in the study presented in Section 4, we use self-report measures and salivary cortisol to measure

the physiological effects of time pressure in written translation and their connection to translation trainees' self-efficacy. Before delving into that section, a critical literature review on self-efficacy beliefs and self-efficacy in CTIS is provided.

3. Self-efficacy beliefs

It is well-known that self-efficacy mitigates the detrimental effects of stress (cf. Fida et al., 2015). Despite this, the assessment of self-efficacy during translation training has been rather overlooked in translation research with a few exceptions (e.g., Bolaños-Medina, 2014; Bolaños-Medina, 2017, 2018; Haro-Soler, 2018, 2022; Jiménez Ivars et al., 2014; Konttinen, 2021; Yang et al., 2021).

The concept of self-efficacy stems from Bandura's Social Cognitive Theory (1986, 1987, 1997), which proposes that cognitive and behavioural aspects of an individual interacts with the environment in a two-way process. In his own words, self-efficacy consists of the "beliefs in one's capacity to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). Self-efficacy beliefs can have an effect on decision-making processes, motivation and goal-setting, the effort and persistence when performing a certain task in the face of difficulties or failure, as well as on emotional states, such as anxiety, which can negatively affect problem solving (Bandura, 1986, 1987, 1997). Moreover, Bandura (1977) stated that self-efficacy can influence an individual's level of anxiety in relation to the task to be accomplished. This means that individuals with low self-efficacy will experience higher anxiety whereas those with high self-efficacy will experience lower anxiety.

Bandura (1994, 2006) describes how self-efficacy is the perception of competence rather than actual performance, a distinction that greatly impacts an individual's course of action since it influences the way a person thinks, feels and acts. The construct of self-efficacy should be distinguished from other self-perception concepts (Haro-Soler, 2018; Pajares, 1997), such as self-esteem, which refers to a person's perception of his or her own worth, and self-confidence whose focus is on the perception of one's capabilities in general. Moreover, Bandura (1997) argues that self-efficacy is more or less durable but does not require stability over time. Self-efficacy, thus, could be modified through teaching intervention or training since individuals could be trained to engage in cognitive appraisal to interpret their environment (e.g., translation and interpreting situations) in constructive ways to build up their self-efficacy beliefs (Atkinson, 2014).

One's self-efficacy can generate from four different sources: enactive mastery experience, vicarious experience, verbal persuasion, and physiological and emotional states (Bandura, 1997). First, mastery experience relates to one's self-perceived ability to (un)successfully perform a task based on previous achievements or failures (Bandura,

1997, p. 80; Zhang & Ardasheva, 2019). Second, vicarious experience is linked to the impact that others' performance has on oneself; in other words, by observing a successful or unsuccessful performance of other individuals with similar abilities, one can strengthen or undermine his or her self-efficacy beliefs. Third, verbal persuasion refers to the feedback for performance on a given task and it is the source of self-efficacy most commonly used by teachers, with positive constructive comments enhancing self-efficacy and negative comments undermining it (Haro Soler, 2017, 2022; Torre, 2007). Last, physiological and emotional states refer to the individual's ability to manage their emotions (e.g., stress, anxiety) during performance. According to Bandura (1997), these four sources show how one's beliefs in his or her self-efficacy can affect performance.

In relation to the effects of self-efficacy on performance and behaviour, it has been noted that individuals with low levels of self-efficacy tend to avoid performing activities which they feel unprepared for, preventing them from having the possibility of modifying the beliefs that they are not capable of coping with (Bandura, 1986, 1997). In contrast, individuals who are more confident in their abilities to perform a particular task invest more effort, persevere longer in the face of challenges, and are more resilient to failures than those with lower self-efficacy (Bandura, 1995, 1997; Zeldin & Pajares, 2000). Moreover, it has also been pointed out that when a task is unchallenging, self-efficacy has less effect on performance, whereas, when a task is challenging and attainable, self-efficacy has a positive effect on performance (e.g., Beattie, Fakehy, & Woodman, 2014).

Psychological research on the interplay between self-efficacy and performance in different fields has also examined affect, anxiety, and stress. As far as affect is concerned, it has been shown that affect can influence persistence and success on a task, with positive affect leading to higher performance and self-efficacy (e.g., Kavanagh & Bower, 1985; Kavanagh & Hausfeld, 1986; Thelwell, Lane, & Weston, 2007). Moreover, there is empirical evidence indicating that teaching intervention can reduce anxiety and negative affect, increasing in turn positive affect and self-efficacy (e.g., Randler et al., 2016). In musical performing, studies indicate that lower levels of self-efficacy are related to higher music performance anxiety (MPA) in both in adult and young musicians (e.g., Hendricks, Smith, & Legutki 2015; Orejudo et al. 2017). MacAfee and Comeau (2020) found that MPA can have performance-enhancing and performance-impairing effects on self-efficacy, but no association was found between MPA and behavioural anxiety, indicating that music students might appear less anxious than they actually feel. Moreover, research has suggested that individuals have generally an optimal level of anxiety that favours performance (Hanin, 2000; Mor et al., 1995). In sports contexts, Costa, Serrano, and Salvador (2016) found that women with high levels of self-efficacy have better performance, better positive mood, and lower anxiety, suggesting that self-efficacy influences the emotional experience

of the situation. In their study, salivary cortisol (i.e., a steroid hormone secreted into salive when an individual is under stress) is also measured. Their results suggested that cortisol response did not depend on self-efficacy beliefs, thus, contradicting Bandura's (1997) claim that individuals with low self-efficacy would have greater stress responses. However, other studies have found significant relationships between cortisol and self-efficacy, pointing out the importance of this stress hormone to better understand the entire process (Suay et al., 1999). Therefore, cortisol is not only a stress marker that is used to explain the response to competition, but also another variable affecting performance since it depends on more subjective factors (Lautenbach & Laborde, 2016; Costa, Serrano, & Salvador, 2016).

Within translation studies, some research attempts have been made to include the study of self-efficacy as translation is considered a higher-order cognitive process in which cognitive, affective and emotional aspects are to be taken into consideration (e.g., Angelone, 2010; Hansen, 2010; Shreve & Lacruz, 2014). Self-efficacy is considered to be included in the concept of self-concept (Kiraly, 1995; Muñoz Martín, 2014). The latter pertains to "a sense of the purpose of the translation, an awareness of the information requirements of the translation task, a self-evaluation of the capability to fulfil the task, and a related capacity to monitor and evaluate translation products for adequacy and appropriateness" (Kiraly, 1995, p. 100); whereas the former, self-efficacy in translation, could be defined as the confidence that translators have in their abilities to perform well or fulfil in translation tasks (Haro-Soler & Kiraly, 2019). This is the definition we are adopting in this piece of research.

Pioneering research on self-confidence (that was used as synonym of self-efficacy in Kussmaul (1995)) in the early 1990s indicated that it is a prerequisite for creative translation and that adequate solutions seems to be subsequently lost in the target text due to the translator's insecurity (Kussmaul, 1995). Furthermore, self-efficacy and translation quality seem to be positively related (Tirkkonen-Condit & Laukkanen, 1996) with translators with higher self-efficacy producing better target texts in terms of quality.

More recent research on self-efficacy overall suggests that translators' self-efficacy is associated with their performance, motivations as well as with their competences and/or skills, professional success and job satisfaction (Albin, 2012; Araghian et al., 2018; Atkinson 2012, 2014; Bolaños-Medina, 2014; Haro-Soler, 2018, 2019a). Atkinson (2012) focused on freelance translators' psychological skills. According to his model, psychological skill consists of three components: self-efficacy, attribution style and locus of control (internal and external). Attribution style refers to the way an individual explain behaviour or past events, in other words, what attributions a person makes for the reason of behaviour (i.e., hard work, luck, task difficulty, etc.). Locus of

control is related to attribution style as it also focuses on the perceptions of causation, but the difference lies on the fact that locus of control and self-efficacy centre on future expectations for performance whereas attribution style on explanations of past performance. His results overall suggested that self-efficacy relates to motivation, job satisfaction and posterior job performance. Albin (2012) found that high self-efficacy and a favourable attribution style (i.e., attribution of success to internal and stable causes, such as one's capabilities) correlates with high level management skills and high levels of performance on the use of CAT tools. Moreover, her data indicated that translators with high self-efficacy seem to evaluate themselves against money and prestige criteria. Bolaños-Medina (2014. p. 212) found that self-efficacy correlated positively with tolerance of ambiguity, perceptions of meeting the needed requirements to become a professional translator, source language reading comprehension, the ability to find background documentary information and to be aware of when to stop searching for a solution for a translation problem. Much in the same vein, Araghian et al. (2018) concluded that translation trainees with lower self-efficacy spend too much time translating because of their repeated attempts at production and exhaustive revision.

The growing interest in self-efficacy in translation has led to scholars to design scales for measuring self-efficacy. By way of illustration, self-efficacy scales have been designed by Lee (2014) for consecutive interpreting and by Bolaños-Medina and Núñez (2018), Haro-Soler (2018, 2022), Yang et al. (2021) and Kottinen (2021) for translation. The development and validation of these scales has narrowed the gap in the assessment of self-efficacy in translation and interpreting; however, they are targeted to specific audiences depending on their scope (e.g., measuring undergraduates translating self-efficacy, self-efficacy in translation service provision, etc.) with specific translation directions and include, overall, a large number of items. We did not opt for these scales so as to avoid participants' fatigue from spending too much time on filling up questionnaires. We, therefore, measure situational self-efficacy beliefs following Costa, Serrano, and Salvador (2016) as it is a light-weight measuring tool consisting of just three items referring to the capacity, confidence and importance of successfully performing the translation task on a 1-to-5 Likert scale. Accordingly, self-efficacy in relation to the translation task in which they were going to take part was operationalized as the mean of these three items.

4. The study

4.1. Aim and hypotheses

Our main aim is to study the impact of perceived self-efficacy on the hormonal (i.e., salivary cortisol response) and subjective emotional (i.e., anxiety, positive and

negative affective states) responses to translating under time pressure, as well as on translation performance in terms of quality and number of translated words.

Based on findings from existing research, we posed the following hypotheses to fulfil our aim:

1. When translating under time pressure, trainees' cortisol response will be higher and their performance will be worse than when translating under no time pressure.
2. Higher levels of self-efficacy will be associated with trainees' lower cortisol levels and state anxiety responses under time pressure (i.e., Text 2 and Text 3).
3. After finishing the translation tasks, positive affect scores will be lower and negative affect and state anxiety scores will be higher.
4. Higher levels of self-efficacy will be associated with better performance in the translation tasks under time constraints (i.e., Text 2 and Text 3).
5. Higher levels of self-efficacy will be associated with trainees' attribution of results to their capacity and effort.

4.2. Participants

After completing a general health questionnaire, 25 female translation trainees at the University of Murcia (Spain) were selected to participate in the study. They did not have any medical or psychological problem. Besides, we ensured they were not taking the contraceptive pill since it has been demonstrated to increase cortisol levels (e.g., Nielsen et. al, 2013) and could, thus, affect our results. All of them had Spanish as mother tongue and English as second language, and their age ranged from 19 to 20 years (M = 19.32 years; SD = .47). The main characteristics of the sample are shown in Table 1.

Table 1: Descriptive characteristics of the sample (N=25)

Variables	**Minimum**	**Maximum**	**Mean**	**SD**[c]
Age	19	20	19.32	0.47
BMI[a]	16.65	24.72	21.11	2.31
STAI Trait[b]	8	53	28.76	10.93

[a]Body Mass Index.
[b]Trait anxiety.
[c]Standard deviation.

Participants granted their consent according to the Declaration of Helsinki, and the study protocols were approved by the Ethics Committee at the University of Murcia.

Participants were informed of the general purpose of the study and were told that they could leave the experiment at any point.

4.3. Materials

4.3.1. Source texts

We used the materials and procedure followed in Rojo López, Cifuentes-Férez, and Espín López (2021). The source texts were three English literary texts of similar difficulty according to the Flesch-Kincaid Reading Ease and Grade Level, the Gumming Fog Score, the SMOG Index, and the Coleman Liau Index (see Table 2). They were taken from the novel *The Ballroom* (2016) by Ann Hope. The three texts were descriptive passages in which dialogue was absent. The first text (Text 1) contained 150 words, whereas the second (Text 2) and the third (Text 3), 153 words each. We opted for texts shorter than 200 words for two reasons, namely, (1) participants were in their second year of their degree, so they translate at a slower pace than more advanced students and we wanted to prevent them from fatigue since all of them were requested to translate the three texts; and (2) to differentiate between a time pressure condition (Text 2) where the task was affordable within the given 10 minute time limit (cf. Rojo López, Cifuentes-Férez, & Espín López, 2021) and another condition (Text 3) where the 5 minute time limit would make trainees aware of the impossibility to accomplish this translation task.

Table 2: Scores of text difficulty, grade conversion and comprehension of the three texts

	Text 1	Text 2	Text 3	Grade Conversion - Comprehension
Flesch Kincaid Reading Ease	92.7	100.4	96.4	5th Grade - Very easy to read
Flesch Kincaid Grade Level	3	2.5	2.3	5th Grade - Very easy to read
Gunning Fog Score	5.6	4.1	4.9	5th Grade and below - Very easy to read
SMOG Index	3.9	2.7	3.2	5th Grade - Very easy to read
Coleman Liau Index	9.6	7.3	7.6	8th, 9th & 10th Grade - Conversational English

In order to assess translators in training's performance in each translation task, the accuracy of translated texts was assessed in terms of number of errors (see Table

3). The evaluation sheet was adapted from the one designed for the TRANSCREA research project (cf. Rojo López, 2019). We subtracted from 0.25 to 1 point for each error from a total score of 10 points. Accuracy was assessed by marking errors on three different categories: (1) transfer of meaning (i.e., false and opposite meaning, and unnecessary omissions or additions); (2) transfer of pragmatic function (including mainly loss of literary style, cultural references, implied meaning, humour or irony); and (3) correctness (i.e., grammatical errors, errors in the cohesion of the text, typos and punctuation and spelling errors).

Table 3: Evaluation sheet for the assessment of target texts

Transfer of meaning	
False meaning / Not the same meaning	– 0.5
Opposite meaning / Incoherent meaning	– 1
Unnecessary omission / addition of meaning	– 0.5
Transfer of pragmatic function	
Loss of cultural reference and/or implied meaning	– 1
Loss of humor or irony	– 1
Correctness	
Grammatical errors	– 1
Cohesion errors (connectors, loss of repetition)	– 0.5
Orthotypographic errors	
Typos	– 0.25
Written accents and punctuation marks	– 0.5
Serious spelling mistakes	– 1

We also considered the number of translated words across the different translation tasks by counting the number of words translated by each participant in each source text. Because of the time constraints imposed on participants, all of them completed the translation of Text 1, 17 did not finish Text 2 and none did it for Text 3. Therefore, a correction index was applied to the accuracy scores for these two texts in order to increase comparability among the error categories across the three different texts. The correction index consisted of dividing the score for each type of error by the number of translated words and multiplying the result by the total number of words from the source text (153 words).

4.3.2. Measures and instruments

Self-efficacy was measured following Costa, Serrano, and Salvador (2016). We used a brief questionnaire that consisted of just three items, referring to the capacity, confidence and importance of successfully performing a task (in this case, three translation tasks) on a Likert-type scale ranging from 1 (none) to 5 (a lot). Therefore, their self-efficacy beliefs in relation to the translation tasks in which they were going to take part was operationalized as the mean of the ratings for these three items.

Anxiety was measured by the State-Trait Anxiety Inventory (STAI) (Seisdedos, 1998; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). This is a 40-item self-report inventory that measured participants' levels of state anxiety (STAI-S) and trait anxiety (STAI-T) on a 4-point Likert scale, ranging from 1 (almost never) to 4 (almost always).

Positive and negative affect was measured by the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988). This is a 20-item self-reported questionnaire in which 10 items measure positive affect and 10 negative affect on a 5-point Likert scale, ranging from 1 (nothing) to 5 (very much). Scores can range between 10 and 50 and the higher the scores, the higher levels of positive or negative affect.

Salivary cortisol was gathered employing the Salivette® collection apparatus (Sarstedt, Newton, NC). Over the course of an hour, five saliva samples were procured at distinct temporal junctures relative to the initiation of the experimental task (referred to as sample t0): t-20 (baseline, 20 minutes before task commencement), t0 (commencement of the experimental task), t+20 (20 minutes post-task onset), t+35, and t+45. Participants were instructed to place the cotton swab in their mouths for a duration of 2 minutes, refraining from chewing it to prevent potential alterations in salivary protein composition and flow rate (Bosch, Veerman, de Geus, & Proctor, 2011). They were further directed to maneuver the swab in a circular motion to collect saliva from all salivary glands (Rohleder & Nater, 2009). The uncentrifuged saliva specimens were promptly preserved at −80 °C until subsequent analyses. In order to minimize potential sources of variation, all five samples from each participant were subjected to analysis within a unified assay. The samples underwent evaluation via a competitive solid phase radioimmunoassay (tube coated) utilizing the commercial kit Coat-A-Count Cort (DPC, Siemens Medical Solutions Diagnostics). The assay exhibited a sensitivity of 0.5 ng/ml. Cortisol levels were denoted in nmol/l, featuring intra- and inter-assay variation coefficients of less than 10%.

Attribution style was measured by a 4-item self-reported questionnaire on a 5-point Likert scale in which participants were asked to rate factors that may have contributed

in their results, namely, capabilities, bad/good luck, effort invested, and task difficulty (Espín López, 2009). After completing this questionnaire, participants' perception and feelings in relation to effort invested, level of frustration, accomplishment of the tasks, tasks difficulty, importance of doing well, and marker assessment of the tasks on a 5-point Likert scale were also measured.

4.4. Procedure

Participants who met the criteria of not having any physical or mental health problems, and of not being on the pill were contacted and asked to attend the experimental session on an agreed date. Participants were told to maintain their general habits, sleep as long as usual, refrain from heavy physical activity the day before the session, and not to consume alcohol after the previous dinner. Instead, they were requested to drink only water and not to eat chocolate or take coffee, cola, tea in the two hours prior to the session.

Participants entered the experimental room individually and were tested in a single session. Each experimental session lasted approximately 1 hour and 10 minutes for each participant and was held between 2 pm and 6 pm. Once inside, the researcher asked the participants whether they had followed the instructions previously given. They were asked to read the participant information form which informed them that the experiment involved filling up some psychological tests and questionnaires, and translating three texts under different time constraints. They were also informed that five salivary samples would be collected at different times throughout the experimental session and explained how to use the salivette for the own collection of salivary cortisol.

Afterwards, the first sample of salivary cortisol was taken and they completed the self-efficacy three-item questionnaire, the STAI (both trait anxiety and state anxiety), and the PANAS. Once finished, the second salivary cortisol sample was taken just before the start of the translation task. No time limit was given for Text 1, but participants overall spent a maximum of 20 minutes translating this text. On completion, the third cortisol sample was taken. They were then given 10 minutes to translate Text 2. When done, they were given 5 minutes to translate Text 3. A visible countdown timer was displayed in the room, but participants were also told to display one on their own computer screen to keep track of time, as an additional manipulation for time pressure inducement and intensification (cf. Weng & Zheng, 2020). Participants were allowed to use any online documentation resource they wished during the translation task in order to maximize ecological validity. The fourth salivary sample was collected on completion of the translation of Text 3. Next, participants were asked to complete again the STAI-state questionnaire and the PANAS, and were also requested to fill in another brief questionnaire in which they were asked about the factors that may have

contributed in their results (i.e., capabilities, bad/good luck, effort invested, and task difficulty); and about their perception and feelings in relation to effort invested, level of frustration, accomplishment of the tasks, tasks difficulty, importance of doing well, and marker assessment of the tasks. Finally, the fifth salivary sample was taken. Then, they were thanked for the participation and debriefed.

4.5. Results

4.5.1. Data analysis

Salivary cortisol was tested for normal distribution and homogeneity of variance using the Shapiro–Wilk test before the statistical procedures were applied. These analyses revealed significant deviations from normality and were transformed to logarithm.

To assess the cortisol response across the different phases of the protocol as well as the response in PANAS and STAI-S before and after the translation tasks, we conducted separate repeated measures analyses of variance (ANOVAs) with time as a within-subjects factor (five phases for salivary cortisol: t-20, t0, t+20, t+35 and t+45) and two phases for PANAS and STAI-S (pre- and post-task).

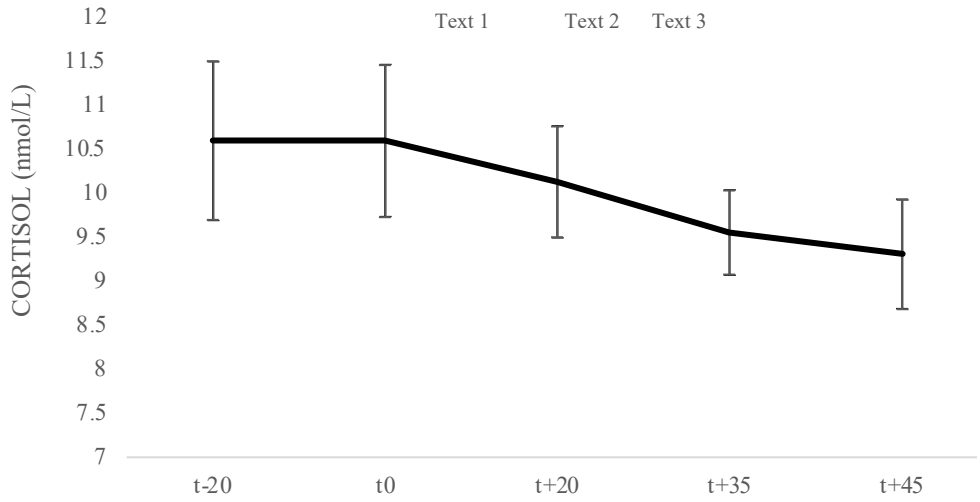
To examine whether cortisol levels could be associated with translation performance scores for the three texts, we conducted Pearson's bivariate correlation analyses between the variables. Moreover, Pearson's bivariate correlation analyses were carried out to test the relationship between self-efficacy, trait-state anxiety, positive and negative affect, and the five cortisol samples. They were also run to examine whether levels of self-efficacy, anxiety, affect and stress-induced cortisol were related to performance scores in the translation tasks. These correlations were conducted for each type of error scale (the total accuracy scale, i.e., the mean score obtained from the three subscales: meaning, pragmatic and correctness errors; and each of the three subscales separately) and for each text (Text 1, Text 2, Text 3). The number of words translated in each text was also included as a performance score.

4.5.2. Results

Salivary cortisol response

A repeated-measures ANOVA was conducted with time (5) as within-subject factor to test differences in salivary cortisol between the different phases. The results did not show a significant main effect for this factor [$F (4; 96) = 1.19$, $p= 0.32$, $\eta 2p = 0.05$], revealing no statistically significant differences in participant's cortisol levels between the different phases of the experimental task (see Figure 1).

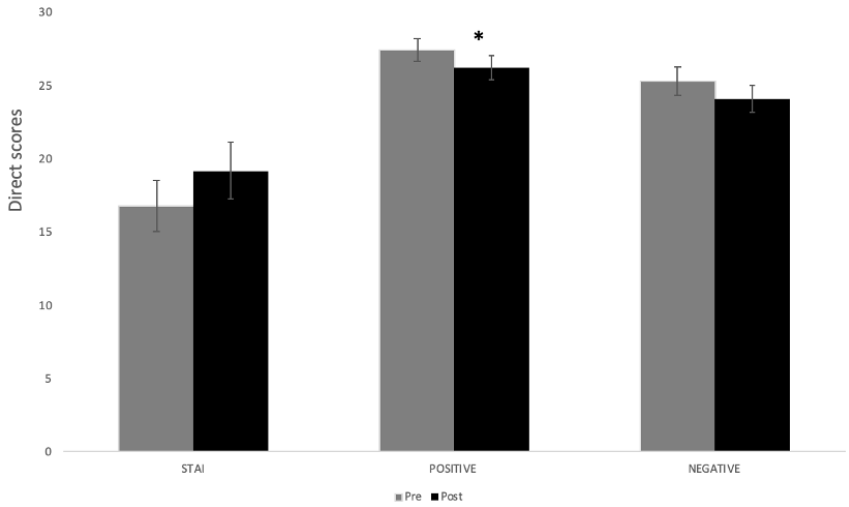
Figure 1: Variation in cortisol response during the experimental session



PANAS (Negative and Positive Affect; NA and PA) AND STAI-S

A repeated-measures ANOVA was conducted with time (2) as within-subject factor to test differences in affect and state anxiety before and after the experimental task. The results did not show a significant main effect for the negative affect [F (1; 24) = 2.15, p= 0.15, η2p = 0.08], but significant results were found for positive affect, [F (1; 24) = 4.14, p= 0.05, η2p = 0.14], with lower scores for positive affect in the post-task as compared with the pre-task (M pre=27.44; M post=26.20. Results for state anxiety did not show a significant main effect [F (1; 24) = 3.47, p= 0.07, η2p = 0.12] (see Figure 2).

Figure 2: Pre- and Post of PANAS (positive and negative affect) and STAI-State



To examine whether the translation performance scores for the three texts was associated with cortisol levels, we conducted Pearson's bivariate correlation analyses between the variables. The results showed a significant negative correlation between t0 and the scores obtained on accuracy (r=-.50**) and meaning (r=-.52**) in Text 1. In addition, we found a positive correlation between t+45 and the scores obtained on correctness (r=.40*) for Text 3 (see Table 4).

Table 4: Correlations between performance of each text and cortisol levels

Type of text		Cortisol				
		t-20	t0	t+20	t+35	t+45
TEXT 1	**Readability**	-.14	.07	.18	-.01	.24
	Total Accuracy	-.26	-.50**	-.35	-.32	-.30
	Meaning	-.34	-.52**	-.35	-.28	-.27
	Pragmatics	.13	.03	-.08	-.15	-.10
	Correctness	.10	-.12	.03	-.02	-.05
	Translated words	-	-	-	-	-
TEXT 2	**Readability**	-	-	-	-	-
	Total Accuracy	.01	-.13	-.06	-.03	.09
	Meaning	-.07	-.09	-.05	-.02	.18
	Pragmatics	.09	-.05	-.005	.06	-.15
	Correctness	.001	-.15	-.21	-.32	-.08
	Translated words	-.27	-.39	-.14	.04	-.10
TEXT 3	**Readability**	-	-	-	-	-
	Total Accuracy	.35	.23	.22	.21	.30
	Meaning	.35	.22	.14	.34	.31
	Pragmatics	.11	.04	.08	-.09	-.17
	Correctness	.11	.12	.18	.03	.40*
	Translated words	-.24	-.30	-.19	.04	-.009

Table 5: Pearson coefficients for associations between self-efficacy and cortisol levels, state-anxiety (STAI) and Negative and Positive Affect (PANAS)

	1	2	3	4	5	6	7	8	9	10	11	12
Self-efficacy	-	-.14	-.37	-.16	-.31	-.52**	-.53**	-.39	-.21	-.22	.18	.07
T-20	-.13	-	.80**	.42*	.41*	.23	.08	.35	.03	.30	-.02	.08
T0	-.37	.80**	-	.67**	.67**	.53**	.20	.26	.21	.39	-.02	.08
T+20	-.16	.42*	.67**	-	.76**	.55**	.13	.13	.27	.20	.11	-.05
T+35	-.32	.41*	.67**	.76**	-	.70**	.02	-.05	.13	.08	.03	-.12
T+45	-.52**	.23	.53**	.56**	.71**	-	.19	.17	.12	.02	-.30	-.34
PR-STAI-S	-.53**	.08	.20	.13	.02	.19	-	.75**	.23	.36	-.25	-.20
PT-STAI-S	-.39	.35	.26	.13	-.05	.17	.75**	-	.18	.40*	-.31	-.09
PR-PANAS-N	-.21	.03	.21	.27	.13	.12	.23	.18	-	.62**	.27	.10
PT-PANAS-N	-.21	.30	.39	.20	.08	.02	.36	.40*	.62**	-	.22	.43*
PR-PANAS-P	.18	-.02	0.2	.11	.03	-.30	-.25	-.31	.27	.22	-	.72**
PT-PANAS-P	.07	.08	.12	.004	.12	.34	.20	.09	.10	.44*	.72**	-

Note: 2 to 6: cortisol samples; PR-STAI-S: state anxiety pre; PT-STATI-S: state anxiety post; STAI-S-POST: state anxiety post; PR-PANAS-N: negative affect pre; PT-PANAS-N: negative affect post; PR-PANAS-P: positive affect pre; PT-PANAS-P: positive affect post.
* Significant at p<.05 level; ** Significant at p=.01 level

Relationship between self-efficacy, cortisol levels, state-anxiety (STAI) and Negative and Positive Affect (PANAS) during the experimental task.

We only found a significant negative relationship between self-efficacy and cortisol levels at t + 45 (r = -. 52 **) and pre-state anxiety levels (r = -. 53 **)

To examine whether self-efficacy and affect before and after the task could be associated with translation performance scores for the three texts, we conducted Pearson's bivariate correlation analyses between the variables. The results showed a significant positive correlation between negative affect pre-task with the score for meaning in Text 2; a significant negative correlation between positive affect before and after the task with the score for correctness in Text 3 and a significant negative correlation between self-efficacy and the score for accuracy in meaning in Text 3 (see Table 6).

Table 6: Correlations between Negative and Positive affect, Self-efficacy and translation performance for each type of text

Type of text		PANAS				SELF-EFFICACY
		NA pre	NA post	PA pre	PA post	
TEXT 1	Total Accuracy	-.22	-.12	.02	-.10	.18
	Meaning	-.05	-.02	.13	-.04	.16
	Pragmatics	-.28	-.13	-.20	-.11	-.03
	Correctness	-.39	-.32	-.14	-.13	.18
	Translated words	-	-	-	-	-
TEXT 2	Total Accuracy	-.22	-.12	.02	-.10	.18
	Meaning	-.05	-.02	.13	-.04	.16
	Pragmatics	-.28	-.13	-.20	-.11	-.03
	Correctness	-.39	-.32	-.14	-.13	.18
	Translated words	-	-	-	-	-
TEXT 3	Total Accuracy	-.20	-.12	-.02	-.08	-.39
	Meaning	-.27	-.08	-.03	.03	-.41*
	Pragmatics	-.06	-.14	.26	.12	-.22
	Correctness	.06	-.001	-.43*	-.42*	.02
	Translated words	-.04	.02	.02	-.12	.23

Note: NA: negative affect; PA: positive affect

* Significant at p<.05 level; ** Significant at p=.01 level

Result attribution and task perception

The scores obtained in the test on result attribution (see Figure 3) showed that the subjects attributed the result in the task mostly to their capacity (M=4.28), effort (M=3.88), and its difficulty (M=3.76) and less due to luck (M=1.92).

Figure 3: Mean scores on result attribution questionnaire for the four scales

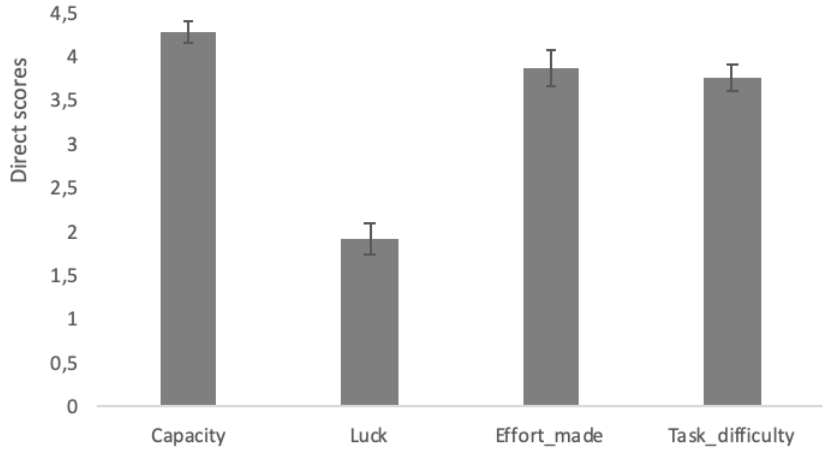


Table 7: Correlations between self-efficacy and questionnaire on result attribution and perceptions about the task

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.Self-efficacy	-	-.06	-.46*	.43*	-.33	.13	-.21	.41*	.05	-.27	.10	-.31	.45*	.28
2.Effort	-.06	-	.39*	-.17	.39	.59**	.22	-.11	-.003	.30	.11	.11	-.31	-.15
3.Frustation	-.46*	.39*	-	-.82**	.70**	.48*	.44*	-.75**	-.28	.47*	.07	.15	-.43*	-.35
4.Achievement	.43	-.17	-.82**	-	-.51**	-.38	-.35	.69**	.22	-.37	-.15	-.03	.20	.25
5.Stressful task	-.33	.39	.70**	-.51**	-	.50**	.19	-.36	-.04	.23	-.05	.11	-.42*	-.51**
6.Difficulty	.13	.59**	.48*	-.38	.50**	-	.35	-.25	-.06	.03	.30	.02	-.14	-.10
7.Importance	-.21	.22	.44*	-.35	.19	.35	-	-.54**	-.22	.20	.08	-.11	-.19	-.13
8.Marker-assessment	.41*	-.11	-.75**	.69**	-.36	-.25	-.54**	-	.30	-.50**	-.10	-.30	.23	.21
9.Results	.05	-.003	-.28	.22	-.04	-.06	-.22	.30	-	-.29	.39	.07	.15	.08
10.Achievement expectations	-.27	.30	.47*	-.37	.23	.03	.20	-.50**	-.29	-	-.34	.17	-.15	-.42*
11.Capacity	.10	.11	.07	-.15	-.05	.30	.08	-.10	.39	-.34	-	.12	.24	.32
12. Luck	-.31	.12	.15	-.03	.11	.02	-.11	-.30	.07	.17	.12	-	-.28	.28
13.Effort made	.45*	-.31	-.43*	.20	-.42*	-.14	-.19	.23	.15	-.15	.24	-.28	-	.06
14.Task difficulty	.28	-.15	-.35	.25	-.51**	-.10	-.13	.21	.08	-.42*	.32	.28	.06	-

As shown in Table 7, correlations were made with the self-efficacy score and the scores for translation trainees' result attribution in addition to their perceptions about the task. Correlation analyses yielded that self-efficacy was positively related to scores on perceived task achievement (r=.43; p=.02) and expected marker assessment (r=.41; p=.04). In contrast, self-efficacy was found to be negatively associated to frustration about the task (r=-.46; p=.02). Moreover, correlation analyses yielded only one statistically significant positive association, i.e., self-efficacy was positively related to effort (r=.45; p=.02), indicating that higher scores on self-efficacy beliefs were associated with higher scores on the effort invested.

Regarding the perception of the task and its relationship with the translation performance scores, as shown in Table 7, the results showed significant correlations only for the Text 3 (the strictest time condition). A positive relationship between accuracy and the scores for perceived frustration (r=.43; p=.03), stressful task (r=.39; p=.04), and importance of the task (r=.42; p=.04), and a significant negative relationship between accuracy and expected marker assessment (r=-.49; p=.01).

4.6. Discussion

The aim of this study was to explore the impact of perceived self-efficacy on the hormonal (i.e., salivary cortisol response) and emotional (i.e., anxiety, positive and negative affective states) responses to translating under time pressure, as well as on translation performance.

Our first hypothesis predicted that when translating under time pressure students' cortisol response would be higher and their performance would be worse than when translating under no time pressure. Results from the repeated-measures ANOVA did not provide support for our hypothesis about the influence of time pressure on cortisol response because no statistically significant differences in students' cortisol levels were found across the different stages of the experimental task. Moreover, as shown in Figure 1, cortisol response displayed the opposite pattern to an average stress response, with more stringent time constraints resulting in a progressive decrease in cortisol. Thus, cortisol response started to decrease from the moment translation students started to translate, which might be due to an increased attentional response to the translation task. This plausible explanation has been also pointed out by Rojo López, Cifuentes-Férez, and Espín-López (2021) on translating under time pressure and other existing results on translation research which indicate that a decrease in cortisol response seems likely to be caused by attentional focus to the task (Rojo López & Naranjo Sánchez, 2021; Rojo López, Ramos Caro, & Espín López, 2021).

As far as the relation between cortisol levels and translation performance,

correlation analyses provided only partial support to our hypothesis. Generally, higher cortisol levels were associated with less accuracy only in the case of pre-task cortisol levels (t0) which emerged as predictor of lower meaning and total accuracy in the condition with no time constraints (Text 1). It seems that translation trainees who entered the experimental session more stressed were more likely to do it worse when translating Text 1, indicating a detrimental effect of stress in performance, even in the absence of time constraints. This result fits well with data supporting the negative effect of stress on task performance (e.g., Lukasik et al. 2019).

In contrast, in the most stringent time condition (Text 3), a positive relation is found between high cortisol levels and accuracy, namely, those trainees who scored higher for correctness under this condition displayed the highest cortisol responses in the last stage of the experiment (t+45, recovery stage), suggesting a beneficial effect of stress on grammatical and orthotypographic aspects of the target text in the most stressful condition.

Our second hypothesis postulated that higher levels of self-efficacy would be associated with trainees' lower cortisol levels and state anxiety responses under time pressure (i.e., Text 2 and Text 3). Results from the correlation analysis partially confirmed our hypothesis for the expected relations between self-efficacy, cortisol response and state anxiety. Data revealed that trainees with higher levels of self-efficacy showed lower levels of cortisol at the end of the experiment (t+45) and lower levels of pre-task state anxiety, suggesting that self-efficacy could be a protective factor against stress induced by time pressure. These data provide support for Bandura's (1977) claims that self-efficacy beliefs may affect people's anxiety in relation to a task and, as a consequence, performance. Nevertheless, no statistically significant results were found for cortisol levels after the translation of texts 1, 2 or 3; in other words, there is not any relation between self-efficacy and cortisol levels after translating the texts, suggesting that focusing on accomplishing the translation tasks could help to reduce stress (Rojo López, Cifuentes-Férez, & Espín-López, 2021) and, thus, aligning with previous findings of the potential impact of attentional focus on the translation tasks (Rojo López & Naranjo Sánchez, 2021; Rojo López, Ramos Caro, & Espín López, 2021).

Our third hypothesis predicted that due to time pressure, after finishing the translation tasks, positive affect scores would be lower whereas negative affect and state anxiety scores would be higher. Despite the fact that state anxiety scores increased after the translation tasks, our data only revealed statistically significant differences between pre- and post-task scores for positive affect, with lower scores after finishing the tasks than before starting them. Our data indicate that translation trainees' negative affect did not change much after translating against the clock, but post-task positive affect

significantly decreased, indicating that translating under time pressure did affect trainees. Our hypothesis was, therefore, corroborated only for positive affect.

Regarding our fourth hypothesis which predicted that higher levels of self-efficacy would be associated with better performance in the translation tasks under time constraints (i.e., Text 2 and Text 3). Correlation analyses showed a significant negative relation between self-efficacy and accuracy in meaning for Text 3 (most stringent time constraints), suggesting a negative effect of trainees' self-efficacy on their translation performance under strict time constraints. It could be likely that trainees with higher self-efficacy beliefs overestimated their capabilities and made more mistakes in terms of the transfer of meaning than those who had a lower self-efficacy and performed significantly better in transference of meaning. This result is consistent with other findings pertaining to the related construct of self-esteem, which have also been pointed out to have a negative impact on translation performance in terms of spelling and punctuation errors (Cifuentes-Férez & Meseguer Cutillas, 2018; Rojo López, Cifuentes-Férez, & Espín-López, 2021). In addition, data showed a positive relation between pre-task negative affect and higher accuracy in terms of meaning in the moderate time condition (Text 2), suggesting that negative affect might have a positive impact on translation performance under given circumstances. Moreover, in line with the previously mentioned research, our data revealed a negative correlation between both pre- and post-task positive affect and correctness (spelling and punctuation errors) in the most stringent time condition (Text 3), suggesting that feeling too positive may hinder performance on spelling and punctuation.

Our fifth hypothesis stated that higher levels of self-efficacy will be associated with trainees' attribution of results to their capacity and effort. Correlation analysis revealed that trainees with higher self-efficacy scores were more likely to attribute their results to their effort made, but no statistically significant correlation was found between scores for self-efficacy beliefs and result attribution to their capacity. However, as shown in Figure 3, it can be observed that trainees mostly attribute their results to their capacity, then to effort made and task difficulty, assigning the lowest scores to luck. Additionally, we were interested in exploring the role of self-efficacy beliefs in trainees' perceptions after the translation tasks. Our data indicated, on the one hand, that translation students with higher scores on self-efficacy felt they performed well in the tasks, expected higher scores in the marker assessment and higher achievement and felt less frustrated. On the other, those with lower self-efficacy beliefs felt they did worse in the tasks, expected lower achievement, felt more frustrated and consider the tasks to be more stressful. These findings in terms of trainees' perceptions on the tasks are in conflict with actual translation performance since those trainees who felt more frustrated, thought the task was more stressful and considered the task of

greater importance are the ones who obtained higher accuracy scores in Text 3, that is, under the strictest time condition. These results reveal intriguing differences between trainees' perception and actual real performance which are of great importance for translator training.

5. Conclusions

Results from this investigation provides evidence for both positive and negative effects of trainees' self-efficacy in translation performance against the clock. On the one hand, its effects on pre-task state anxiety and on cortisol responses after finishing the tasks indicate that self-efficacy could be considered a protective factor against stress and anxiety. On the other, its negative effects on performance under the most stringent time condition suggest in line with extant literature that psychological constructs related to the self, such as self-efficacy seem to be good predictors of lower accuracy under the most stringent time condition. In contrast, under the same time condition, our data suggest a beneficial effect of high cortisol responses on accuracy, namely, trainees who maintained higher cortisol levels, managed to produce better target texts under the most stressful situation. Results also suggest that the way translation trainees feel and think about their performance in translation does not seem to correspond to their actual objective performance. Nevertheless, the present study is exploratory in this regard and further research needs to be conducted to elucidate which other constructs related to the self might influence on these divergences between feelings, thoughts and actual translation performance.

As far as future research is concerned, there are some limitations that should be addressed. First of all, the study should be replicated with larger samples of participants, as well as professional translators. It would be interesting to investigate the effects of different levels of translation competence and expertise. Second, as stated above, other related self-constructs, such as self-esteem or self-concept should be included in future studies to test their effects on performance and trainees' thoughts, feelings and perceptions. Third, although three comparable narrative texts in terms of readability were used as stimuli, it is still possible that any text differences had an effect in performance as readability does not mean that texts have similar translation difficulty. Additionally, it would be interesting to use other text types or even positive and negative emotional texts of the same text type so as to shed light on the effect of text types and of valence when translating under time constraints. Fourth, it is observed that trainees had high cortisol levels when they entered the experimental room. Experimental anxiety, therefore, should be addressed in future research so as to reduce it by starting with some sort of relaxation phase such as sitting comfortably, closing their eyes and listening to relaxing music for a couple of minutes. Last, further research is needed to elucidate whether overall cortisol response was due to stress induced by time constraints or to the attentional focus on the translation task.

Potential applications of the main findings of this study may be found in translator training and in the workplace. Nowadays, the internet, technological changes to translation such as machine translation have significantly altered translation industry. Professional translators are expected to meet tight deadlines and quality of work and so are translation trainees in a lesser extent. Stress caused by time pressure can be temporary or it can continue over a long term, affecting, thus, hormones, mood, and all aspects of translators' health and well-being. Translation trainees and professionals can benefit from awareness of the impact of time pressure on translation performance and of the protective role of self-efficacy against stress and anxiety. Furthermore, being aware of the potential negative effect of high self-efficacy on performance can be helpful for both trainees and professionals so that their high self-efficacy beliefs do not work against them. Finally, it can be of great help to both translation teachers and employers in order to maximize their students and staff's abilities and minimise their weaknesses.

Acknowledgement

This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and FEDER/UE funds (grant number PID2021-123650NB-100). The authors report there are no competing interests to declare.

6. References

Abihssira, L. (2019). *Rasgos de la personalidad y creatividad determinantes en la calidad y la creatividad en traducción [Determining factors of personality and creativity in translation product's quality and creativity].* [Unpublished doctoral dissertation, University of Murcia].

Albin, J. (2012). Competencia y autoeficacia: Estudio de los factores afectivos en el traductor. *Hikma, 11*, 9–33.

Angelone, E. (2010). Uncertainty, Uncertainty Management and Metacognitive Problem Solving in the Translation Task. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 17-40). Amsterdam: John Benjamins. https://doi.org/10.1075/ata.xv.03ang

Araghian, R., Ghonsooly, B., & Ghanizadeh, A. (2018). Investigating problem-solving strategies of translation trainees with high and low levels of self- efficacy. *Translation, Cognition and Behaviour, 1*, 74–97. https://doi.org/10.1075/tcb.00004.ara

Atkinson, D. (2012). *Freelance Translator Success and Psychological Skill: A Study of Translator Competence with Perspectives from Work Psychology.* [Doctoral dissertation, University of Auckland]. Retrieved from: https://researchspace.auckland.ac.nz/handle/2292/18723. [14 July 2021).

Atkinson, D. P. (2014). Developing psychological skill for the global language industry: an exploration of approaches to translator and interpreter training. *Translation Spaces, 3*, 1–24. https://doi.org/10.1075/ts.3.01atk

Bandura, A. (1995). *Self efficacy in changing societies.* Cambridge: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales in Self-efficacy beliefs of adolescents. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich; CT: Information Age Publishing.

Bayer-Hohenwarter, G. (2009). Methodological reflections on the experimental design of time-pressure studies. *Across Languages and Cultures, 10*(2), 193–206. https://doi.org/10.1556/Acr.10. 2009.2.2

Beattie, S., Fakehy, M., & Woodman, T. (2014). Examining the moderating effects of time on task and task complexity on the within person self-efficacy and performance relationship. *Psychology of Sport and Exercise, 15*, 605–610.

Bolaños-Medina, A. (2012). Las pruebas psicométricas en la investigación empírica sobre los procesos cognitivos del traductor. In S. Cruces Colado, M. del Pozo Triviño, A. Luna Alonso, & A. Álvarez Lugrís (Eds.), *Traducir en la Frontera* (pp. 469–482). Spain: Atrio.

Bolaños-Medina, A. (2016). Translation psychology within the framework of translator studies: new research perspectives. In C. Martín de León & V. González-Ruíz (Eds.), *From the lab to the classroom and back again perspectives on translation and interpreting training* (pp. 59–100). Frankfurt: Peter Lang.

Bolaños-Medina, A., & Núñez, J. (2018). A preliminary scale for assessing translators' self-efficacy. *Across Languages and Cultures, 19*, 53–78. https://doi.org/10.1556/084.2018.19.1.3

Bontempo, K. & Napier, J. (2014). Evaluating emotional stability as a predictor of interpreter competence and aptitude for interpreting. In F. Pöchacker & M. Liu (Eds.), *Aptitude for Interpreting* (pp. 87–106). Amsterdam: John Benjamins. https://doi.org/10.1075/bct.68.05bon.

Bosch, J. A., Veerman, E. C. I. Geus, E. J. D., & Proctor, G. B. (2011). A-amylase as a reliable and convenient measure of sympathetic activity: don't start salivating just yet! *Psychoneuroendocrinology, 36*, 449–453. https://doi.org/10.1016/j.psyneuen.2010.12.019

Cifuentes-Férez, P., & Fenollar-Cortés, J. (2017). On the impact of self-esteem, emotion regulation and emotional expressivity on student translators' performance. *Vigo International Journal of Applied Linguistics, 14*(1), 71-97.

Cifuentes-Férez, P., & Meseguer, P. (2018). Can self-esteem and creative intelligence foster accuracy and creativity in professional translators? *Translation, Cognition and Behaviour, 1*(1), 341–360. https://doi. org/10.1075/tcb.00015.cif

Costa, R., Serrano, M. A., & Salvador, A. (2016). Importance of self-efficacy in psychoendocrine responses to competition and performance in women. *Psicothema, 28*(1), 66-70.

De Rooze, B. (2003). *La traducción, contra reloj [The translation, against the clock].* [Unpublished doctoral dissertation, University of Granada].

Espín López, L. (2009). *Respuesta hormonal y cardiovascular a un estresor social de laboratorio: efectos sobre la memoria declarativa [Hormonal and cardiovascular response to a social laboratory stressor: effects on declarative memory].* [Unpublished doctoral dissertation, University of Murcia].

Fida R., Paciello M., Tramontano C., Barbaranelli C., & Farnese M. L. (2015). "Yes, I Can": the protective role of personal self-efficacy in hindering counterproductive work behavior under stressful conditions. *Anxiety Stress Coping, 28*(5),479-99. https://doi.org/10.1080/10615806.2014.969718

Ghobadi, M., Madadi, G., & Najafian, B. (2017). A study of the effects of time pressure on translation quantity and quality. I*nternational Journal of Comparative Literature and Translation Studies, 5*(2), 7–13. https://doi.org/10.7575/aiac.ijclts.v.5n.2p.7.

Gouadec D. (2007). *Translation as a Profession.* Amsterdam and Philadelphia: John Benjamins.

Hanin, Y. L. (2000). Individual Zones of Optimal Functioning (IZOF) Model. In Y. L. Hanin (Ed.), *Emotions in Sport* (pp. 65–89). Champaign, IL: Human Kinetics.

Hansen, G. (2010). Integrative Description of Translation Processes. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 189–212). Amsterdam: John Benjamins. https://doi.org/10.1075/ata.xv.11han

Haro-Soler, M. M. (2017). Teaching practices and translation students' self- efficacy: the teachers' perceptions. *Current Trends in Translation Teaching and Learning, 4*, 198–228.

Haro-Soler, M. M. (2018). Translation students' self-efficacy beliefs: An X of their development. [Doctoral dissertation, University of Granada]. Available at: https://digibug.ugr.es/handle/10481/53590 (Accessed February 20, 2021).

Haro-Soler, M. M. (2019a). Vicarious learning in the translation classroom: how can it influence students' self-efficacy beliefs? *English Studies at NBU, 1*, 92–113. https://doi.org/10.33919/esnbu.19.1.5

Haro-Soler, M. M. (2019b). Translator's self-confidence versus self-efficacy: a terminology proposal and literature review. *Cadernos de Tradução, 39*, 204–226. https://doi.org/10.5007/2175-7968.2019v39n2p204.

Haro-Soler, M. M. (2020). Propuesta de escala de creencias de autoeficacia del traductor: versión en español y en inglés. *Tonos digital, 42*, 1–28.

Haro-Soler, M. M., & Kiraly, D. (2019). Exploring self-efficacy beliefs in symbiotic collaboration with students: an action research project. T*he Interpreter and Translator Trainer, 13*, 255–270. https://doi.org/10.1080/1750399X.2019.1656405

Hendricks, K. S., Smith, T. D., & Legutki, A. R. (2016). Competitive comparison in music: influences upon self-efficacy beliefs by gender. *Gender and Education, 28*(7), 918-934.

Hitches, E., Woodcock, S., & Ehrich, J. (2023). Shedding light on students with support needs: Comparisons of stress, self-efficacy, and disclosure. *Journal of Diversity in Higher Education, 16*(2), 205–2014. https://doi.org/10.1037/dhe0000328

Hubscher-Davidson, S. (2016). Trait Emotional Intelligence and Translation: A Study of Professional Translators. *Target, 28*(1), 132–157. https://doi.org/10.1075/target.28.1.06hub

Jääskeläinen, R. (2000). Focus on methodology in think-aloud studies on translating. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on empirical research* (pp.71-82). Amsterdam: John Benjamins.

Jensen, A. (1999). Time pressure in translation. In Hansen G. (Ed.), *Probing the Process in Translation. Methods and Results* (pp. 103-119). Copenhagen: Samfundslitteratur.

Jensen, A.. (2000). The effects of time on cognitive processes and strategies in translation. [Unpublished doctoral dissertation, Copenhagen Business School].

Jensen, A. & Jakobsen, A. L. (2000). Translating under time pressure: an empirical investigation of problem-solving activity and translation strategies by non-professional and professional translators. In A. Chesterman, N. G. S. Salvador, & Y. Gambier (Eds.), *Translation in context: Selected contributions from the EST congress, Granada 1998* (Vol. 39, pp. 105–116). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.39.13jen

Jiménez Ivars, A., Pinazo Catalayud, D., & Ruiz i Fores, M. (2014). Self-efficacy and language proficiency in interpreter trainees. *The Interpreter and Translator Trainer, 8*(2), 167-182. https://doi.org/10.1080/1750399X.2014.908552

Kavanagh, D. J., & Bower, G. H. (1985). Mood and self-efficacy: Impact of joy and sadness on perceived capabilities. *Cognitive therapy and research, 9*(5), 507-525.

Kavanagh, D., & Hausfeld, S. (1986). Physical performance and self-efficacy under happy and sad moods. *Journal of Sport and Exercise Psychology, 8*(2), 112-123.

Khalzanova, S. (2008). *Revision and time constraints in Translation.* [Unpublished MA dissertation, Universitat Rovira i Virgili, Spain.]

Kiraly, D. (1995). *Pathways to Translation. Pedagogy and Process.* Kent, OH: Kent State University Press.

Kleiner S. (2014). Subjective time pressure: General or domain specific? *Social Science Research* 47, 108–120. https://doi.org/10.1016/j.ssresearch.2014.03.013 PMID:24913948

Konttinen, K. (2021). A self-efficacy scale for measuring student progress in translation company simulations. *Across Languages and Cultures, 22*(1), 64-81. https://doi.org/10.1556/084.2021.00004

Kourouni, K. (2012). *Translating under time constraints in an undergraduate context: A study of students' products, processes and learning styles.* [Unpublished doctoral dissertation, Universitat Rovira i Virgili].

Korpal, P. (2016). Interpreting as a stressful activity: Physiological measures of stress in simultaneous interpreting. *Poznan Studies in Contemporary Linguistics, 52*(2), 297-316.

Korpal, P., & Jankowiak, K. (2021). On the potential impact of directionality on emotion processing in interpreting. *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile, 8*, 43-60.

Künzli, A. (2007). Translation Revision: a study of the performance of ten professional translators revising a legal text. In Y. Gambier, M. Schlesinger, & R. Stolze (Eds.), *Doubts and Directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004* (pp. 115-126). Amsterdam: John Benjamins.

Kussmaul, P. (1995). *Training the translator.* Amsterdam: John Benjamins.

Lautenbach, F., & Laborde, S. (2016). The influence of hormonal stress on performance. In M. Raab, B. Lobinger, S. Hoffmann, A. Pizzera, & S. Laborde (Eds.), *Performance Psychology* (pp. 315-328). San Diego, USA: Academic Press.

Lee, S.B. (2014). An Interpreting Self-Efficacy (ISE) Scale for Undergraduate Students Majoring in Consecutive Interpreting: Construction and Preliminary Validation. *The Interpreter and Translator Trainer 8*(2), 183–203.

Lehka-Paul, O. (2020). *Personality Matters: The Translator's Personality in the Process of Self-Revision.* Leuven: Leuven University Press.

Lehr, C. (2014). *The influence of emotion on language performance: Study of a neglected determinant of decision-making in professional translators.* [Unpublished doctoral dissertation, University of Geneva].

Lorenzo, M.P. (2002). Competencia Revisora y Traducción Inversa. *Cadernos de Tradução, 2*(10), 133-166.

Lukasik, K. M., Waris, O., Soveri, A., Lehtonen, M., & Laine, M. (2019). The relationship of anxiety and stress with working memory performance in a large non-depressed sample. *Frontiers in psychology, 10*, 1-4. https://doi.org/10.3389/fpsyg.2019.00004 PMID: 30728790

MacAfee, E., & Comeau, G. (2020). Exploring music performance anxiety, self-efficacy, performance quality, and behavioural anxiety within a self-modelling intervention for young musicians. *Music Education Research, 22*(4), 457-477.

McPherson, G., & J. McCormick. (2006). Self-efficacy and Music Performance. *Psychology of Music, 34*(3), 322–336.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in cognitive sciences, 7*(3), 141-144.

Mor, S., Day, H. I., Flett, G. L., & Hewitt, P. L. (1995). Perfectionism, Control, and Components of Performance Anxiety in Professional Artists. *Cognitive Therapy and Research, 19*, 207–225.

Moorkens, J. (2017). Under pressure: translation in times of austerity. *Perspectives, 25*(3), 464–477. https://doi.org/10.1080/0907676X.2017.1285331

Muñoz Martín, R. (2014). Una instantánea movida de la investigación en procesos de traducción. *Monografias de Traducción e Interpretación, Special Issue – Minding Translation, 2014, 9-47.* http://dx.doi.org/10.6035/MonTI.2014.ne1.1

Muñoz Martín, R., & Xiao, K. (Eds.) (2020). Cognitive Translation Studies: Theoretical models and methodological criticism [Special issue]. *Linguistica Antverpiensia, New Series–Themes in Translation Studies, 19*, 1–24.

Nielsen, S. E., Segal, S. K., Worden, I. V., Yim, I. S., & Cahill, L. (2013). Hormonal contraception use alters stress responses and emotional memory. *Biological Psychology, 92*(2), 257-66. https://doi.org/10.1016/j.biopsycho.2012.10.007

Ordóñez, L. D., Benson, L., & Pittarello, A. (2015). Time-pressure perception and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making, 2* (pp. 517-542). Chichester, UK: Wiley-Blackwell. https://doi.org/10.1002/9781118468333.ch18

Orejudo, S. F., Azarza-Alzugaray, F. J., Casanova, O., Rodríguez-Ledo, C., & Mazas, B. (2017). The Relation of Music Performance Anxiety (MPA) to Optimism, Self-efficacy, and Sensitivity to Reward and Punishment: Testing Barlow's Theory of Vulnerability on a Sample of Spanish Music Students. *Psychology of Music, 45*(4), 570–583.

Pajares, F. (1997). Current Directions in Self-efficacy Research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in Motivation and Achievement* (pp. 1-49). Greenwich, CT: JAI Press.

Randler, C., Demirhan, E., Wüst-Ackermann, P., & Desch, I. H. (2016). Influence of a dissection video clip on anxiety, affect, and self-efficacy in educational dissection: A treatment study. *CBE–Life Sciences Education, 15*, 1-8.

Reigal, R. E., Vázquez-Diz, J. A., Morillo-Baro, J. P., Hernández-Mendo, A., & Morales-Sánchez, V. (2020). Psychological profile, competitive anxiety, moods and self-

efficacy in beach handball players. *International journal of environmental research and public health, 17*(1), 241.

Rohleder, N., & Nater, U. M. (2009). Determinants of salivary a-amylase in humans and methodological considerations. *Psychoneuroendocrinology, 34*, 469–485. https://doi.org/10.1016/j.psyneuen.2008.12.004

Rojo López, A. M. (2019). *La investigación de la creatividad en traducción: Resultados del proyecto TRANSCREA.* Granada: Comares.

Rojo López, A. M., Cifuentes-Férez, P., & Espín López, L. (2021). The influence of time pressure on translation trainees' performance: Testing the relationship between self-esteem, salivary cortisol and subjective stress response. *PLoS ONE, 16*(9), e0257727. https://doi.org/10.1371/journal.pone.0257727

Rojo López, A. M., & Naranjo Sánchez, B. (2021). Translating in times of crisis: A study about the emotional effects of the COVID19 pandemic on the translation of evaluative language. *Journal of Pragmatics, 176*, 29–40. https://doi.org/10.1016/j.pragma.2021.01.018

Rojo López, A. M., Foulquié-Rubio, A. I., Espín López, L., & Martínez Sánchez, F. (2021). Analysis of speech rhythm and heart rate as indicators of stress on student interpreters. *Perspectives*, *29*(4), 591-607. https://doi.org/10.1080/0907676X.2021.1900305.

Rojo López, A. M., Ramos, M., & Espín López, L. (2021). Audio Described vs. Audiovisual Porn: Cortisol, Heart Rate and Engagement in Visually Impaired vs. Sighted Participants. *Frontiers in Psychology, 29*(4), 591-607. *12*, 1–15, 661452. doi: 10.3389/fpsyg.2021.661452

Seisdedos N. (1988). *Adaptación Española del STAI, cuestionario de ansiedad estado-rasgo* [Spanish adaptation of the STAI, state-trait anxiety inventory]. Madrid: TEA Ediciones.

Sharmin, S., Špakov, O., Räihä, K.-J., & Jakobsen, A. L. (2008). Where and for how long do translators look at the screen while translating? In S. Göpferich, A. Jakobsen, & I. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (pp. 31–51). Copenhagen: Samfundslitteratur.

Shreve, G. M., & Lacruz, I. (2014). Translation as higher-order text processing. In S. Bermann & C. Porter (Eds.), *A companion to translation studies* (pp.107-118). Clevedon: Multilingual Matters.

Shreve, Gregory M., & Angelone E. (2010). Translation and Cognition: Recent Developments. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 1–13). Amsterdam: John Benjamins. https://doi.org/10.1075/ata.xv.01shr

Spielberger, C. D., Gorsuch, R. L., Lushene R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Suay, F., Salvador, A., González-Bono, E., Sanchís, C., Martínez, M., Martínez-Sanchis, S., Simón, V. M., & Montoro, J. B. (1999). Effects of competition and its outcome on serum testosterone, cortisol and prolactin. *Psychoneuroendocrinology, 24*, 551-566.

Thelwell, R. C., Lane, A. M., & Weston, N. J. (2007). Mood states, self-set goals, self-efficacy and performance in academic examinations. *Personality and Individual Differences, 42*(3), 573-583.

Tirkkonen-Condit, S., & Laukkanen, J. (1996). Evaluations: a key to understanding the affective dimension of translational decisions. *Meta, 41*, 45–59. https://doi.org/10.7202/002360AR

Torre Puente, J. C. (2007). *Una triple alianza para un aprendizaje universitario de calidad*. Madrid: Universidad Pontificia Comillas.

Travis, J., Kaszycki, A., Geden, M., & Bunde, J. (2020). Some stress is good stress: The challenge-hindrance framework, academic self-efficacy, and academic outcomes. *Journal of educational Psychology, 112*(8), 1632.

Udayar, S., Fiori, M., & Bausseron, E. (2020). Emotional intelligence and performance in a stressful task: The mediating role of self-efficacy. *Personality and Individual Differences, 156*, 109790.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063

Weng, Y., & Zheng, B. (2020). A multi-methodological approach to studying time-pressure in written translation: Manipulation and measurement. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 218–236.

Xiao, K. & Muñoz, R. (2020). Cognitive Translation Studies: Models and methods at the cutting edge. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 1–24.

Yang Y., Cao X., & Huo, X. (2021). The Psychometric Properties of Translating Self-Efficacy Belief: Perspectives From Chinese Learners of Translation. *Frontiers in Psychology, 12*, 1–11, 642566. https://doi.org/10.3389/fpsyg.2021.642566

Zeldin, A. L. & Pajares, F. (2000). Against the Odds: Self-efficacy Beliefs of Women in Mathematical, Scientific, and Technological Careers. *American Educational Research Journal, 37*(1), 215-246.

Zhang, X., & Ardasheva, Y. (2019). Sources of college EFL learners' self-efficacy in the English public speaking domain. *English for Specific Purposes, 53*, 47-59.

Association of lexical and collocation knowledge: A comparative analysis of a learner corpus of English and a native speaker corpus

Sung-Yeon Kim
Department of English Education
Hanyang University
sungkim@hanyang.ac.kr

Dongkwang Shin
Department of English Education
Gwangju National University of Education
sdhara@gmail.com

Kyung-Sook Kim
College English Education Center
Hanyang University
cindytesol@hanyang.ac.kr

Abstract

This study compared L2 learners' (EFL) use of single words and collocations with that of native speakers (NSs). The study compiled two text corpora, one from an existing native corpus and the other with texts from Korean EFL learners at three proficiency levels. The study found a significant difference in the lexical diversity of single words between the two corpora. The difference between the advanced EFL learners and NSs was, however, non-significant, which means that advanced learners were on par with NSs in terms of using a diverse range of words. In a comparison of lexical distribution of single words, however, these advanced learners were found to use words between the 1K (the first 1,000 words) and 4K level (4,001 ~ 5,000 words) whereas NSs used words beyond this level. This suggests the need to cover a wider range of words in classroom teaching and assessment. Regarding collocational knowledge, the difference between the NSs and EFL group was apparent and statistically significant, regardless of learners' proficiency levels. Namely, EFL learners used far fewer collocations in a smaller range than their counterparts. The learners' limited collocational competence indicates that collocations should be considered an integral component of the curriculum, instruction, and assessment.

Keywords: single words, collocations, lexical competence, collocational knowledge, corpus.

Resumen

Este estudio comparó el uso de palabras sueltas y colocaciones de los estudiantes de L2 (ILE) con el de los hablantes nativos (HNs). El estudio compiló dos corpus de texto, uno de un corpus nativo existente y otro con textos de estudiantes coreanos de ILE en tres niveles de competencia. El estudio encontró una diferencia significativa en la diversidad léxica de palabras individuales entre los dos corpus. Sin embargo, la diferencia entre los estudiantes avanzados de ILE y los HNs no fue significativa, lo que significa que los estudiantes avanzados estaban a la par con los NS en términos de uso de una amplia gama de palabras. Sin embargo, en una comparación de la distribución léxica de palabras individuales, se encontró que estos estudiantes avanzados usaban palabras entre el nivel 1K (las primeras 1000 palabras) y 4K (4001 ~ 5000 palabras), mientras que los HNs usaban palabras más allá de este nivel. Esto sugiere la necesidad de cubrir una gama más amplia de palabras en la enseñanza y la evaluación en el aula. Con respecto al conocimiento de colocación, la diferencia entre el grupo HNs y ILE fue evidente y estadísticamente significativa, independientemente de los niveles de competencia de los alumnos. Es decir, los estudiantes de ILE utilizaron muchas menos colocaciones en un rango más pequeño que los otros. La competencia de colocación limitada de los alumnos indica que las colocaciones deben considerarse un componente integral del plan de estudios, la instrucción y la evaluación.

Palabras clave: palabras sueltas, colocaciones, competencia léxica, conocimiento colocacional, corpus.

1. Introduction

There are two common generalizations about lexical competence that deserve attention in the era of globalization where language acquisition is facilitated with internet technology and transnational mobility (Duff, 2015; Ma, 2017; Yu & Trainin, 2022). First, native speakers are superior to EFL learners in terms of vocabulary knowledge (Demir, 2017; Zareva 2007); and second, lexical knowledge automatically entails collocational competence (Nation, 2001; Zareva et al., 2005). The first argument has been received as a fact based on a considerable body of research. The second, however, has been questioned by a substantial number of L2 vocabulary studies. Many researchers have cautioned that lexical competence does not imply collocational knowledge (Bahns & Eldaw, 1993; Laufer & Waldman, 2011; Paquot & Granger, 2012). For instance, Siyanova-Chanturia (2015) claimed that L2 learners' general

vocabulary knowledge is not transferred to collocational competence, reporting diverse problems that second language learners often experience when using collocations. This is a particularly thorny problem for language learners that creates a sizable gap between them and their native counterparts.

Learner difficulty with collocations partly arises because they are "a type of formulaic expression made of strongly associated pairs of words characterized by restricted substitutability", (Bestgen, 2017, p. 67), as in the following examples: "make a mistake" vs. "*do a mistake", "strong tea" vs. "*powerful tea", and "heavy drinker" vs. "*extravagant drinker" (Supasiraprapa, 2018). This makes it more difficult for learners to acquire collocations as they have to differentiate the collocational restrictions and use the right combinations. However, classroom instruction and assessment generally focus on single words (SWs) rather than on collocations. In addition, learners' lexical competence is often evaluated based on their knowledge of SWs (Laufer & Nation, 2001; Nation, 2001, 2006; Paquot, 2007; Shin, 2015). This negligence of teaching and assessing collocations is, however, problematic as it can lead to unbalanced development of lexical competence, which is not complete without collocational knowledge. Collocations should not be neglected as they comprise 20 to 50 percent of the spoken and written discourse of native speakers (Erman & Warren, 2000; Foster, 2001; Laufer & Waldman, 2011). In addition, a strong relationship was found between formulaic competence and writing quality, and thus formulaic measures are known to predict writing quality better than single-word measures (Bestgen, 2017). Yet, as discussed in previous studies, L2 learners have problems with collocations, and as a result, underuse, overuse, or misuse them (González-Fernández & Schmitt, 2015; Siyanova-Chanturia, 2015). Against this backdrop, collocation deserves special attention, particularly in an EFL context like Korea, where English teaching disregards language areas such as collocation that are not measured in the college entrance exam (Kim, 2021). However, given that collocational knowledge is known as a yardstick for distinguishing L1 users from L2 users of English, it is critical to assess L2 learners' collocational competence and teach collocations as well as single words (Siyanova-Chanturia, 2015).

In recent years, numerous studies have been conducted on learners' use of collocations; however, they have focused on a certain number of lexical bundles or a limited set of combinations, such as adjective-noun (e.g., "heavy rain"), verb-noun (e.g., "tell lies"), or intensifier-adjective (e.g., "deeply rooted") collocations (Altenberg & Granger, 2001; Boers et al., 2014; Cross & Papp, 2008; Durrant & Schmitt, 2009; García Salido & Garcia, 2018; Granger, 1998a; Kashiha & Chan, 2015; Li & Schmitt, 2010; Nesselhauf, 2003; Siyanova-Chanturia, 2015; Siyanova-Chanturia & Schmitt, 2008). Also, few studies have investigated single words and collocations together (Shin et al., 2018; Vedder & Benigno, 2016), and those have been limited in scope, simply

describing the frequencies of either single- or multi-words. This is unfortunate because they offer an incomplete account of L2 learners' linguistic competence (Bahns & Eldaw, 1993; Read & Nation, 2006). Consequently, little is known about L2 learners' collocational competence.

Given that vocabulary and collocations have rarely been examined together, this study aims to analyze the two using the BNC-COCA25 Range (a vocabulary analysis program loaded with 25 wordlists constructed from the Corpus of Contemporary American English or COCA, and British National Corpus or BNC, Nation, 2012) and COCA_MWU20 ColloGram (a collocation analysis program based on 20 collocation lists constructed from COCA, Shin et al., 2018), respectively, to yield a thorough assessment of L2 learners' lexical competence and examine the association between lexical and collocational knowledge. Considering that these two areas of lexical competence deserve equal attention, they should be included as units of lexical assessment. In addition, L2 learners' knowledge of these two areas should be compared to that of their native counterparts. To ensure the validity of the research, the learner corpus used in the study was carefully compiled from three graded proficiency levels (high, mid, low), and a reference corpus of approximately the same size as the learner corpus was selected from a native corpus (LOCNESS) to facilitate a comparison of the two corpora (Granger, 1998b). Such a comparison allows us to measure what learners across proficiency levels can do relative to their native counterparts in terms of using single words and collocations in their written text production.

2. Literature review

It is generally accepted that lexical knowledge plays a crucial role in shaping target language proficiency (Laufer & Nation, 1995; Lewis, 1993; Milton, 2013; Nation, 2001, 2006; Olinghouse & Leaird, 2009; Schmitt, 2008). This awareness of the role of vocabulary has led to a surge in vocabulary research, further accelerated by computer-aided language analysis (Cobb, 2010; Coxhead, 2000, 2016; Read, 2007). For example, corpus analysis is widely used to measure the size of learner vocabulary using frequency-based wordlists. Researchers have divided English vocabulary into fourteen different levels according to their frequency and the first 3,000-word families (1K to 3K) are generally considered as high-frequency vocabulary, words beyond the 9K frequency bands are treated as advanced words, and those between the first 3K and 9K as mid-frequency words (Leech et al., 2001; Nation, 2006; Schmitt & Schmitt, 2014). An analysis of lexical profiles can be also used to identify the distribution of L2 learners' vocabulary and guide them to learn low-frequency words as well as high-frequency words (Johnson et al., 2016). Ha (2021) analyzed the lexical profiles of input texts used in the IELTS listening and reading tests using the BNC-COCA25 Range program, and found that the 3K words cover 95% of the words in the listening texts, and the 5K

words have 95% coverage of the running words in the reading passages.

Vocabulary research has focused on estimating the vocabulary size required for performing language tasks, describing learner language according to frequency-based wordlists, or analyzing L2 learner language against a native corpus in terms of lexical properties, such as word frequencies, lexical diversity, and lexical density (Dang & Webb, 2014; Gregori-Signes & Clavel-Arroita, 2015; Kao & Wang, 2014; Laufer & Nation, 2001; Nation, 2006; Paquot, 2007; Shin, 2015). For instance, Nation (2006) used fourteen one-thousand-word-family lists (1K though 14K) created from the BNC to analyze the vocabulary size of nine different written and spoken corpora and suggested that L2 learners need receptive vocabulary of the most frequent 8,000 (8K) to 9,000 (9K) words to adequately comprehend written text and 6,000 (6K) to 7,000 (7K) words for spoken text. Recently, Dabbagh and Enayat (2019) investigated the association between productive vocabulary size and writing score using the Vocabulary Levels Test (VLT, Schmitt et al., 2001) to measure learner knowledge of vocabulary at four frequency bands (2000, 3000, 5000, and 10,000 words). They found that mid- (the 5,000-word or 5K) and low (the most frequent 3,000 families or 3K) level vocabulary contributed significantly to EFL students' writing scores. This finding supports an earlier suggestion that teachers and learners should pay attention to 5K words as they cover a large portion of different types of texts (González, 2013; Schmitt & Schmitt, 2014).

There are also other studies that have compared the lexical diversity and distribution of L2 learners with those of L1 writers. Doro (2007), for example, compared Hungarian EFL learners' texts with L1 users' and found that L2 learners used a narrower range of words than L1 users, mostly confined to high-frequency words from the first 2K. In a study conducted with graduate students, Nasseri and Thompson (2021) also reported that EFL graduate students incorporated significantly less word types in their dissertation abstracts compared to L1 users and ESL graduate students. Native speakers (NSs) of English also differed from non-native speakers (NNSs) in terms of lexical distribution (Coxhead & Boutorwick, 2018). While NSs mastered K2 and K3 words by Grade 6 (G6), K5 words by G8, and most of K10 words by G10, NNS acquired K2 words by G6, K3 words by G8, and K5 words by G10. Notably, these studies concentrate only on single words. The question is whether single words alone are sufficient, particularly for processing connected discourse in written language production. This is where knowing a word means "much more than knowing the form-meaning link" (Tsai, 2015, p. 724) through word connections. This connected feature is often observed in the way that words are combined, and collocation is one example of this.

Collocational knowledge plays a pivotal role in shaping lexical competence, given that collocations constitute about half of the native written English corpus (Ehrman

& Warren, 2000; Vedder & Benigno, 2016). According to Ehrman and Warren (2000), more than half (52.3% to be exact) of written discourse consists of formulaic sequences or "collocations" (Dickins, 2020, p. 34). Accordingly, it would be incomplete to estimate lexical competence without considering the ability to understand and use collocations. Unfortunately, collocation poses enormous challenges to L2 learners, who are known to make collocational errors when using language, regardless of proficiency or duration of learning (Kuo, 2009; Laufer & Waldman, 2011; Nesselhauf, 2005; Shitu, 2015; Vedder & Benigno, 2016). For example, from an analysis of 900 essays written by 300 Nigerian college students, Shitu (2015) found that even advanced ESL learners make collocational errors due to L1 interference, overgeneralization, and lack of collocational knowledge; among the six subtypes of lexical errors, a verb-noun (or prepositional phrase) pattern was the most problematic. There are also other studies that have examined native speakers and L2 learners in terms of their collocational competence. For instance, in a comparison of L1 and L2 postgraduate students' English writing, Durrant and Schmitt (2009) found that L2 writers used significantly more collocations from the high-frequency bands, but fewer collocations from the low frequency bands compared to the native speakers. More recently, García Salido and Garcia (2018) found that unlike native speakers, advanced learners of Spanish underused low frequency collocations with high mutual information scores (MIS), or strongly associated words with MIS ≥ 5, while overusing high frequency collocations with low mutual information scores (MIS < 5). Demir (2017) also compared L1 and L2 writers by analyzing the research articles published in top-tier journals in the field of ELT in terms of seven categories (verb + noun, verb + adj./adv., noun + verb, noun + noun, adjective + noun, adverb + adjective, and adverb + verb). He found that the L1 writers used three times (n = 1,548 vs. n = 499) as many collocations as non-native counterparts. Interestingly, the L1 writers used more collocations than the L2 writers in all categories except for noun + verb. The fact that L2 learners lack collocational competence may be because collocations are difficult to acquire (Siyanova-Chanturia, 2015). Collocations are known to be acquired late (Vedder & Benigno, 2016), and their development is "slow and uneven" (Laufer & Waldman, 2011, p. 664).

The significance of collocations and their learnability has ignited interest in studying them. The research in this area can be classified into several categories according to their purposes. While some studies attended to describing the distribution of collocations in L2 written or spoken texts (Kim et al., 2020; Tsai, 2015), others focused on teaching and learning collocations (Boers et al., 2014; Nurmukhamedov, 2017). Furthermore, some studies have compared native speakers to learners with the same L1 (Chen & Baker, 2010; Forsberg, 2010; Henderson & Barr, 2010), along with those conducted with learners of different native languages (Alejo-González, 2010; Cross & Papp, 2008; Waibel, 2008). Another difference is noted in the target of the analysis: some studies

focused on spoken production (Crossley & Salsbury, 2011; Gotz & Schilk, 2011) and others on written production (Bestgen, 2017; Siyanova-Chanturia, 2015).

Despite the diverse range of studies, these do not present a complete picture of L2 learners' lexical competence. This is primarily because their analyses have been restricted to specific lexical bundles and collocational combinations, such as adjective-noun or verb-noun (Altenberg & Granger, 2001; Cross & Papp, 2008; Durrant & Schmitt, 2009; Granger, 1998a; Kashiha & Chan, 2015; Laufer & Waldman, 2011; Li & Schmitt, 2010; Siyanova-Chanturia, 2015; Siyanova-Chanturia & Schmitt, 2008). This tendency may have to do with the inadequacy of existing word lists for multi-word (MW) analysis (Martinez & Schmitt, 2012) or the dearth of tools for the extensive analysis of collocation. While there have been endeavors to develop collocation checker programs (Kuo, 2009) or multi-word lists, such as the Academic Formulas List (Simpson-Vlach & Ellis, 2010), and the Phrasal Expressions List (Martinez & Schmitt, 2012), they seem to be incomplete in terms of coverage.

Another problem with vocabulary research is that only a handful of studies have examined both single words and collocations simultaneously (Bestgen, 2017; Kim et al., 2020; Shin et al., 2018; Vedder & Benigno, 2016), and even these are limited in scope. For example, Bestgen (2017) reported formulaic and single-word measures of lexical richness from an analysis of learner language data, but the study focused exclusively on bigrams (e.g., private correspondence, target audience, good example, and more than). Kim et al. (2020) also considered both single words and collocations, and analyzed Korean college students' lexical competence using the BNC-COCA25 Range, COCA_MWU20 ColloGram, and CLAWS Web Tagger. The problem is, however, that the study focused on the learners' lexical competence according to proficiency levels, and thus little is known about their lexical and collocational knowledge in relation to native speakers. Shin et al. (2018) attempted to compare an EFL learner corpus and a native corpus simply in terms of the distribution of single words and multi-words using the BNC-COCA25 Range and COCA_MWU20 ColloGram. This urges us to describe learners' lexical competence systematically and thoroughly by considering both single words and collocations as the unit of analysis and using a tool that facilitates comprehensive analysis.

This study is an attempt to accurately estimate EFL learners' lexical competence in comparison with their native counterparts in terms of SWs and collocations used in writing. This investigation is based on the two areas of lexical competence with carefully graded written texts from beginning to advanced, following Granger and Bestgen's (2014) suggestion to include texts at different levels when sampling. The study used the BNC-COCA25 Range program (Nation, 2012) and a multi-word analysis program, COCA_MWU20 ColloGram (Shin et al., 2018). These programs have explanatory

power because they can systematically analyze the frequency and types of SWs and collocations across the graded lists constructed from BNC-COCA and COCA, respectively. The present study is significant in that it examined the role of collocation in explaining lexical competence in an EFL context, where excessive attention is paid to single words in terms of curriculum, instruction, and materials. Examining the relation between the two constituents of lexis (single words and collocations) will enable us to capture a holistic picture of EFL learners' lexical competence, aside from filling the gap in prior studies. In addition, the comparison of learner language with native data offers solid comparative ground for mapping out the degree and direction of vocabulary instruction in a more balanced way. The two specific research questions to pursue are as follows:

(1) Are there differences between native English speakers and Korean EFL learners in their knowledge of SWs? How different are native speakers from EFL learners at different proficiency levels (high, mid, and low)?

(2) Are there differences between L1 users of English and L2 learners in their collocational knowledge? How different are native speakers from EFL learners with different levels of proficiency (high, mid, and low) in terms of their collocational competence?

3. Methods

3.1. Data

The data used in this study comprised two text corpora: an EFL learner (EFL-KR) corpus and a native speaker (Native-US) corpus. The learner (EFL-KR) corpus consisted of 150 argumentative essays written by Korean college freshmen for an English placement test. The essays were randomly selected from three (high, mid, low) proficiency groups (50 from each).

Table 1 shows the range of scores for each proficiency level. For instance, advanced learners' scores ranged from 15 to 16, which was the maximum score obtainable.

Table 1: Range of scores for the three proficiency levels in the EFL-KR corpus

Level	Score range	CEFR Level	N
High	15~16	C1	50
Mid	10~11	B2	50
Low	6~7	B1	50

The Native-US corpus was compiled from the LOCNESS (Louvain corpus of native English essays), a corpus constructed by the Université Catholique de Louvain (Granger, 1998b). LOCNESS was chosen because it contains academic English texts written by native speakers between the year 1900 and 2000.

Table 2 summarizes the number of tokens for the learner corpus (n = 46,190) and the native corpus (n = 46,791), respectively.

Table 2: Composition of the two corpora

EFL-KR corpus			Native-US corpus		
Level	**N**	**Token**	**Source**	**N**	**Token**
High	50	20,102	Argumentative essays from University of South Carolina	30	34,474
Mid	50	15,995			
Low	50	10,093	Argumentative essays from Presbyterian College, South Carolina	6	12,317
Total		46,190	Total		46,791

3.2 Data collection

For a comparison of vocabulary and collocation used in writing by EFL learners and native speakers, data were compiled from two different sources: a learner corpus and a native corpus. First, for the learner corpus, 150 texts were randomly selected from a collection of essays (n = 1,141) written by college freshmen for an in-house English placement test at a university in Seoul. Before the test, the students provided consent for their texts to be used for research purposes. For the test, the students were asked to take a position regarding the following statement and write an argumentative essay for 30 minutes: "Television, newspapers, magazines, and other media pay too much attention to the personal lives of famous people such as public figures and celebrities. Use specific reasons and details to explain your opinion".

The students' essays were then scored by English-speaking professors, who used a rubric developed by the director and coordinator of the English program at the university. The rubric was designed to measure four dimensions of writing: content, organization, accuracy, and vocabulary. The score for each area ranged from 1 to 4, and thus the maximum score the student could earn was 16 points. To ensure suitability for a comparative analysis of the three proficiency levels, essays that received a minimum of six points were selected because essays with scores lower than that were

roughly 50 words long. The students were generally more proficient than average EFL learners, as they had been admitted to one of the most prestigious colleges in Korea.

For a native speaker corpus, argumentative essays written by American university students were chosen, given that Korean schools and universities adopt American English as the norm. To make this comparable with the EFL-KR corpus (46,190 words), 36 essays (30 from the University of South Carolina and 6 from Presbyterian College, South Carolina) were chosen from the LOCNESS corpus so that its size (46,791 words) was almost the same. All the texts were argumentative essays, and thus identical in genre to the EFL learner texts. The topics, however, varied, as they were written on a variety of topics, such as gender discrimination, drug control, media and self-esteem, role of college, conflict and divorce, adolescent suicide, and water pollution. These texts were submitted by native speakers of English as college writing assignments, and thus their written output may have been different from the EFL learners' due to the nature of the tasks: writing assignments versus placement tests. Given that assignments are scored for course grades, and placement tests generally have lower stakes, hopefully, the psychological pressure associated with these tasks may not have been so different from the students' perspectives.

3.3 Research instruments

BNC-COCA25 Range Program

For lexical analysis, we used a vocabulary analysis program, BNC-COCA25 Range (Nation, 2012), loaded with 25 wordlists extracted from the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). The program can analyze up to 25 grades (25,000 words) by 1,000 words per grade. The analysis of the present study is confined to the first 20 grades, as the number of words that belonged to the higher levels (21 through 25) was found to be minimal in the preliminary analysis.

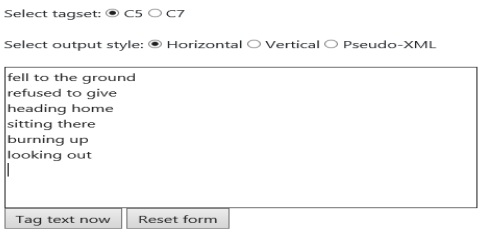
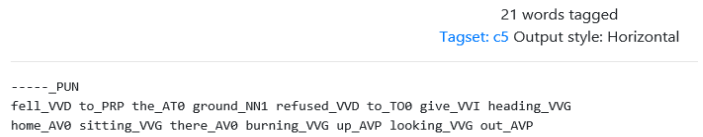
COCA_MWU20ColloGram

To compare the collocational diversity and distribution of the two text corpora, COCA_MWU20 ColloGram was used (Shin et al., 2018). The program is similar to the BNC-COCA Range program, in that it is loaded with 10,000 collocations (20 graded lists, 500 on each) extracted from the COCA and can analyze the collocational distribution by grade.

CLAWS Web Tagger

For an analysis of the part-of-speech (PoS) combination pattern of collocates, a program called CLAWS Web Tagger (University of Lancaster, n.d.) was used.

Figure 1: Part-of-speech tagging using CLAWS Web Tagger

Submitting Collocations into CLAWS Web Tagger	
Output of CLAWS Web Tagger	
Definition of 13 POS Tags	AJ: Adjective AV: Adverb AVP: Adverb particle (e.g., up, off, out) AVQ: WH-adverb (e.g., when, why) CJC: Conjunction DT: Determiner (e.g., a/an, the, this, these) NN: Noun PNQ: WH-pronoun (e.g., who, whoever) POS: The possessive (or genitive morpheme) 's or ' PRP: Preposition VB: Verb TO: Infinitive marker (e.g., to –infinitive) DTQ: WH-determiner (e.g., whose, which)

CLAWS Web Tagger is a free PoS tagging program with two types of tagsets: c5 and c7. For this study, the c5 tagset, a basic tagset (C5) with 62 tags, was adopted. It is sufficiently comprehensive as it was used for tagging the 100-million-word BNC (Garside, 1996). The tagset grouped different varieties of verbs, such as VVI (infinitive of lexical verb), VVD (past tense form of lexical verb), and VVG (-ing form of lexical verb) and merged them into one. Consequently, the 62 PoS were integrated into 13 parts of speech as shown in Figure 1 (Kim et al., 2020).

4. Results

4.1. Korean EFL learners and native English speakers: Knowledge of single words

To compare the lexical knowledge of the two groups, this study examined the distribution of single words used in the two written corpora: an EFL learner corpus (EFL-KR, 46,190 words) and a native speaker corpus (Native-US, 46,791 words). Table 3 summarizes the total number of words (tokens), types, and type-token ratios (TTR) for the native and the learner corpus with three levels: high, mid, and low.

Table 3: Lexical diversity of EFL-KR and Native-US corpora

Group		Token	Type	TTR
EFL-KR	High	20,102	2,468	0.12
	Mid	15,995	2,188	0.14
	Low	10,093	1,625	0.16
EFL-KR Total		46,190	3,936*	0.09
Native-US Total		46,791	5,501	0.12

Note: The number marked with an asterisk (*) is smaller than the sum of the high-, mid-, and low-level data because the same word, repeated across levels, is counted as one.

Given that TTR is a measure of lexical diversity, notably the words used in the native corpus (TTR = 0.12) were more varied than those of the learner corpus (TTR = 0.09). To check whether the two type-to-token proportions differed significantly between the native and EFL learners, a *z*-test was run to compare the two proportions with a web-based *z*-score calculator (Social Science Statistics, 2022). The type-token proportion of the native corpus (12%) was found to be significantly higher than that of the learner corpus (9%) in the test [z (df =1) = 16.34, $p < .001$]. That is, the TTR, a measure of lexical diversity, was significantly higher for native speakers than for EFL learners.

The differences were also noted in the comparisons of native speakers and EFL learners at mid and low levels. Specifically, the two type-to-token proportions were found to be different between the NSs and mid-level [z (df =1) = 6.40, $p < .001$], and between the NSs and low-level [z (df =1) = 11.96, $p < .001$] learners. However, in a comparison of advanced EFL learners with native speakers, it was found that advanced learners used diverse types (n = 2,468) of words for the size of their corpus (n =20,102). Considering that the advanced learner corpus size was 43% of the native speaker corpus, it is surprising that advanced learners did not differ much from their counterparts,

i.e., native speakers, in terms of lexical diversity, as shown by the identical TTR. The seemingly non-significant difference was confirmed by a z-test of the proportions from the two groups [z (df =1) = 1.91, p =.056]. The type-to-token proportion of the advanced learners' vocabulary was identical to that of the native speakers. This means that the advanced learners were as competent as the native speakers in terms of lexical diversity. This finding merits attention, as it deviates from a general assumption that EFL learners are inferior to native speakers with regard to their linguistic repertoire, including lexical knowledge. It is particularly notable that the TTR measures of the low- and mid-level students were higher than that of the native speakers, which was due to the size of the corpora. The small size of the two text corpora seems to have contributed to boosting the TTR measures.

For a more detailed analysis of the profiles of words from the corpora, a BNC-COCA 25 Range program was used, because it includes 25 wordlists (BNC-COCA 25) graded from 1K (the most frequent 1,000 words) through 25K according to frequency. Then, the lexical profiles of the two groups (native speakers and EFL learners) were compared across the 20 graded bands. Table 4 shows the distribution of words across the groups and word levels.

Table 4: Lexical profiles of EFL-KR and Native-US corpora in token (%)

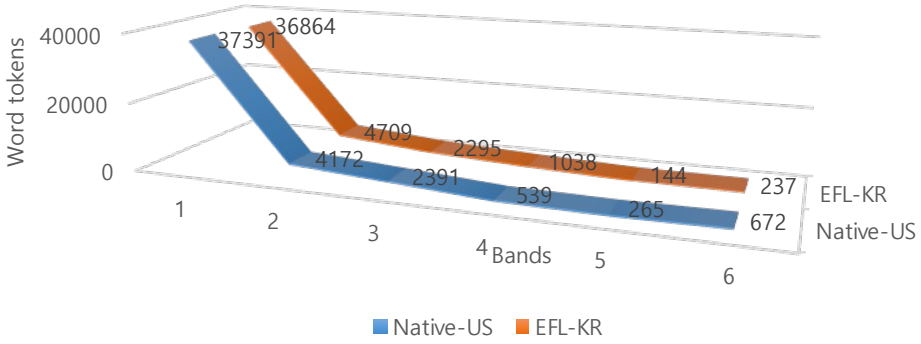
Level	Band	EFL-low	EFL-mid	EFL-high	EFL-KR	Native-US
One	1K	8,023 (79.49)	12,795 (79.99)	16,046 (79.82)	36,864 (79.81)	37,391 (79.91)
Two	2K	1,063 (10.53)	1,641 (10.26)	2,005 (9.97)	4,709 (10.19)	4,172 (8.92)
Three	3K	501 (4.96)	730 (4.56)	1,064 (5.29)	2,295 (4.97)	2,391 (5.11)
Four	4K	198 (1.96)	362 (2.26)	478 (2.38)	1,038 (2.25)	539 (1.15)
Five	5K	32 (0.32)	55 (0.34)	57 (0.28)	144 (0.31)	265 (0.57)
Six-twenty	6K and Above	37 (0.37)	81 (0.51)	119 (0.59)	237 (0.51)	672 (1.44)

Note: The shaded lists (6K and above) were combined for analysis; off-list words were excluded; thus, the total percentage for each column is not 100.

As seen in the table, level 1 through level 5 contain one thousand words frequently used at the respective level, and they are expressed as 1K band through 5K band, respectively. As there are few words beyond level six, level six through level 20 were combined for analysis, which was then categorized as the '6K and above' band. For this reason, the 6K and above band seems to contain more words than the 5K band, but in general, as the band goes up (e.g., from 1 to 2), the frequency of words declines.

The table also shows that the native corpus contained more words from the 5K and the 6K and above levels, which was almost double and triple the number of words in the learner corpus, respectively. However, interestingly, the lexical distribution of the two corpora displayed similar patterns (see Figure 2).

Figure 2: Word tokens: EFL-KR and Native-US corpora

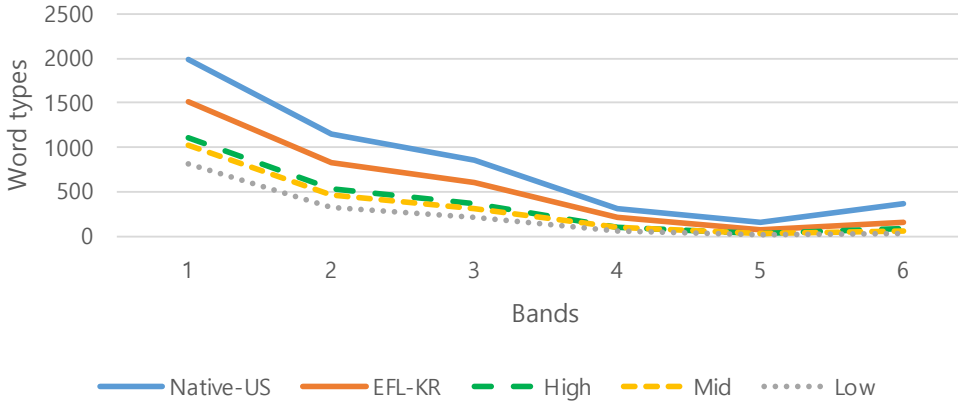


Considering that the two corpora were almost the same size and thus were directly comparable, the similar pattern that was observed deserves attention. First, regardless of the groups, about 80% of the words used were from the 1K band. In fact, more than 90% of the words were from the first two thousand words, regardless of group. Second, while the EFL learner corpus showed a steady decrease between the 2K and 5K bands and a plateau in the levels above, the native corpus displayed a steep decrease in the 4K band, and then a gradual increase in the subsequent levels. Third, native speakers used the 5K words about twice as much and the 6K words almost three times as much as the EFL learners. In other words, the differences are more pronounced in the 5K and 6K and above bands. With this exception, the lexical profiles or distributions were similar across the groups. This may have been because the students who contributed to the learner corpus were considerably competent in using productive vocabulary compared to average EFL learners, particularly the high-level learners.

Figure 3 shows the comparison of word types across the proficiency groups in relation to the native speakers. Types are used instead of tokens because the three

proficiency groups comprised of the EFL learner corpus, and thus were much smaller than the native corpus in terms of tokens.

Figure 3: Word types: High-, mid-, and low-level learners and native speakers



As seen in Figure 3, EFL learners used varied types of words up to 4K, and showed a gradual decrease thereafter in the learner corpus, although the native corpus showed slight increase from the 5K to 6K and above bands. The difference in learners' vocabulary use differed according to their proficiency, although the difference between the high and the mid-level learners was marginal. Interestingly, the differences across proficiency levels were larger in the 1K through 4K bands than in the sequential levels, such as the 5K and 6K and above bands, where the differences were minimal.

4.2. Korean EFL learners and native English speakers: Collocational knowledge

4.2.1 Collocational diversity and distribution

In addition to the learners' knowledge of single words, their knowledge of collocations was compared with that of native speakers for a holistic measurement of lexical competence. To analyze collocational diversity and distribution, the study used COCA_MWU20, loaded with 20 graded-multiword-unit (MWU) lists from the COCA, with each list containing 500 MWU families. Table 5 presents the tokens, types, and TTR of collocations used by native speakers and EFL learners in three levels: high, mid, and low.

Table 5: Collocational diversity of EFL-KR and Native-US corpora

Group		Token	Type	TTR
EFL-KR	High	798	350	0.44
	Mid	590	304	0.52
	Low	392	208	0.53
EFL-KR Total		1,780	670*	0.38
Native-US Total		1,749	1,091	0.62

Note: The number marked with an asterisk (*) is smaller than the sum of the high-, mid-, and low-level data because the same word, repeated across levels, is counted as one occurrence.

As seen in the table, the tokens and types of collocations in the learner corpus increased in proportion to learner proficiency. In other words, the more proficient the learners were, the more tokens and types of collocations they used. Compared to the learner corpus, the native speaker corpus contained a more diverse range of collocations. Although the collocation size of the native speaker corpus (n = 1,749) was approximately the same as that of the learner corpus (n = 1,780), there were far fewer types of collocations used by EFL learners (n = 670) than by native speakers (n = 1,091), and the difference was significant in the statistical comparison of the two proportions. The type-to-token proportion of the native corpus (62%) was significantly higher than that of the learner corpus (38%) in the z-test [z (df =1) = 14.70, p <.001]. This indicates that the native speaker corpus differed significantly from the learner corpus in terms of collocational variety, and that native speakers employed (about 1.6 times) more diverse types of collocations than EFL learners.

The study also examined whether there are statistical differences between the native speakers and EFL learners at different proficiency levels, and found significant differences between NSs and learners at all levels. The type-to-token proportions for the two groups were found to be statistically different [z (df =1) = 8.75, p <.001] with the proportion of the native speaker corpus (62%) being significantly higher than that of the advanced learner corpus (44%). The type-to-token proportions were also different between the NSs and EFL learners at the intermediate level [z (df =1) = 4.65, p <.001] or at the low-level [z (df =1) = 3.41, p <.001]. This shows that native speakers had a greater range of collocations than EFL learners of all proficiency levels.

From the two comparisons, we can see that EFL learners, regardless of proficiency level, used a more limited set of collocations than native speakers. The difference in the type-to-token proportions (62% vs. 38%) for the two groups manifests the disparity in their collocational competence. Although the difference was slightly smaller for

advanced learners versus native speakers (62% vs. 44%), the two corpora were still significantly different in terms of the type-token proportions, with the native speakers employing more diverse types of collocations than their counterparts. Table 6 shows how the collocations in the two corpora were distributed across the 20 graded lists.

Table 6: Collocational profiles of EFL-KR and Native-US corpora in token (%)

Level	Band	EFL-low	EFL-mid	EFL-high	EFL-KR	Native-US
One-two	1K	194 (49.49)	280 (47.46)	365 (45.74)	839 (47.13)	775 (44.31)
Three-four	2K	49 (12.50)	74 (12.54)	103 (12.91)	226 (12.70)	280 (16.01)
Five-six	3K	31 (7.91)	49 (8.31)	80 (10.03)	160 (8.99)	200 (11.44)
Seven-eight	4K	47 (11.99)	82 (13.9)	122 (15.29)	251 (14.10)	111 (6.35)
Nine-ten	5K	39 (9.95)	51 (8.64)	48 (6.02)	138 (7.75)	102 (5.83)
Eleven-twenty	6K and Above	32 (8.16)	54 (9.15)	80 (10.03)	166 (9.33)	281 (16.07)

Note: As each collocational level has 500 collocation families, two levels were combined to yield 1,000; Levels 1 and 2 were combined to make the 1K collocation family list. Likewise, Levels 3 and 4 (2K), 5 and 6 (3K), 7 and 8 (4K), 9 and 10 (5K), 11 to 20 (6K and above) were merged to make the collocation groups comparable to the wordlist groups; the shaded lists (6K and above) were combined for analysis.

In general, the frequency of collocations decreased as the collocation level increased, although there were occasional exceptions. In the learner corpus, the lower the learners' level, the higher the frequency of the 1st band collocations (High: 45.74%; Mid: 47.46%; Low: 49.49%). In addition, the frequency of the 2nd band collocation was consistently low in the EFL learner corpus regardless of proficiency (High: 12.91%; Mid: 12.54%; Low: 12.50%), as opposed to the native speaker corpus, which contained 16.01% of the 2nd band collocations. Notably, 42% (n = 740) of the collocations that the EFL learners used were from the first 500 collocation families. For instance, the collocate "too much" was used 208 times, taking up 28% of the 740 occurrences in the 1st level. Furthermore, the collocational distribution showed an unusual rise in level 7 or band 4 (level 7 and 8 combined) due to the writing topic. Namely, the students' use of topic-related collocates (such as "public figure") inflated the frequency of the 4th band or the 7th level collocation.

Figure 4 delineates the differences in collocational distribution between the native and learner corpora. The collocations of the native corpus were spread across all the levels and displayed a steady decrease from the 2K to 5K band, followed by a rise in the 6K and above band. By contrast, the learner corpus showed a drop from the 5K to 6K band and some fluctuations between the 3K and 5K band, as the number of collocations in percent drastically increased due to the frequent use of topic-related collocation, "public figure".

Figure 4: Collocation tokens: EFL-KR and Native-US corpora

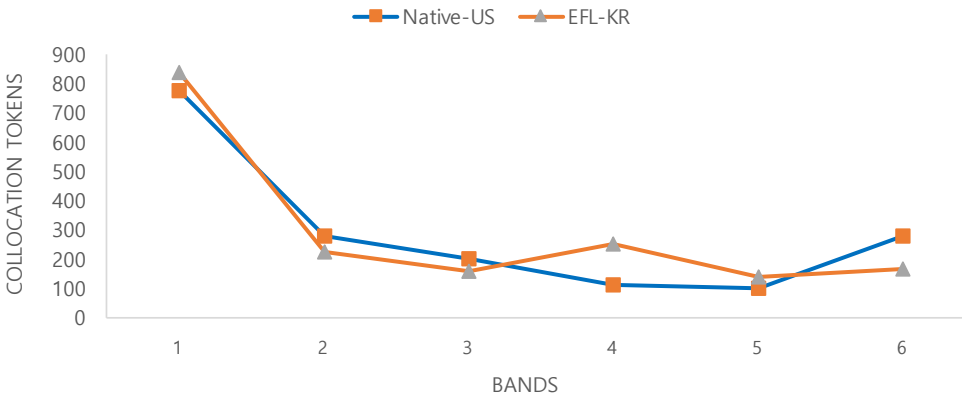
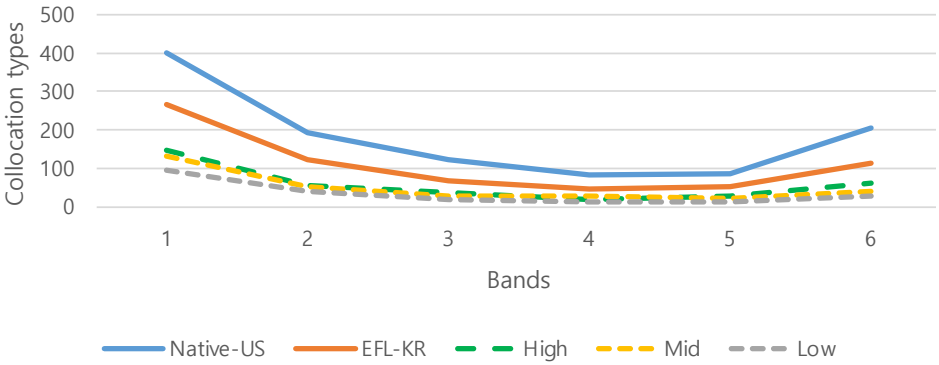


Figure 5 presents the distribution of collocations across the proficiency groups in comparison with the native corpus. It is particularly noteworthy that the EFL learners used far fewer collocations than single words and in a limited range only. The number of collocations used by the three proficiency groups ranged from 13 to 147; for the entire learner groups, it ranged from 53 to 266. This indicates that EFL learners used substantially less varied forms of collocations than native speakers, whose collocation types ranged from 84 to 400.

Figure 5: Collocation types: High-, mid-, and low-level learners and native speakers



Although the collocational distributions were seemingly similar up to the 5K band, the native speakers used far more diverse forms of collocations, whereas the three proficiency groups used collocations in a narrow range. In other words, regardless of proficiency, the three groups of learners used a limited range of collocations (Low: 13~94, Mid: 22~131, High: 18~147) across the graded bands.

4.2.2 Collocational patterns

In addition to the collocational diversity and distribution, collocational patterns were analyzed by tagging Part-of-Speech (PoS) combinations in each text corpus. As seen in Table 7, the native speaker corpus contained slightly more diverse combinations of collocations.

Table 7: Total number of POS combinations of collocations

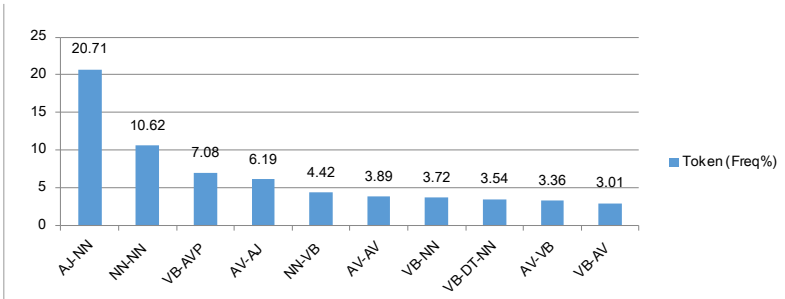
Native-US	EFL-KR	High	Mid	Low
108	95	61	65	55

Note: There were 86 overlaps in the POS combinations of collocations across the three levels.

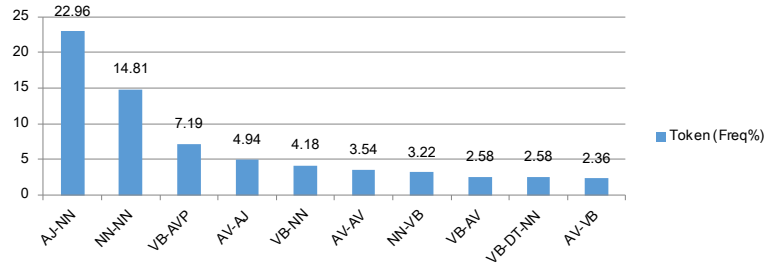
The number of PoS combinations extracted was 95 for the learner corpus, whereas it was slightly greater (n = 108) for the native speaker corpus. Figure 6 presents the distribution of different PoS combinations of collocations. Despite the differences between the learner and native speaker corpora in terms of collocational diversity and distribution, the collocational combinations displayed similar patterns. For example, the most frequent PoS combination was AJ-NN (e.g., "romantic comedy") both in the learner corpus (20.71%) and native corpus (22.96%), and the top four PoS combinations (AJ-NN, NN-NN, VB-AVP, and AV-AJ) were identical. Although there were slight differences in the frequency rank, the top 10 patterns were the same across the two corpora.

Figure 6: Comparison of top 10 collocational patterns: PoS combinations of the two corpora

EFL-KR



Native-US



The collocational patterns observed in the two corpora were in line with five of the six PoS combinations noted by Lewis (2000). Likewise, the five patterns proposed by McCarthy and O'Dell (2005) were also included in the top 10 PoS patterns. Consequently, the following seven patterns from Lewis (2000) and McCarthy and O'Dell (2005) matched the top 10 POS patterns in this study: AJ-NN (e.g., "bright color"), NN-NN (e.g., "radio station"), AV-AJ (e.g., "extremely inconvenient"), NN-VB (e.g., "economy boomed"), VB-AV (e.g., "smiled proudly"), VB-DT-NN (e.g., "submit a report"), and AV-VB (e.g., "happily married").

To summarize, as shown in the distribution of the top 10 collocation patterns, the PoS combinations of collocations were the same between the EFL learner and the native corpus. There were, however, some differences between the groups in the way they used combinations of words. For instance, the native speakers used the NN-NN combination more frequently, whereas the high-, mid-, and low-level EFL learners used the VB-, ADV-, and ADJ-associated collocational patterns more frequently than their counterparts, respectively. This finding may be because L2 learners tend to prefer certain combinations of high frequency collocations and overuse them (Durrant & Schmitt, 2009; García Salido & Garcia, 2018; Granger, 1998a).

5. Discussion

The study focused on comparing native speakers and L2 learners in terms of their knowledge of single words (research question 1) and collocations (research question 2). As expected, native speakers excelled against EFL learners in terms of lexical diversity. The EFL learners relied greatly on the first 2K high-frequency words, whereas the native speaker group used many low-frequency words from the 5K, and the 6K and above bands. In other words, the EFL group underused low-frequency words, unlike their native-speaking counterparts who utilized advanced vocabulary in addition to high-frequency words from the first 2,000 words.

In a comparison of the NSs and the learners at different proficiency levels, the native speaking counterparts outperformed the mid- and the low-level EFL students in

their lexical and collocational knowledge. This is not surprising as it is in line with the findings from earlier studies (Ädel & Erman, 2012; Dabbagh & Enayat, 2019; Doro, 2007; Lessard-Clouston, 2006; Siyanova-Chanturia & Schmitt, 2007). It is, however, interesting to note that advanced EFL learners were as competent as native speakers in terms of lexical variety. This finding departs from the general expectation that L2 learners, regardless of proficiency levels, are less skilled than NSs. In that regard, the non-significant difference between the NSs and advanced learners is promising, in that it signals the potential development of L2 learners' interlanguage in terms of lexis. In other words, L2 learners' knowledge of SWs can grow to the level of native speakers, and this growth can be facilitated with the advancement of internet technology (e.g., synchronous computer-mediated communication) and transnational mobility in the era of globalization (Duff, 2015; Kim & Kim, 2022; Ma, 2017). L2 learners are now able to readily access target language input on the Internet and easily travel to English-speaking countries. Consequently, they can gain sufficient exposure to L2 input and increased opportunities for interaction and communication. As there is room for development of L2 learners, we should be wary of regarding them as inferior to native speakers, let alone promoting native speakerism.

Despite their excellence in using a diverse range of words, the advanced learners were not as competent as native speakers in their collocational knowledge, as indicated in the significant difference between the two groups, and also corroborated by earlier studies (Ädel & Erman, 2012; Demir, 2017; Shitu, 2015). The difference was also found between the EFL learners as a whole and the native speakers. Also notable was that the learners, regardless of proficiency levels, used far fewer collocations than native speakers and in a narrower range. This result confirms the findings from Durrant and Schmitt (2009) who demonstrated that L2 postgraduate writers used substantially more collocations from the high-frequency bands in contrast to native speakers. This finding also supports Vedder and Benigno's (2016) claim that collocations are acquired late. In addition, the finding that the EFL learners lack collocational competence seems to be in line with Kim et al. (2020) who examined the collocational competence of Korean college students at high, mid, and low levels, and found that the distribution of collocations used by advanced learners was similar to the other groups' collocational distribution. This means that L2 learners, regardless of proficiency, used a limited repertoire of collocations, mostly confined to the 2K collocation families.

Tsai (2015) also reported that Taiwanese EFL learners used far fewer and less varied collocations than native speakers. As Granger (1998a) noted, learners tend to adopt a safe approach in their use of collocations, that is, using a small set of formulaic sequences that they are certain about. Likewise, Siyanova-Chanturia (2015) suggested that L2 learners have trouble using collocations and are inferior to native speakers in their collocational knowledge. This has been also addressed in other previous studies

(Kim et al., 2020; Kuo, 2009; Laufer & Waldman, 2011; Nesselhauf, 2005; Shitu, 2015; Vedder & Benigno, 2016). This difficulty may be because collocational development occurs gradually at a slow pace (Laufer & Waldman, 2011).

Considering that collocations are acquired later than single words (Vedder & Benigno, 2016), it is crucial to design a curriculum and teaching method that go beyond SWs. As Laufer and Waldman (2011) put, collocations are "a necessary component of second-language (L2) lexical competence in addition to the knowledge of single words" (p. 648). For this reason, Bestgen (2017) suggested that formulaic competence or "the native-like use of ready-made sequences of words" (p. 65) should be considered in L2 writing assessment, as it best predicted the quality of learner texts. Bestgen argued that formulaic competence has been neglected in automated L2 writing assessment programs, such as ETS' e-Raters. This negligence can have a harmful influence on teachers and learners alike. As Henriksen (2013) cautioned, teachers are likely to place less focus on collocations and fail to use materials effective for learner awareness raising for collocations.

Similarly, curriculum wise, little attention has been paid to collocation in the primary through secondary school context in Korea (Kim et al., 2020). Given that the non-native participants of this study were newly admitted college freshmen, it is lamentable that they learned English under the secondary school curriculum that trivialized the role of collocation. The Korean national curriculum seems to place more emphasis on SWs than collocations, in that it has specified wordlists but not collocation lists for different grades. This curricular negligence is manifest in the learning materials that the students use to study for the college SAT. Considering that the test-preparation materials often contain low-frequency idioms, collocations are not likely to be adequately level-based and evenly covered. In addition, classroom teaching focuses on preparing for the Korean SAT (KSAT), the college entrance exam (Kim, 2021). As the KSAT has a harmful washback on classroom instruction, particularly in the secondary school context, teachers only attend to the skill areas that are measured in the test. As knowledge of individual words suffices for students to excel on the test, they do not acknowledge the need for studying multi-words or collocations. However, as Bahns and Eldaw (1993) cautioned, tests that do not assess collocational knowledge do not yield comprehensive measures of learners' lexical competence. In particular, given that written discourse is composed of formulaic sequences (Ehrman & Warren, 2000), it is critical to teach and assess collocations along with lexis, and maintain a balance between word knowledge and collocational knowledge (Kim et al., 2020). Once collocational knowledge is measured in a language test, it will receive adequate attention from teachers and students, who will then change how they teach and learn for lexical competence development (Bestgen, 2017). This will eventually help to make a positive contribution to L2 writing proficiency development.

6. Conclusion

The present study is significant in that it analyzed L2 learners' use of single words and collocation in their writing, in comparison to that of native speakers. The study found significant differences between the native speaker and the whole learner group in both lexical and collocational knowledge. The difference between the advanced learners and the NSs was, however, significant only in their collocational competence. These findings are meaningful, as they offer holistic information about learners' knowledge of SWs and collocations. They also have pedagogical implications for curriculum design and vocabulary instruction and assessment. First, it is of paramount importance that single words across the graded lists be used in teaching and assessment. It was found from the lexical distribution that L2 learners used a limited range of words compared to native speakers. This indicates the need for sampling a diverse range of words across the graded lists and covering them in classroom instruction. Learners can then acquire words from the advanced bands, such as 5K and 6K and above, and increase their lexical repertoire. Moreover, curriculum designers should consider collocation an integral component of the curriculum and develop collocation lists in addition to wordlists. Once collocation is included in the curriculum as a core unit of learning, material designers should sample collocational expressions from different bands, so that a diverse range of collocations can be used in textbooks. These expressions should then be covered in classroom teaching to ensure that students be exposed to a vast array of collocational combinations. Once they become familiar with collocational expressions through repeated exposure, they are likely to reduce the number of collocational errors and acquire collocational knowledge. In addition, classroom tests should be constructed to measure collocational knowledge and provide complete information about learner's lexical competence (Bahns & Eldaw, 1993). In doing so, collocational expressions should be sampled across the graded lists, and various combinations of collocations should be used to ensure beneficial washback effect of the tests.

Notwithstanding the pedagogical merits of these findings, they are not without limitations. First, despite the attempt to equalize the size and text genre of the two corpora, they were not completely comparable, as the texts were written on different topics, and the nature of the tasks was different. The learner corpus was composed of texts from a timed-writing task in a test-taking situation whereas the native speaker corpus comprised college writing assignments. Thus, it is plausible that the difference in the nature of the tasks may have affected the students' written output. Another limitation is that the native corpus was available in its entirety without individual writers' demographic information, and thus it was not possible to run a statistical comparison of the two corpora in terms of the lexical and collocational distribution.

Thus, a follow-up study is recommended with two groups of text corpora that are tightly controlled in terms of topics, task type, and writers. Finally, as the non-native corpus was constructed with written texts by Korean college students who were generally more proficient than average college students, the findings of the study may not be directly applicable and generalizable to groups of learners in Korea or other EFL contexts. It should also be noted that the learner corpus was compiled from a context where the KSAT has a negative washback on the learning of collocations. Thus, the findings of the study should be interpreted with caution, and a further study should be deigned to resolve the limitations by controlling for the effects of the writer, tasks, topics, and learning contexts.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92.

Alejo-González, R. (2010). L2 Spanish acquisition of English phrasal verbs: A cognitive linguistic analysis of L1 influence. In M. C. Campoy, B. Belles-Fortuno, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 149-166). London, UK: Continuum.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics, 22*(2), 173–195.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System, 21*(1), 101–114.

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System, 69*, 65-78.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research, 18*(1), 54-74.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*, 30-49.

Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language, 22*(1), 181-200.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A. (2016). Reflecting on Coxhead (2000), "A new academic word list." *TESOL Quarterly, 50*(1), 181-185.

Coxhead, A., & Boutorwick, T. J. (2018). Longitudinal vocabulary development in an EMI international school context: Learners and texts in EAL, maths, and science.

TESOL Quarterly, 52(3), 588-610.

Cross, J., & Papp, S. (2008). Creativity in the use of verb + noun combinations by Chinese learners of English. In G. Gilquin, S. Papp, & M. B. Diez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 57–81). Amsterdam, Netherlands: Rodopi.

Crossley, S. A., & Salsbury, T. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics in Teaching, 49*, 1-26.

Dabbagh, A., & Enayat, M. J. (2019). The role of vocabulary breadth and depth in predicting second language descriptive writing performance. *The Language Learning Journal, 47*(5), 575-590.

Dang, T., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*, 66-76.

Demir, C. (2017). Lexical collocations in English: A comparative study of native and non-native scholars of English. *Journal of Language and Linguistic Studies, 13*(1), 75-87.

Doró, K. (2007). The use of high-and low-frequency verbs in English native and non-native student writing. In Z. Lengyel & J. Navracsics (Eds.), *Second language lexical processes: Applied linguistic and psycholinguistic perspectives* (pp. 117-129). Clevedon, UK: Multilingual Matters.

Duff, P. A. (2015). Transnationalism, multilingualism, and identity. *Annual Review of Applied Linguistics, 35*, 57-80.

Durant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching, 47*, 157-177.

Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text, 20*(1), 29-62.

Forsberg, F. (2010). Using conventional sequences in L2 French. *International Review of Applied Linguistics, 48*(1), 25-50.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 75-93). Harlow, UK: Longman.

García Salido, M., & Garcia, M. (2018). Comparing learners' and native speakers' use of collocations in written Spanish. *International Review of Applied Linguistics in Language Teaching, 56*(4), 401-426.

Garside, R. (1996). The robust tagging of unrestricted text: The BNC experience. In J. Thomas & M. H. Short (Eds.), *Using corpora for language research: Studies in honour of*

Geoffrey Leech (pp. 167–180). London. Longman.

González, M. C. (2013). *The intricate relationship between measures of vocabulary size and lexical diversity as evidenced in non-native and native speaker academic compositions.* [Unpublished doctoral dissertation]. University of Central Florida.

González-Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics, 166*(1), 94–126.

Gotz, S., & Schilk, M. (2011). Formulaic sequences in spoken ENL, ESL, and EFL. In J. Mukherjee & M. Hundt (Eds.), *Exploring second language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 79-100). Amsterdam: John Benjamins.

Granger, S. (1998a). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford: Oxford University Press.

Granger, S. (1998b). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). London: Addison-Wesley Longman.

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics, 52*(3), 229-252.

Gregori-Signes, C., & Clavel-Arroita, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia-Social and Behavioral Sciences, 198*, 546-556.

Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia, 11*(1), 1-20.

Henderson, A., & Barr, R. (2010). Comparing indicators of authorial stance in psychology students' writing and published research articles. *Journal of Writing Research, 2*, 245-264.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 29-56). Amsterdam: Eurosla.

Johnson, M. D., Acevedo, A., & Mercado, L. (2016). Vocabulary knowledge and vocabulary use in second language writing. *TESOL Journal, 7*(3), 700-715.

Kao, S. M., & Wang, W. C. (2014). Lexical and organizational features in novice and experienced ELF presentations. *Journal of English as a Lingua Franca, 3*(1), 49-79.

Kashiha, H., & Chan, S. H. (2015). A little bit about: Differences in native and non-native speakers' use of formulaic language. *Australian Journal of Linguistics, 35*(4), 297-310.

Kim, S.-Y. (2021). Developing disciplinary writing proficiency of pre-service English teachers; Writing-to-learn through online PBL. *Multimedia-Assisted Language Learning, 24*(1), 37-63.

Kim, S.-Y., Shin, D., & Kim, K.-S. (2020). Korean college students' use of vocabulary and collocation according to writing proficiency. *Multimedia-Assisted Language Learning, 23*(2), 215-235.

Kim, S.-Y., & Kim, K.-S. (2022). Vocabulary transfer from reading to writing: A comparison of essay writing and synchronous CMC. *TESL-EJ, 26*(1), 1-21.

Kuo, C. (2009). An analysis of the use of collocation by intermediate EFL college students in Taiwan. *ARECLS, 6*, 141-155.

Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*, 307-322.

Laufer, B., & Nation, I. S. P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook, 1*, 7-28. https://doi.org/10.1075/eurosla.1.05lau

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647-672.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British national corpus.* London: Longman.

Lessard-Clouston, M. (2006). Breadth and depth specialized vocabulary learning in theology among native and non-native English speakers. *Canadian Modern Language Review, 63*(2), 175-198.

Lewis, M. (1993). *The lexical approach.* Hove, UK: Language Teaching Publications.

Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach.* Hove, UK: Language Teaching Publications.

Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 2-46). London: Continuum.

Ma, Q. (2017). Technologies for teaching and learning L2 vocabulary. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 45-61). Hoboken, NJ: Wiley.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299-320.

McCarthy, M., & O'Dell, F. (2005). *English collocations in use* (5th ed.). Cambridge: Cambridge University Press.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57-78). EuroSLA Monograph 2. The European Second Language Association.

Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing, 47*. https://doi.org/10.1016/j.asw.2020.100511

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82.

Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved January, 17 2020, from http://www.victoria.ac.nz/lals/about/staff/publications/BNC_COCA_25000.zip

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242.

Nesselhauf, N. (2005). *Collocations in a learner corpus.* Amsterdam, Netherlands: John Benjamins.

Nurmukhamedov, U. (2017). The contribution of collocation tools to collocation correction in second language writing. *International Journal of Lexicography, 30*(4), 454–482.

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing, 22*, 545–565. https://doi.org/10.1007/s11145-008-9124-z

Paquot, M. (2007). Towards a productively-oriented academic word list. In J. Walinski, K. Kredens & S. Gozdz-Roszkowski (Eds.), *Practical applications in language and computers 2005* (pp. 127–140). Frankfurt & Main: Peter Lang.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32*, 130-149.

Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies, 7*(2), 105-125.

Read, J., & Nation, P. (2006). An investigation of the lexical dimension of the IELTS speaking test. *IELTS Research Reports, 6*, 207-231.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*, 329–363.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the vocabulary levels test. *Language Testing, 18*(1), 55-88.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503.

Shin, D. (2015). The influence of vocabulary regulations in the 2009 national curriculum of English on English textbooks. *The Journal of Foreign Studies, 34*, 47-76.

Shin, D., Chon, Y., Lee, S., & Park, M. (2018). A comparison of single word and multi-word unit profiles in spoken and written corpora of Korean learners and English native speakers. *Journal of the Korea English Education Society, 17*(2), 93-112.

Shitu, F. M. (2015). Collocation errors in English as a second language (ESL) essay writing. *International Journal of Cognitive and Language Sciences, 9*(9), 3270-3277.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*, 487-512.

Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System, 53*, 148-160.

Siyanova-Chanturia, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *IRAL, 45*, 119-139.

Siyanova-Chanturia, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. Canadian Modern Language Review, *64*(3), 429-458.

Social Science Statistics. (2022, July 1). *Z score calculator for 2 population proportions.* https://www.socscistatistics.com/tests/ztest/

Supasiraprapa, S. (2018). Second language collocation acquisition: Challenges for learners and pedagogical insights from empirical research. *The Journal of Asia TEFL, 15*(3), 797-804.

Tsai, K.-J. (2015). Profiling the collocation use in ELT textbooks and learner writing. *Language Teaching Research, 19*(6), 723-740.

University of Lancaster. (n.d.). *Free CLAWS web tagger.* http://ucrelapi.lancaster.ac.uk/claws/free.html

Vedder, I., & Benigno, V. (2016). Lexical richness and collocational competence in second language writing. *International Review of Applied Linguistics in Language Teaching, 54*(1), 23-42.

Waibel, B. (2008). *Phrasal verbs: German and Italian learners of English compared.* Saarbrücken, Germany: VDM.

Yu, A., & Trainin, G. (2022). A meta-analysis examining technology-assisted L2 vocabulary learning. *ReCALL, 34*(2), 232-252.

Zareva, A. (2007). Structure of the second language mental lexicon: How does it compare to native speakers' lexical organization? *Second Language Research, 23*(2), 123–153.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition, 27*(4), 567–595.

Teaching critical literacy in the classroom: A comparison of CLIL and EFL across contexts

Yolanda Ruiz de Zarobe
English, German and Translation & Interpretation Department
University of the Basque Country, UPV/EHU
yolanda.ruizdezarobe@ehu.es

Alba Gutiérrez
Fundación Universitaria Iberoamericana
alba.gutierrez.mtz@gmail.com

Abstract

Critical reading (Fairclough, 1989, 1995; Kress, 1985; Wallace, 2003) is considered to be one of the instruments that gives the necessary resources to construct meaning in discourse. This paper presents the results of a quasi-experimental study with a pre-test post-test design on students in the 6th year of primary education (10-11 years old) enrolled in two schools in Spain, one school located in the Basque Country (CLIL group) and the other in Cantabria (EFL group). Altogether four groups were selected: CLIL-experimental, CLIL-control, EFL-experimental and EFL-control. The experimental groups received critical reading strategy instruction for seven weeks while the control groups continued with regular classes. The one-way ANCOVA results showed that students in both the CLIL and EFL programmes developed greater mastery in critical reading after the teaching protocol. However, and contrary to our initial hypothesis, experimental students from the EFL context outperformed the CLIL learners. The novelty of the critical reading awareness protocol and motivation seemed to have a greater influence on the EFL students, who, unlike the CLIL students, were not so familiar with strategy learning. The new training procedure helped them to promote higher order critical reading skills, taking more advantage of the whole strategy learning protocol.

Keywords: critical reading, reading instruction, CLIL, EFL, learning strategies.

Resumen

La lectura crítica (Fairclough, 1989, 1995; Kress, 1985; Wallace, 2003) se considera uno de los instrumentos que proporciona los recursos necesarios para construir significado en el discurso. En este trabajo se presentan los resultados de

un estudio cuasi-experimental con un diseño pre-test post-test con alumnado de 6º curso de educación primaria (10-11 años) matriculado en dos colegios de España, uno situado en el País Vasco (grupo AICLE) y otro en Cantabria (grupo EFL). En total se seleccionaron cuatro grupos: AICLE-experimental, AICLE-control, EFL-experimental y EFL-control. Los grupos experimentales recibieron instrucción en estrategias de lectura crítica durante siete semanas, mientras que los grupos de control continuaron con las clases regulares. Los resultados del ANCOVA unidireccional mostraron que los alumnos de los programas AICLE y EFL desarrollaron un mayor dominio de la lectura crítica tras el protocolo de enseñanza. Sin embargo, y contrariamente a nuestra hipótesis inicial, los alumnos experimentales del contexto EFL superaron a los alumnos del programa CLIL. La novedad del protocolo de sensibilización a la lectura crítica y la motivación tuvieron una mayor influencia en los alumnos de EFL, quienes, a diferencia de los alumnos de AICLE, no estaban tan familiarizados con el aprendizaje de estrategias. El nuevo procedimiento de formación les ayudó a fomentar las destrezas de lectura crítica de orden superior, sacando más partido de todo el protocolo de aprendizaje de estrategias.

Palabras clave: lectura crítica, enseñanza de la lectura, AICLE, EFL, aprendizaje de estrategias.

1. Introduction

Critical literacy (Fairclough, 1989, 1995; Kress, 1985; Wallace, 2003) refers to a higher order thinking skill which considers the different mechanisms that are used by authors in order to influence readers. Researchers such as McLaughlin and DeVoogd (2004a) or Molden (2007) claim that critical reading is a constituent of critical literacy. Critical literacy begins with reading itself by helping learners of all ages to read in a more meaningful way and to become more actively involved in the process. This implies engaging the reader by making them ask questions such as the author's intention or the main argument that is being presented. It encourages the questioning of ideas by analysing and interpreting the texts, exercising their judgement about what they read and not taking anything at face value.

Reading from a critical perspective involves making connections between the text and the world, moving beyond the text itself to comprehend such issues as the author's intentions or opinions. For some scholars (Dozier, Johnston & Rogers, 2006), this involves understanding literacy as a tool for social action. It further encourages the use of power to construct meaning, implying that communication is a social act that can lead to social change (Comber & Simpson, 2001). However, when we focus on critical reading, we must first acknowledge that it can actually involve many different concepts, which range from merely understanding the definitions of words or the

structure of sentences to a more metacognitive approach (Bosley, 2008; Hermida, 2009; Li, 2016; McLaughlin & DeVoogd, 2004a) in which readers are expected to contrast different viewpoints, solving problems in order to understand how meaning has been constructed and making their own judgements based on what they read. This requires a deep and active engagement with the text (Fleming, 2020; Vaseghi, Gholami & Barjestech, 2012).

This article deals with critical reading in the case of second or foreign language learning. As Wallace (2003, p. 49) argues, "any educational activity must address the issue of what kind of knowledge is being transmitted or constructed in classroom settings". In relation to this, critical literacy does not consider language just as a set of explicit messages, but takes into account those hidden, where the critical reader takes an active role by questioning both the implicit and explicit information and reflecting on the content or processes involved (Endres, 2001; Palincsar & David, 1991). Thus, in the case of a foreign language, the reader may be more likely to be easily manipulated because of their lower language level or partial and/or total lack of cultural skills. In our study not only do we analyse the effect of a critical reading intervention on students learning English as a foreign language (EFL), but we also compare them with another cohort of learners following a content and language integrated learning (CLIL) approach in the learning of English with the aim of looking at the similarities and differences between them. As is well known, CLIL refers to those programmes that use a foreign or additional language in the teaching of content subjects (see Ruiz de Zarobe, 2016).

Before presenting the study, we will devote the following section to previous research that has addressed the importance of teaching critical reading in the L2 classroom and the different approaches to critical reading. We will then present the study, its objectives, research questions and hypotheses, as well as the methodology used. This is followed by a discussion of the results and the conclusions drawn from them. The article concludes with future lines of research, a description of some of the limitations and the pedagogical implications of the study.

2. Literature review

2.1. Critical reading research in the classroom

In this section we will describe some of the research that has been conducted on EFL reading from a critical stance. These studies can help us to understand some of the mechanisms and challenges of critical reading. Zabihi and Pordel (2011) analysed whether three popular reading textbooks designed for EFL intermediate (B2) and advanced (C1) learners promoted critical reading. In order to do so, three criteria

were considered: critical thinking items, the use of appropriate tasks, and strategic instruction. Their results showed that the three textbooks met the first criterion to some extent, but seriously lacked the other two.

A year later, Fahim, Barjesteh and Vaseghi (2012) analysed the effect of critical thinking strategy training on reading comprehension at university. In order to do so, 240 EFL students were selected and divided into two proficiency levels based on their TOEFL results. Each proficiency group was further subdivided into two cohorts: a critical and a non-critical group. The results suggested that critical thinking skills significantly affected EFL learners' reading comprehension performance. However, no effect of the critical thinking strategy training was found across proficiency levels. As the authors argue, these findings provide empirical support for the facilitative effect of critical thinking strategy training on the reading comprehension performance of EFL learners.

In a university context as well, Zin, Eng and Rafik-Galea (2014) conducted research on critical reading practices on learners of English from various disciplines in a Malaysian university. Their study analysed learners' ability to identify the author's purpose and the main idea of the text by asking critical reading questions. The outcomes showed that 66% of the students failed to identify the author's intention with the text and only 34% of the participants could correctly identify the main ideas. These results indicate that students lacked analytical and inference skills, implying problems with critical reading strategies.

To illustrate the effectiveness of the instructional approach on critical reading, argument analysis and metacognitive monitoring skills, Bensley and Spero (2014) compared three groups of university students who received different types of instruction: direct and explicit teaching of critical thinking through rule application, critical reading infused into the course and content knowledge acquisition. Compared to the other two groups, the first group showed significantly better results in the tests that measured argument analysis and critical reading skills, as well as improved metacognition. These results suggest that direct and explicit teaching of critical thinking skills can improve both critical reading and metacognitive knowledge.

Much of the research examining critical reading has been conducted in university settings and much of it also portrays a lack of critical reading skills on the part of students unless they are explicitly taught. In that respect, Azizi-Fini et al. (2015) arrived at similar results to those of Bensley and Spero (2014) in their study of critical thinking skills when they compared first-year students (a freshman at the university) and senior nursing students at Kashan University of Medical Sciences in Iran. Their results showed that both the freshmen and senior nursing students had low critical thinking

skills, and those skills did not change during their nursing degree. Furthermore, no correlation was found between their critical thinking skills and their age, gender, high school grade point average (GPA), rank in university entrance examination or interest in the nursing profession, which shows that no change is expected unless critical thinking is tackled explicitly and directly in the classroom.

On the other hand, part of the research has been concerned with exploring teachers' opinions and perceptions of critical reading. By way of illustration, Defianty and Wilson (2019) reported on the use of teacher questions to promote critical thinking in seven high school English classrooms in Indonesia. Their data showed that teachers often prompted thinking about the language (64% of questioning sequences) and through the language (36% of sequences) but they frequently missed the opportunity to promote higher order critical thinking skills.

More recently, Sutherland and Incera (2021) described faculty perceptions about critical reading skills and their usefulness for students. In general, faculty considered more complex skills (e.g., applying) as more useful, while simpler skills (e.g., skimming) were considered less useful. They also claimed to spend more time teaching the critical reading skills they considered to be more necessary. The authors acknowledge the importance of understanding the usefulness of teaching critical reading skills to improve learning in university settings.

The focus of the research and the context change completely in Díaz Iso et al. (2022), who, on a more general level, conducted a systematic review to analyse the wide range of interventions to improve reading literacy in primary school students. The following characteristics of the interventions were identified: (1) they were remedial, (2) they targeted students in disadvantaged situations, and (3) they focused on executive functions. However, this review was based on reading literacy, without a specific focus on critical reading, the basis of this study.

In general, research related to critical reading in the classroom defends the need for students to engage actively with the text and to be explicitly taught how to do so in order to become progressively more critical and autonomous when reading. However, much of the research conducted has been undertaken at university level, rather than in school settings, and in relation to EFL contexts. In fact, while there is little research analysing critical reading, even less has been conducted in CLIL settings comparing CLIL and EFL contexts. One of those studies was carried out by Nieto Moreno de Diezmas (2017), who examined how CLIL programmes affect the development of reading comprehension in the mother tongue, Spanish. In order to do so, she selected two groups of students, CLIL and EFL in secondary education (13-14 years old). The results showed that the acquisition of literal reading comprehension

and inferential reading comprehension in Spanish significantly benefitted from CLIL. Apart from that, global comprehension, lexical comprehension, understanding of space-time relationships, integration of extra-textual information and identification of extra-textual relations were the skills that benefitted the most. However, no significant differences were found in critical reading comprehension. The author attributes the positive results of the CLIL group to the potential transfer of reading strategies across languages and the importance of reading in CLIL settings. One year later, Nieto Moreno de Diezmas (2018) analysed the effect of CLIL on the provision of reading proficiency in Spanish as L1 in primary school. She also compared reading comprehension in two different cohorts, a CLIL group (bilingual group) and a non-CLIL group. The overall results indicated that the bilingual programme did not negatively affect the acquisition of reading proficiency in Spanish as L1, suggesting the transfer of strategies between languages. There were also no differences between the two groups in the acquisition of literal and inferential reading, although there were significant differences in favour of the non-CLIL group in the level of critical reading. This could be explained by the difficulty of reading in an L2, which may result in teachers paying more attention to the literal than to the critical comprehension of academic texts.

Finally, Ruiz de Zarobe and Zenotz (2018) report on an intervention in reading comprehension among learners of English as a third language (L3) in a multilingual (Spanish-Basque-English) context in the Basque Country. The study involved a pre-test post-test design, with an intervention of 7 weeks using two intact groups of participants that served as experimental and control groups in CLIL classrooms, where a number of subjects are taught through the L3: English. Findings indicated that reading awareness, reading comprehension, which included critical reading, and strategy use were enhanced following the training. While both the control and the experimental group improved over time, the experimental group showed statistically significant gains in reading comprehension. Furthermore, this difference was maintained over a two-year period, with the experimental group showing a marked improvement in reading competence. That study supported previous research (Ruiz de Zarobe & Zenotz, 2014), that examined if and how strategic instruction could provide the specific techniques needed for successful language learning in CLIL settings. More precisely, it analysed how reading strategic instruction could increase metalinguistic awareness through the teaching of critical reading strategies to plurilingual students of primary education in the Basque Country.

In sum, the literature on critical reading underscores the importance of engagement in order to enhance critical reading. This may be achieved through strategy instruction, which can be explicitly taught and learned in the classroom, as we will see later on.

2.2. Reading from a critical stance

In the previous sections we have seen how critical reading involves reflecting on the content of a text, scrutinising the author's opinion and interpreting the points of view they may have. In this respect, different authors give an account of approaching the reading process in different ways. For instance, in her framework for critical literacy, Janks (2010) follows three perspectives: (1) decoding, (2) reading with the text and (3) reading against the text. As the author claims, as critical readers we should be able first to understand the text, taking notes and paraphrasing the key points, in brief, "decoding" the text. "Reading with the text" involves interpreting it, giving a meaning to the text and interpreting and attaching our ideas and values to it. Lastly, "reading against the text" relates to the critical slant of the text, combining individual and social experiences to understand and challenge the ideologies behind it. In short, reading critically involves engaging with the text as an indispensable tool for success. "Essentially, critical literacy is about enabling young people to read both the word and the world in relation to power, identity, difference and access to knowledge, skills, tools and resources", according to Janks (2013, p. 227).

Hughes's model (2014), which is to some extent related to that of Anderson and Krathwohl's (2001), uses the image of a staircase to represent the five linear sub-skills that progress upwards in terms of difficulty in critical reading. The five sub-skills are as follows: (1) "understanding", which may be easier in the mother tongue but more difficult and time consuming in a foreign language, (2) "Applying" refers to the need to apply what has already been learned to new information. In language learning, it may refer to task completion using the information already gathered. The next three stages relate more directly to key higher-order critical thinking skills. The third step is (3) "analysing", where we study the text to understand how the author's information is presented, without taking it at face value. The following stage is (4) "evaluating", which refers to the ability to evaluate the relevance and validity of the author's opinions and views, differentiating fact and opinion. The final stage (5) "creating" involves applying your knowledge to create something new in new situations and contexts.

Both models are helpful to develop our students' ability to read critically. In this study, we share the ways they approach critical reading, in our case for young children. The next section will discuss our research for the improvement of critical reading which as Raman, Sharma and Collins (2013) claim, can be mainly achieved through the effective and regular application of reading strategies.

3. The present study

3.1. Objectives, research questions and hypotheses

The main objective of this paper is to investigate whether there is a connection between the teaching of critical reading strategies in the classroom and the improvement of critical reading skills in EFL and CLIL contexts. Apart from the effect of the teaching protocol, we are also interested in studying the differences between both learning contexts (EFL and CLIL).

Based on the objectives mentioned above, the following research questions were considered:

1. To what extent do students in a CLIL and in an EFL programmes develop a greater mastery in critical reading in L2 after an intervention on critical reading strategies?

   A significant relationship between the critical reading intervention and the development of critical reading is expected, following previous research (Ruiz de Zarobe & Zenotz, 2014, 2018, 2019; Bensley & Spero, 2014; Fahim, Barjesteh & Vaseghi, 2012), so that experimental students, regardless of the context, would outperform their control peers.

2. Do CLIL experimental students show a greater mastery of critical reading competence than EFL experimental students?

   Our hypothesis is that CLIL may have some positive effects on the acquisition of critical reading skills when compared to EFL. In other words, it is reasonable to expect that CLIL students would benefit more from the critical reading intervention than EFL students. There are different arguments that may support our hypothesis: firstly, in the case of CLIL settings, the integration of content goes hand in hand with the integration of the language of the discipline, which involves the activation of different cognitive and metacognitive processes (Ruiz de Zarobe & Smala, 2020). The cognitive effort that students need to make to learn the content and the language in CLIL settings may eventually enable them to take more advantage of the teaching protocol. Furthermore, CLIL is often associated with active and student-centred teaching methodologies (van Kampen et al., 2018).

3.2 Methodology

3.2.1. Participants and context of the study

The data in this study correspond to students in Year 6 of primary education (10-11 years old) enrolled in two schools in Spain.[1] One of the schools, from now on "CLIL school", was located in the Basque Autonomous Community (BAC). The Basque Country is one of the smallest autonomous communities in Spain but, with a population of over 2,000,000 inhabitants, it ranks eighth in terms of population. It is a bilingual community with two official languages, Basque, a pre-Indo-European language unrelated to any of the language families existing in Europe today, and Spanish. Both are official languages, but while Spanish is the majority language, Basque is the minority language of the community. However, Basque is widely spoken and taught in schools, and has a strong presence in the media and public institutions (Ruiz de Zarobe, 2015).

The other school, from now on "EFL school", was located in Cantabria, an adjoining region to the Basque Country in the north of Spain. Cantabria is a monolingual community, with Spanish as their official language.

Our study had a quasi-experimental pre-test post-test design with four groups: CLIL-experimental, CLIL-control, EFL-experimental and EFL-control. In Table 1 we can find the total number of participants for each of the groups. The total study sample included a percentage of 46.3 boys and 53.7 girls.

Table 1: Total number of participants

Group	N	Percentage
CLIL		
Experimental	52	25%
Control	75	36%
EFL		
Experimental	54	26%
Control	26	13%
TOTAL	207	100%

[1] In Spain there are 17 autonomous communities which have authority on education. In the case of our two schools in both communities, all primary school students followed the same educational protocol, without any selection for CLIL groups. This is not always the case in secondary education.

The basic level of English in Year 6 both in the Basque Country and Cantabria is A2 according to the Common European Framework of Reference for Languages (CEFR). This level implies that students can deal with everyday situations, use common phrases and talk about familiar subjects. No placement test was conducted prior to the intervention but, as we will see in the data analysis section, the pre-test score was included in the analysis as a covariate to control for possible initial differences.

In both schools, the students filled out a biographical questionnaire (see Appendix 1) which, among other questions, collected information about parents' studies as well as about their work, in order to find out about the social context from which the participants came.

The CLIL school was a private school partly subsidised by the Government. The social background of the students at the school can be said to be middle class. It was noted that 71% of the participants' fathers and 73% of the mothers had university degrees.

Table 2: Level of education of the participants' parents from CLIL school

Group	Frequency	Percentage
Fathers		
Primary/Secondary	9	6,5%
Baccalaureate	25	18%
University	105	75.5%
TOTAL	139	100%
Mothers		
Primary/Secondary	5	3.5%
Baccalaureate	26	18.5%
University	110	78%
TOTAL	139	100%

The EFL school was also a private school partly subsidised by the Government. With regard to the socio-economic level of the participants' families, we can see that it is also medium, as most of the parents have completed higher education or baccalaureate studies. There is a slight difference between the contexts of Cantabria and the Basque Country: while in Cantabria the fathers have higher levels of education, in the Basque Country it is the mothers who have a higher level.

Table 3: Level of education of the participants' parents from EFL school

Group	Frequency	Percentage
Fathers		
Primary/Secondary	12	8.3%
Baccalaureate	44	30.3%
University	89	61.4%
TOTAL	145	100%
Mothers		
Primary/Secondary	23	16.1%
Baccalaureate	42	29.4%
University	78	54.5%
TOTAL	143	100%

In the case of the CLIL school, the students' L1 was mainly Spanish in terms of general proficiency, but a high percentage of them used both Spanish and Basque in the school context (both in and outside the classroom). Furthermore, pupils had all been learning English formally at school since they were 4 years of age, In fact, students learned each of the three languages: Basque, Spanish and English for at least 20% of the teaching hours (6 hours per week). Our sample received 3 hours per week of EFL, and science, physical education and arts and crafts were also taught through English. Basque, music, religion, and tutorials were carried out in Basque, and all other subjects (Spanish and mathematics) were undertaken in Spanish.

In the EFL school, formal instruction only involved EFL 3 times a week. Table 4 shows the time of instruction in each context.

Table 4: Level of education of the participants' parents from EFL school

	CLIL school[2] Hours per week	EFL school Hours per week
Pre-primary	9 hours EFL/CLIL	1 hour EFL
Primary Education Grades 1-2	4 hours EFL 4.5 hours CLIL	2.5 hours EFL

[2] As can be seen, the "CLIL group" not only learns English in the content classes, but also receives more formal EFL classes. In fact, we could have called this group "CLIL+EFL group". However, we have limited ourselves to calling it "CLIL group" for the sake of simplicity.

Primary Education Grades 3-4	4 hours EFL 4.5 hours CLIL	3 hours EFL
Primary Education Grades 5-6	4 hours EFL 4.5 hours CLIL	3 hours EFL
TOTAL	2808 hours	826.5 hours

3.2.2. Treatment procedure

In each context, intact classes of Year 6 of primary education were selected. These classes were randomly assigned to the experimental groups and to the control groups. The experimental groups received instruction on critical reading strategies in English during EFL classes for 7 weeks (1 hour per week), and the control groups continued with regular classes.

The critical reading programme was developed by the research team following the Cognitive Academic Language Learning Approach (CALLA), a model that has been widely used for strategy instruction (see, for instance, Chamot & Harris, 2019). This model has different phases: preparation/awareness raising, presentation/modelling, practice, evaluation and expansion/transfer, which allow for both content and language learning. The aim of this model is for the teacher to gradually shift the responsibility towards the student.

The strategies were selected because they focus specifically on understanding and properly evaluating the information read. These strategies help students to have a clearer understanding of what the text says and what it really means. As a result, when students know more about how to interpret a text, they become critical readers and thinkers. The strategies that were selected were:

- Strategy 1. Distinguish true from false: The aim of this strategy is to help students to discern authentic from biased information. It involves evaluating the validity of the information presented in a text, identifying the contradictions. It helps learners to avoid being misled by false content.

- Strategy 2. Identify the main ideas: Students discover the most important ideas, finding the relevant points. It requires using keywords to express central ideas. It helps learners to understand the purpose and organization of the text.

- Strategy 3. Distinguish facts from opinions: Students distinguish factual and objective information from subjective statements. It helps readers to recognise the relevance of a statement.

- Strategy 4. Discover the author's intentions: Students come to understand what the writer intended when creating a text. It helps learners to comprehend the implications of the author's message.

Strategy instruction followed a similar pattern for the seven sessions. Firstly, the strategies were worked on individually and, during the last sessions, combined to make the training more effective (Wharton-McDonald & Swiger, 2009). The pattern for strategy instruction included:

1. Awareness raising of strategies and strategy use, which involved explicit information on why it is important to use strategies and the advantages of their use for students.[3]

2. Modelling by the researcher with explicit information on how, when and why to use the strategy.

3. Practice that included individual, pair or group work through different tasks and games in which learners initially practised the strategies with scaffolding support before the scaffolding was gradually removed under the supervision of the researcher.

4. Finally, an assessment of the effectiveness of the strategy that incorporated a self-assessment learning diary of each of the seven training sessions.

The intervention was carried out by the researchers in English, although Spanish was used in some cases to clarify comprehension. The teaching sessions took place during the EFL lessons for the experimental groups. Meanwhile, the CG continued with their regular classes and were in no way aware of any kind of strategic intervention.

Although this intervention took place in Year 6 of primary education, learners had already worked with reading strategies in English the year before with the same students (see Ruiz de Zarobe & Zenotz, 2018, for a review of the intervention protocol in Year 5). However, we believed it was a suitable moment now to begin with critical reading strategies once they had already learned other reading strategies.

[3] Explicit presentation and practice of critical reading skills has been shown to be beneficial for students' development of critical thinking (Abrami et al., 2008).

In order to show how critical reading interventions were undertaken in Year 6, we present an example using one of the strategies: "Discover the author's intentions". Students were told that authors can sometimes show their values and intentions in a very obvious way, but other times they are ironic and say one thing and mean the opposite. This can become a more effective tool to attract the reader's attention. Some examples were presented to the students, as shown below:

A picture shows a brother and a sister fighting inside a car and the mother replies to the children:

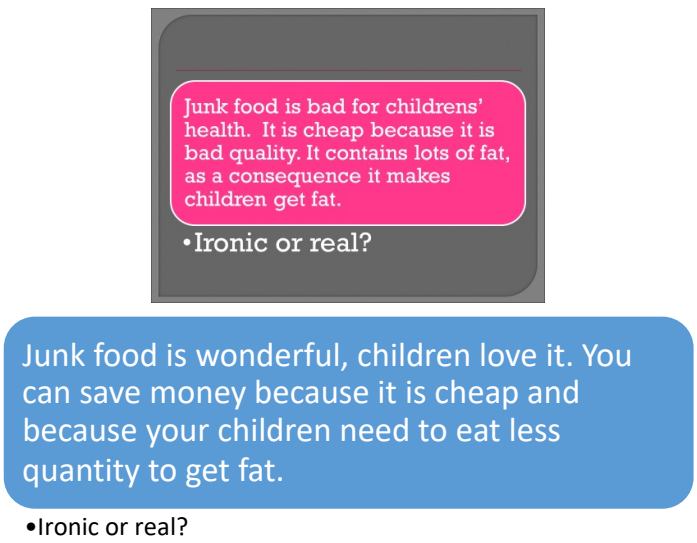
Apparent intention: Congratulate children for fighting.

- *That's very nice. You're behaving like a big girl and a big boy. Fantastic!!*
- *Continue like this and you will go to Disneyland in summer.*

Real intention: Make children stop fighting.

- *That's wrong. Stop behaving like this, kids!!*
- *Continue like this and you will not go to Disneyland in summer.*

Figure 1: Example of a strategy: "Discover the author's intentions"



After raising the students' awareness of the strategy, which involved explicit information on the benefits of strategy use, students worked with several ironic texts (see Appendix 2 for an example) where they had to discover the real intention of the author and what he or she initially seemed to stand for. The possibility to engage actively in what they read, analysing the author's intentions, evaluating what is being said, and even questioning and challenging the arguments that were being presented made students take a critical stance, contributing to effective reading comprehension. In our view, this engagement with the text is a way to encourage critical reading strategies, to promote a critical approach and turn young readers into critical readers.

Furthermore, in order to avoid any instructional effect, the researchers followed a didactic protocol so that both the EFL and CLIL experimental groups received the same explicit, systematic and scaffolded instruction on critical reading skills. This instructional protocol involved the group's perception of the relevance, usefulness and applicability of critical reading skills.

3.2.3. Materials and procedure

In order to answer our research questions, a critical reading test (see Appendix 3) was administered before and after the strategy intervention.[4] The test was administered by one of the researchers in the classroom during normal lessons. Both the control and the experimental groups completed the test in the pre- and post-test phases, which took place a week before and after the intervention, and otherwise received the usual curriculum for 6th year of primary education. The experimental group, however, also followed a strategy intervention, spanning three months a year. The strategy intervention consisted of seven sessions in which the students learned and practiced the critical reading strategies.

Week 1 and 2. Distinguish true from false.
Week 3. Identify the main ideas.
Week 4 and 5. Distinguish facts from opinions.
Week 6 and 7. Discover the author's intentions.

The critical reading test consisted of the reading of a text about child soldiers. The title was: Welcome to the web site of the coalition to promote the use of child soldiers. The text advocated the use of children as soldiers, but it was obviously a hoax. The test had two parts: the first part had seven open-ended questions that students had to answer using only the information from the text. Two examples are:

[4] All tests had been piloted by the research team prior to the intervention in order to suit the level of language proficiency and prior knowledge of the learners.

- *Write two key words to summarize the text.*
- *Where has this idea (the idea to recruit children under the age of 18 as soldiers) been applied?*

The second part consisted of nine open-ended questions that students had to answer using information from the text and about the text. Some examples include:

- *Who has written the text?*
- *What is the author's intention?*
- *Can you trust this information? Are there any elements to distrust this text?*

Instructions were given in English and participants had 22 minutes to answer both parts, dividing the time as they preferred. Each question was given 1 point so that the maximum score they could obtain was 16.

As can be seen, the protocol undertaken approaches reading practices in a similar way as that presented, for instance, by Janks (2010), and discussed in the introduction to the study. Students needed to be able to "decode" information, which requires linguistic knowledge and proficiency, they needed to "read with the text", interpreting the written word and, quite significantly, they needed to "read against the text", providing a deeper critique of the text, challenging what is said.

3.2.4. Data analysis

Once the data were collected, they were analysed using SPSS (Statistical Package for Social Sciences). For reliability index, Cronbach's alpha was used. The reliability index was 0.862 and, therefore, the internal consistency and reliability of the test were high (Larson-Hall, 2015).

Normal distribution was also studied to best decide on the statistical model for the analysis. The values of skewness and kurtosis were explored to consider the normality of the sample. As the values of these two measures were between 1 and -1, we considered that our sample was normally distributed (George & Mallery, 2010; Paltridge & Phatiki, 2015) and, therefore, it was possible to conduct parametric tests.

To analyse the data, a one-way ANCOVA was conducted to determine statistical differences. The independent variable was the group (EFL-experimental, EFL-control, CLIL-experimental and CLIL-control) and the dependent variable was the score on the critical reading post-test. The score of the pre-test was included in the analysis as a covariate to control for possible initial differences. The fact that the pre-test was used as a covariate removes variability. That is, in ANCOVA, the dependent variable is the post-test measure. The pre-test measure is not an outcome, but a covariate. Therefore, this model assesses the differences in the post-test means after accounting for pre-test values.

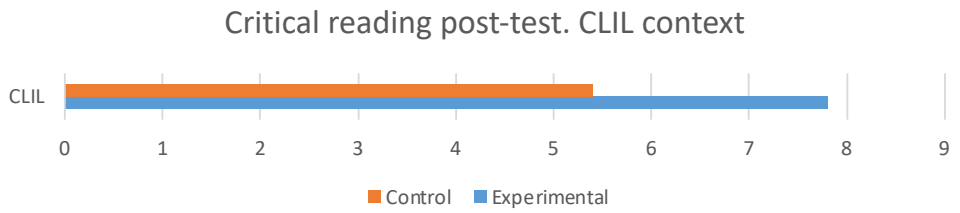
To meaningfully interpret the univariate F test for the different groups, we determined whether any statistical assumptions underlying the use of the ANCOVA were violated in the dataset. An examination of Levenne's Test of equality of error of variance of the score was carried out, and the assumptions of the ANCOVA were met. Results are discussed in the following section.

4. Results

4.1. Research Question 1: To what extent do students in a CLIL and in an EFL programmes develop a greater mastery in critical reading in L2 after an intervention on critical reading strategies?

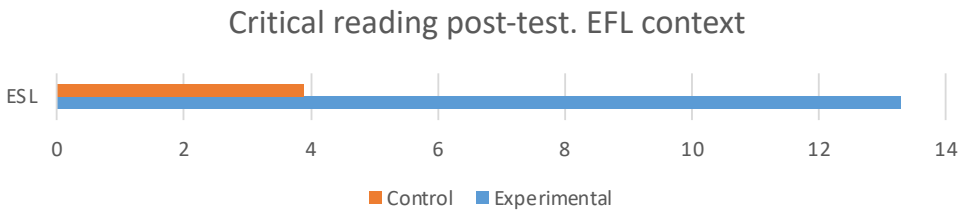
In answer to this question, the results of CLIL and EFL programmes were studied independently. Firstly, we are going to consider the CLIL context. CLIL experimental students scored a marginal mean of 7.8 out of 16 possible points, whereas the control students obtained a marginal mean of 5.4 (see Figure 2). The marginal mean is the mean adjusted for the covariate. As a result, CLIL experimental students scored higher than their peers in the critical reading post-test, and the difference of 2.4 points was considered significant ($p<0.000$).

Figure 2: Example of a strategy: "Discover the author's intentions"



On the other hand, focusing on the EFL context, experimental students scored a marginal mean of 13.3 and the control group obtained 3.9 (see Figure 3). This difference was considered significant ($p<0.000$).

Figure 3: Results of the critical reading post-test. EFL context



In sum, the results of the analyses related to the first research question indicate that the groups (CLIL and EFL) receiving instruction on critical reading outperformed control groups on the post-test.

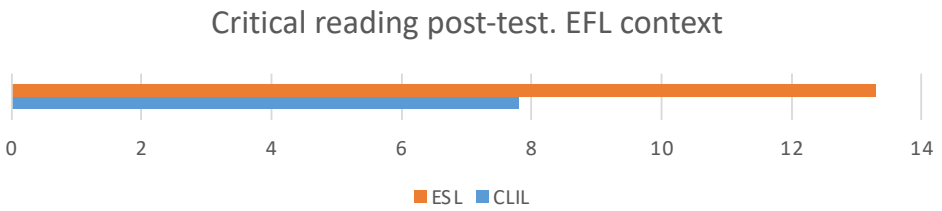
The next research question considers differences between experimental groups to study whether the treatment is more effective for one group of students.

4.2. Research question 2: Do CLIL experimental students show a greater acquisition of critical reading competence than EFL experimental students?

In order to answer this question, we compared the results of the CLIL-experimental and the EFL-experimental groups. Our hypothesis suggested that CLIL experimental students would outperform EFL-experimental students.

If we compare the marginal means of CLIL-experimental and EFL-experimental groups, we see that the former scored 7.8 and the latter 13.3 (see Figure 4). This difference was found to be significant ($p<0.000$) in favour of EFL experimental students. Therefore, our hypothesis was not confirmed as, unexpectedly, experimental students from the EFL context outperformed CLIL learners. Several reasons for this result are suggested in the next section.

Figure 4: Results of the critical reading post-test. Experimental groups



Finally, if we compare the results of the control groups, CLIL and EFL, there were no statistically significant differences between them ($p<0.088$).

5. Discussion

From the analysis of the data of this study, it can be inferred that the critical reading strategic intervention is an effective tool to improve critical reading both in EFL and CLIL contexts (research question 1). Our results show statistically significant differences between experimental and control groups in CLIL and EFL, which indicates that the intervention has been successful in both contexts.[5] This positive

[5] Although no placement test was performed, the fact that the pre-test score was included in the analysis as a covariate controlled for possible initial differences.

result is consistent with previous findings (Ruiz de Zarobe & Zenotz, 2014, 2019; Azizi-Fini et al., 2015; Bensley & Spero, 2014; Fahim, Barjesteh & Vaseghi, 2012), which have demonstrated the importance of teaching explicitly critical thinking and critical reading strategies.

This study also shows the importance of supporting students in the learning process by providing them with appropriate scaffolding. As Wilson and Devereux (2014, p. 98) argue, "scaffolding is not merely a synonym for support, and that while scaffolding involves support, it is the nature of that support that is crucial. Rather than 'dumbing down' the curriculum, we maintain that scaffolding involves challenging students to make leaps forward into their ZPD. Intellectual challenge must be high, while explanations must be explicit in terms of what is expected, how to achieve it and why it is important".[6] As Wilson and Devereux (2014) claim, scaffolded tasks require high challenge and high support, with practice in the skills, with the necessary feedback from the teacher and with support in their mother tongue, if necessary. In brief, this is the protocol we have followed in our study, which has provided positive outcomes irrespective of the pedagogical approach.

In addition, the results indicate that EFL experimental students benefitted significantly more than their CLIL peers (research question 2). When the two experimental groups were compared, we found out that there were also statistically significant differences between CLIL-experimental and EFL-experimental, in favour of the latter. It seems that our teaching protocol was more effective in the EFL context. In general, as commented above, studies show that CLIL students have better metacognitive awareness (Ruiz de Zarobe & Zenotz, 2018; Coyle, Hood & Marsh, 2010; Meyer et al., 2015). In fact, the European Commission (2014) claims that "CLIL students may also gain in cognitional development and other aptitudes. CLIL presents an opportunity to students for using knowledge learned in one context as a knowledge base in other contexts because it helps learners to apply, integrate and transfer knowledge while fostering critical thinking (Duverger, 1995 cited in Gravé-Rousseau, 2011)". Furthermore, Meyer (2010), when dealing with quality CLIL methodology, claims that learning skills and strategies "(...) are the pillars of CLIL learning and their potential for promoting language as well as higher order and critical thinking skills has long been neglected" (p. 16).

However, our study presents a different scenario: while it is true that in the CLIL context there was significant improvement in the experimental group, our results indicate that EFL students benefitted significantly more. This unexpected and somewhat surprising outcome could be attributed to the "novelty effect" (Marek,

[6] ZPD stands for "Zones of Proximal Development" (Vygotsky, 1978), whereby Vygotsky addresses the potential of a learner, how much they can do, with the assistance from others.

2019; Marek & Wu, 2021), whereby the results may to some extent be a consequence of the introduction of something new in the process. The reason might lie in the fact that this teaching protocol was a completely new way of working for this group of EFL learners, who were not so familiar with strategy learning or critical reading strategies. That is, the EFL group may have been more motivated and interested by the new intervention, which involved a more systematic and scaffolded teaching of critical reading skills. Therefore, the students became more engaged in the tasks, improving their attention and retention of critical reading skills. The novelty effect may have also triggered more curiosity and feedback-seeking behaviours in the EFL group, which may have facilitated their metacognitive and self-regulatory processes.

Apart from the novelty effect, other factors may have played a role in these results, especially motivation. Previous research in this field had successfully linked motivation level and strategy use in second languages (Schmidt & Watanabe, 2001; Gutiérrez & Ruiz de Zarobe, 2019). In fact, Gutiérrez and Ruiz de Zarobe (2019) demonstrated that the level of motivation played a significant role in improving students' performance when they underwent metacognitive reading instruction. Those students who reported higher levels of motivation obtained significantly better results than those with lower levels of motivation. This suggests that motivation is an important factor for the success of students who undergo this type of instruction. In addition, students' motivation to learn languages increased when topics to which they could relate were addressed and when adequate scaffolding was provided. This is also the case for the experimental groups in this study, whose teaching protocol was also effective.

This can also be connected with the conclusions drawn from Meyer (2010) and Tedick and Lyster (2020), who suggest that students can greatly benefit from scaffolding, even in the case of students with low proficiency. Furthermore, it seems that their motivation increases as they understand better how and in which order to proceed, which could also explain how our EFL sample improved so much. As the reading process was explicitly handled, techniques on how to approach texts were explained, modelled, taught, and feedback was given, EFL learners, despite having fewer hours of instruction than their CLIL peers, made the most of the teaching protocol, practising these newly-learned strategies and applying them to new activities.

6. Conclusion

This study points to the effectiveness of a reading intervention in improving critical reading regardless of the pedagogical approach followed, both in EFL and CLIL contexts. The seven-week intervention followed by both cohorts proved to be effective in improving critical reading. Furthermore, our results showed that, even though students may sometimes have had fewer hours of foreign language instruction,

the results can still be positive in their use of critical strategies. Factors such as the novelty effect and motivation may account for some of these results.

Apart from this, these results relate to models such as those of Janks (2010) or Hughes (2014), discussed above, which uses the image of a staircase to represent the different stages that progress in terms of difficulty in critical reading and thinking. To progress, teachers must engage students in activities that can turn them into "open-minded, active, strategic readers who are capable of viewing text from a critical perspective" (McLaughlin & DeVoogd, 2004b, p. 56). In sum, teachers can help students to develop their capacity to read critically. When students are encouraged to learn from a critical stance, they will end up reading critically.

Therefore, these results have an important pedagogical implication. It shows how it may be interesting for teachers to become aware of the relevance of this type of interventions, in which explicit and scaffolded strategy instruction should be part of the protocol. The steps to be followed move from the presentation and explanation of the new strategy, the demonstration of its effectiveness, individual or group practice of the new strategy by the students and a final evaluation of the effectiveness of the strategy. This study suggests that it may be worth allocating some time to teaching strategies in the classroom if it leads to significant outcomes in the learning process.

However, this research suffers from some limitations, which should be pointed out. The first is the difficulty involved in developing research in the classroom in two different geographical contexts. Although the schools were located in similar socio-economic areas in both geographical contexts, the Basque Country and Cantabria, it is possible that some conditioning factors beyond the location itself may have had an impact. In order to avoid these, a fixed teaching protocol was designed and followed for both cohorts, piloting the tests in order to reduce the differences, but other factors may have had an effect on these results. The second limitation relates to the fact that, although both cohorts were in the same school year, no placement test was conducted in both contexts, which did not allow us to control for proficiency level as such. Other factors not taken into consideration include, for example, the gender of the students, which was outside the scope of this study. These are two of the main limitations of this study that could be addressed in future research.

Needless to say, there is still much that needs to be done in the teaching and learning of reading and thinking critically. The study presented here is just a small sample of the type of interventions that can be carried out to foster critical reading, and by extension, critical thinking in the classroom. We have seen how critical reading involves engaging in the text without taking what we read at face value. Both critical reading and critical thinking provide the foundations of what sound learning is and

should therefore be encouraged from a very early age. Future research on the subject will help us to delimit with greater clarity the paths to follow in order to proceed in the learning and teaching of reading, and more precisely, of reading with a critical stance, a domain that undoubtedly deserves to be investigated in today's world.

Acknowledgments

We would like to express our gratitude to the principals, teachers and students who have taken part actively in this project.

We acknowledge the funding of the research projects FFI2015-63715-P and PID2021-122689NB-I00 financed by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe", and IT1426-22 (Department of Education, Basque Government).

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102-1134. https://doi.org/10.3102/0034654308326084

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition.* New York: Longman.

Azizi-Fini I., Hajibagheri, A., & Adib-Hajbaghery, M. (2015). Critical thinking skills in nursing students: a comparison between freshmen and senior students. *Nursing and Midwifery Studies, 4*(1), e25721. https://doi.org/ 10.17795/nmsjournal25721.

Bensley, D. A., & Spero, R. A. (2014). Improving critical thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity, 12*, 55-68. https://doi.org/10.1016/j.tsc.2014.02.001

Bosley, L. (2008). "I don't teach reading": Critical reading instruction in composition courses. *Literacy research and instruction, 47*(4), 285-308, https://doi.org/10.1080/19388070802332861

Chamot, A. U., & Harris, V. (Eds.). (2019). *Learning strategy instruction in the language classroom: Issues and interpretation.* Bristol: Multilingual Matters. https://doi.org/10.21832/9781788923415

Comber, B., & Simpson, A. (2001). *Negotiating critical literacies in classrooms.* New York: Routlege. https://doi.org/10.4324/9781410600288

Coyle, D., Hood, P., & Marsh, D. (2010). CLIL– *Content and language integrated learning.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009024549

Defianty, M., & Wilson, K. (2019). Fostering critical thinking through questioning in EFL: An Indonesian study. In L. Li (Ed.), *Thinking skills and creativity in second language education* (p. 74-94). Abingdon, Oxon: Routledge. https://doi.org/10.4324/9781315098920-5

Díaz Iso, A., Velasco, E. & Meza, P. (2022). Interventions carried out to improve reading competence: A systematic review. *Revista de Educación, 398*, 249-281. https://doi.org/10.4438/1988-592X-RE-2022-398-559

Dozier, C., Johnston, P., & Rogers, R. (2006). *Critical literacy critical teaching: Tools for preparing responsive teachers.* New York: Teachers College Press.

Endres, B. (2001). A critical read on critical literacy: From critique to dialogue as an ideal for literacy education. *Educational Theory, 51*(4), 401-413. https://doi.org/10.1111/j.1741-5446.2001.00401.x

European Commission. (2014). *Improving the effectiveness of language learning: CLIL and computer assisted language learning,* 1-36. Retrieved from: https://ec.europa.eu/education/content/improving-effectiveness-language-learning-cliland-computer-assisted-language-learning_en

Fahim, M., Barjesteh, H., & Vaseghi, R. (2012). Effects of critical thinking strategy training on male/female EFL learners' reading comprehension. *English Language Teaching, 5*(1), 140-154. https://doi.org/10.5539/elt.v5n1p140

Fairclough, N. (1989). *Language and power.* London: Longman.

Fairclough, N. (1995). *Critical discourse analysis.* London: Longman.

Fleming, G. (2020) *What does critical reading really mean?* Retrieved from: http://www.thoughtco.com/critical-reading-basics-1857088

George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference 17.0 update* (10th edition). Boston: Pearson.

Gravé-Rousseau, G. (2011). L'EMILE d'hier à aujourd'hui: une mise en perspective de l'apprentissage d'une discipline en langue étrangère. In *Charte européenne du plurilinguisme,* 1-10. Retrieved from: http://www.emilangues.education.fr/files/parrubriques/L_EMILE_d_hier_a_aujourdhui_G_Grave-Rousseau.pdf.

Gutiérrez, A., & Ruiz de Zarobe, Y. (2019). Training CLIL students to be strategic readers: Is motivation important? In A. Jiménez-Munoz, & A. C. Lahuerta Martínez (Eds.), *Empirical studies in multilingualism* (pp. 47-71). Bern, Berlin, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang.

Hermida, J. (2009). *The importance of teaching academic reading skills in first-year*

university courses, 1-12. Retrieved from: https://ssrn.com/abstract=1419247

Hughes, J. (2014). *Critical thinking in the classroom*, 1-27. Retrieved from: https://cdn.ettoi.pl/pdf/resources/Critical_ThinkingENG.pdf

Janks, H. (2010). *Literacy and power.* New York: Routledge.

Janks, H. (2013). Critical literacy in teaching and research. *Education Inquiry, 4*(2), 225–242. https://doi.org/10.3402/edui.v4i2.22071

Kress, G. R. (1985). *Linguistic processes in sociocultural practice.* Deakin, Australia: Deakin University Press.

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R.* New York: Routledge. https://doi.org/10.4324/9781315775661

Li, L. (2016). Integrating thinking skills in foreign language learning: What can we learn from teachers' perspectives? *Thinking Skills and Creativity, 22*, 273–288. https://doi.org/10.1016/j.tsc.2016.09.008

Marek, M. W. (2019). *Understanding the novelty effect: How long should an experimental intervention last?*, 1-2. Retrieved from: https://www.researchgate.net/publication/336831965_Understanding_the_Novelty_Effect_How_long_should_an_Experimental_Intervention_Last

Marek, M. W., & Wu, W. C. V. (2021). Motivational affordances of the novelty effect in TELL. In N. Chen & M. Chang (Eds.), *2021 International Conference on Advanced Learning Technologies (ICALT)* (pp. 268–269). IEEE. https://doi.org/10.1109/ICALT52272.2021.00086

McLaughlin, M., & DeVoogd, G. L. (2004a). *Critical literacy: Enhancing students' comprehension of text.* New York: Scholastic Teaching Resources.

McLaughlin, M., & DeVoogd, G. L. (2004b). Critical literacy as comprehension: Expanding reader response. *Journal of Adolescent and Adult Literacy, 48*, 52-62.

Meyer, O. (2010). Towards quality-CLIL: successful planning and teaching strategies. *Pulso: Revista de Educación, 33*, 11-29. https://doi.org/10.58265/pulso.5002

Meyer, O., Coyle, D., Halbach, A., Schuck, K., & Ting, T. (2015). A pluriliteracies approach to content and language integrated learning - mapping learner progressions in knowledge instruction and meaning-making. *Language, Culture and Curriculum, 28*(1), 41–57. https://doi.org/10.1080/07908318.2014.1000924

Molden, J. (2007). Critical literacy: The right answer for the reading classroom. Strategies to move beyond comprehension for reading improvement. *Reading Improvement, 44*(1), 50-56.

Nieto Moreno de Diezmas, E. (2017). How does CLIL affect the acquisition of reading comprehension in the mother tongue? A comparative study in secondary

education. *Investigaciones Sobre Lectura, 8*, 7-26. https://doi.org/10.37132/isl.v0i8.214

Nieto Moreno de Diezmas, E. (2018). Adquisición de la lectura en L1 en programas bilingües de Educación Primaria. Un estudio comparativo. *Ocnos. Revista De Estudios Sobre Lectura, 17*(1), 43-54. https://doi.org/10.18239/ocnos_2018.17.1.1471

Palincsar, A., & David, Y. (1991). Promoting literacy through classroom dialogue. In E. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 122-140). New York: Teachers College Press.

Paltridge, B., & Phatiki, A. (2015). *Research methods in applied linguistics.* London: Bloomsbury.

Raman, M., Sharma, S., & Collins, G. (2013). *Communication for engineers.* Oxford: Oxford University Press.

Ruiz de Zarobe, Y. (2015). Basque Country: Plurilingual education. In P. Mehisto, & F. Genesee (Eds.), *Building bilingual education systems: Forces, mechanisms and counterweights* (pp. 97–108). Cambridge: Cambridge University Press.

Ruiz de Zarobe, Y. (2016). *Content and language integrated learning: Language policy and pedagogical practice.* New York: Routledge.

Ruiz de Zarobe, Y., & Zenotz, V. (2014). Strategic instruction in primary education: A pathway to successful learning in content-based contexts. In R. Breeze, C. Llamas, C. Martínez, & C. Tabernero (Eds.), *Integration of theory and practice in CLIL* (pp. 37-54). Amsterdam: Rodopi. https://doi.org/10.1163/9789401210614_004

Ruiz de Zarobe, Y., & Zenotz, V. (2018). Learning strategies in CLIL classrooms: how does strategy instruction affect reading competence over time? *International Journal of Bilingual Education and Bilingualism, 21*(1), 1-13. https://doi.org/10.1080/13670050.2017.1391745

Ruiz de Zarobe, Y., & Zenotz, V. (2019). Effective learning interventions in young children: The impact of critical reading strategies. In J. Rokita-Jaskow & M. Ellis (Eds.), *Early instructed second language acquisition: Pathways to comprehension* (pp. 142-152). Bristol: Multilingual Matters. https://doi.org/10.21832/9781788922517-011

Ruiz de Zarobe, Y., & Smala, S. (2020). Metacognitive awareness in language learning strategies and strategy instruction in CLIL settings. *Journal for the Psychology of Language Learning, 2*(2), 20-35. https://doi.org/10.52598/jpll/2/2/3

Sutherland, A., & Incera, S. (2021). Critical reading: What do faculty think students should do? *Journal of College Reading and Learning, 52*(2), 113-129. https://doi.org/10.1080/10790195.2021.1887777

Schmidt, R., & Watanabe, Y. (2001). Motivation, strategy use, and pedagogical preferences in foreign language learning. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation*

and second language acquisition (pp. 313-359). Honolulu: Univer¬sity of Hawaii, Second Language Teaching and Curriculum Center.

Tedick, D., & Lyster, R. (2020). *Scaffolding language development in immersion and dual language classrooms.* New York: Routledge. https://doi.org/10.4324/9780429428319

Van Kampen, E., Admiraal, W., & Berry, A. (2018). Content and language integrated learning in the Netherlands: teachers' self-reported pedagogical practices. *International Journal of Bilingual Education and Bilingualism, 21*(2), 222-236. https://doi.org/10.1080/13670050.2016.1154004

Vaseghi, R., Gholami, R., & Barjestech, H. (2012). Critical thinking: An influential factor in developing English reading comprehension performance. *Advances in Asian Social Science, 2*(1), 401-410.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard: Harvard University Press.

Wallace, C. (2003). *Critical reading in language education.* London: Palgrave Macmillan. https://doi.org/10.1057/9780230514447

Wharton-McDonald, R., & Swiger, S. (2009). Developing higher order comprehension in the middle grades. In S. Israel, & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 510-530). New York: Routledge. https://doi.org/10.4324/9781315759609-38

Wilson, K., & Devereux, L. (2014). Scaffolding theory: High challenge, high support in Academic Language and Learning (ALL) contexts. *Journal of Academic Language and Learning, 8*(3), A91-A100. Retrieved from: https://journal.aall.org.au/index.php/jall/article/view/353

Zabihi, R., & Pordel, M. (2011). An investigation of critical reading textbooks: A qualitative analysis. *International Education Studies, 4*, 80-87. https://doi.org/10.5539/ies.v4n3p80

Zin, Z., Eng, W. B., & Rafik-Galea, S. (2014). Critical reading ability and its relation to L2 proficiency of Malaysian ESL learners, *The Southeast Asian Journal of English Language Studies, 20*(2), 43-54. https://doi.org/10.17576/3L-2014-2002-04

Appendix 1: Bibliographical questionnaire.

POR FAVOR, CONTESTA A ESTAS PREGUNTAS

1. NOMBRE: _____ 2. CURSO: _____

3. CHICA ☐ CHICO ☐ 4. EDAD: _____

5. ¿VIVES EN SANTANDER/VITORIA?
SÍ ☐ NO ☐ ¿DÓNDE VIVES? _____:_____

6. ¿CUÁL ES EL TRABAJO DE TUS PADRES?
PADRE_____
MADRE_____

7. ¿QUÉ ESTUDIOS TIENEN TUS PADRES? PON UNA 'X' EN LA CASILLA QUE CORRESPONDA

	PRIMARIOS	BACHILLERATO	UNIVERSITARIOS
PADRE			
MADRE			

8. ¿HAS PARTICIPADO EN UN CAMPAMENTO DE INGLÉS EN LOS ÚLTIMOS AÑOS?
SÍ ☐ NO ☐ ¿DÓNDE? _____
¿CUÁNDO? _____

9. ¿HAS ESTADO ALGUNA VEZ EN UN PAÍS DE HABLA INGLESA?
SÍ ☐ NO ☐ ¿DÓNDE? _____
¿CUÁNDO? _____
POR FAVOR, EXPLICA CUÁNTO TIEMPO TE QUEDASTE EN ESOS PAÍSES Y SI FUISTE A CLASES DE INGLÉS _____

10. ¿VAS A CLASES EXTRAESCOLARES DE INGLÉS?
SÍ ☐ NO ☐ ¿DÓNDE? _____
¿CUÁNDO? _____
POR FAVOR, EXPLICA CUÁNTAS HORAS A LA SEMANA HAS DADO O DAS INGLÉS FUERA DEL COLEGIO: _____

Appendix 2: Example of a strategy. "Discover the author's intentions".

My visit to Honduras

Source: www.pbs.org/now/shows/309/pericles-students.html

The first week in Honduras was eye-opening for me. They say that the poverty is impressing since Honduras is the second poorest nation in Central America, but this was not important for me because I stayed in an expensive area.

It is true most houses are in bad condition, rubbish covers the ground, and people's appearances don't take a top priority, but we also saw a few very nice houses, which are hidden behind huge walls. Crime rates are high and security guards are everywhere, therefore this is not a problem if you can pay security. It is common for both men and women to be visibly armed with a weapon. At first I was impressed when I saw a mother walking along the street with her children, carrying a rifle. By the end of the trip, it didn't surprise me, it was a fantastic opportunity to practise with arms, I would like to do that in my country.

Hospital conditions in Honduras were far worse than any I'd ever seen. The emergency room was one large, open room, over-crowded with hospital beds and people. The patients were given just two meals per day and often ate nothing for over 12 hours.Fortunately, this only happens in public hospitals. If something happened to me, I would go to a private clinic as they have very good services and real luxury!!

When I heard how high malnutrition is in Honduras, I was surprised to see so many adults who looked healthy or even fat. I saw some children who looked thin sitting beside parents who looked like they hadn't been hungry a day in their lives. I discovered that one of the main causes of malnutrition is not eating a healthy diet. Many Hondurans don't receive the proper nutrients they need because they only consume soda and junk food.

From failure to eat a healthy, balanced diet, many of the people are dying. I think this is not really serious because they do not look thin. Malnutrition is also the reason children looked so much younger than they actually were. Predicting a child's age was impossible—most four-year-olds were equivalent to the size of a baby in Europe. In my opinion this is not a real problem because looking younger is fantastic!

1. Write the key words for each paragraph.

2. Write the key words for the whole text.

3. What is the main idea in the text?

4. Who has written the text?

5. Where was the text published?

6. What type of text is it?

7. Can you trust this information? Are there any elements to distrust this text?

8. Apparently what are the author's values? (Find sentences with his/her opinion)

9. In reality what are the author's values?

10. Apparently what is the author's intention?

11. In reality what is the author's intention?

Appendix 3: Critical reading test.

Welcome to the web site of the Coalition to Promote the Use of Child Soldiers

Source: http://www.boring.ch/childsoldiers/why.html

We are a not for profit organization dedicated to helping humanity by encouraging countries around the world to use children between the ages of 0 and 18 as soldiers.

Our primary objectives are to educate the citizens and governments of the world that the use of child soldiers can be a benefit to societies everywhere, by helping children to help their communities - and giving them food, education, practical skills, an income, and maturity at the same time.

Around the world the citizens of developing countries face a life of poverty and oppression. Many of these citizens are children. The world's population is growing at exponential rates, meaning that every day there are more and more hungry mouths to feed.

What can be done?

The answer is simple: Recruit children under the age of 18 as soldiers.

Already, around the world, children and their governments are taking an active role in solving these problems by volunteering or recruiting children into the army. By doing so, not only does the country's army have the added benefit of thousands of little hands helping fight for the good of the citizens, but the children themselves learn valuable lessons, and the population is reduced in an efficient way.

This group of chidren in Nicaragua are having lots of fun, and learning too, by participating in the ongoing was in their nation

TASK: Read the text and the questions (in English or Spanish)

PART 1 QUESTIONS IN THE TEXT: Preguntas EN el texto (IN THE FOLLOWING QUESTIONS USE ONLY THE INFORMATION IN THE TEXT) (EN LAS PREGUNTAS SIGUIENTES USA SÓLO LA INFORMACIÓN DEL TEXTO)

1. What does this organization propose? *¿Qué propone la organización?*

2. What are the advantages of the proposal for the children? *¿Cuáles son las ventajas de la propuesta para los/las niños/as?*

3. Does the proposal have any disadvantage? Which? *¿Tiene la propuesta alguna desventaja? ¿Cuál?*

4. Why are there so many hungry people?*¿Por qué hay tanta gente hambrienta?*

5. Where has this idea been applied?*¿Dónde se ha puesto en marcha esta idea?*

6. What does "the population is reduced in an efficient way" mean? *¿Qué quiere decir "the population is reduced in an efficient way"?*

7. Write 2 key words to summarise the text. *Escribe 2 palabras clave que resuman el texto:*

TASK: Read the text and the questions (in English or Spanish)

THE FOLLOWING QUESTIONS USE MORE THAN WHAT APPEARS IN THE TEXT) (EN LAS PREGUNTAS SIGUIENTES UTILIZA MÁS INFORMACIÓN ADEMÁS DE LA QUE APARECE EN EL TEXTO)

1. Who has written the text?

2. Some of the protagonists' voices are not heard. Whose?

3. Where was the text published?

4. What type of text is it?

5. Can you trust this information? Are there any elements to distrust this text?

6. Apparently what are the author's values?

7. In reality what are the author's values?

8. Apparently what is the author's intention?

9. In reality what is the author's intention?

Online processing and vocabulary learning in massed versus spaced repeated reading

Raquel Serrano
Dept. Modern Languages and Literatures and English Studies
University of Barcelona
raquelserrano@ub.edu

Ana Pellicer-Sánchez
Department of Culture, Communication & Media
UCL Institute of Education
University College London
a.pellicer-sanchez@ucl.ac.uk

Abstract

The present exploratory study analyzed the effect of two different implementations of repeated reading (massed vs. spaced) on the online processing of a new text including some unknown vocabulary, as well as on incidental vocabulary learning. The relationship between processing and learning of target vocabulary in massed versus spaced repeated reading was also examined. A group of Spanish undergraduate students read the same text in English three times under either a massed (three times consecutively) or spaced (once a week) schedule, while their eye movements were recorded with an eye tracker. Knowledge of a set of 12 target words from the text was assessed through multiple-choice meaning-recognition tests. The results of the statistical analyses examining processing of the whole text suggest that repeated reading led to a significant decrease in total reading time and number of fixations in the massed group but not in the spaced group. However, in the analyses at the word level, no significant differences were found between the two conditions in the processing of the target words or immediate vocabulary gains. Finally, it was observed that longer processing time on the target words was related to larger immediate vocabulary gains, but only in the spaced group.

Keywords: repeated reading, input spacing, online processing, vocabulary learning, eye-tracking.

Resumen

El presente estudio exploratorio analiza el efecto de dos implementaciones diferentes de lectura repetida (concentrada vs. espaciada) en el procesamiento del texto así como en el aprendizaje incidental de nuevo vocabulario. También se examina la relación entre procesamiento y aprendizaje. Un grupo de estudiantes universitarios españoles leyó el mismo texto en inglés tres veces de forma consecutiva y otro grupo de forma espaciada (una vez a la semana), mientras se registraban sus movimientos oculares. Una prueba de selección múltiple evaluó el reconocimiento del significado de 12 palabras meta incluidas en el texto. Los análisis estadísticos sobre el procesamiento del texto completo muestran que la lectura repetida condujo a una disminución significativa en el tiempo total de lectura y en el número de fijaciones en el caso de la lectura repetida concentrada, pero no en el de la espaciada. Sin embargo, en los análisis a nivel de la palabra no se encontraron diferencias significativas entre las dos condiciones de lectura repetida ni en el procesamiento ni en el aprendizaje a corto plazo de nuevo vocabulario. Finalmente, se observó que un mayor tiempo de procesamiento estaba relacionado con un mejor aprendizaje de vocabulario a corto plazo, pero solo en el grupo espaciado.

Palabras clave: lecturas repetidas, distribución temporal, procesamiento, aprendizaje de vocabulario, seguimiento ocular.

1. Introduction

Developing reading fluency in English as a foreign language is crucial in many educational contexts, especially those that use English as the medium of instruction, which are becoming increasingly popular worldwide, but also in regular EFL contexts. Reading fluency has been defined in many different ways; however, under most definitions it entails fast and efficient processing of a written text (Grabe et al., 2015, p. 75). The most common component of reading fluency is reading rate, usually operationalized as words per minute. However, beyond reading rate, changes in physical aspects of reading (e.g., fixations, fixation durations) are also important to examine the development of reading fluency (Nation, 2009). Eye-tracking has been increasingly used as an objective measurement of reading fluency, providing rich information about processing effort during reading.

Considering the importance of reading fluency, researchers and practitioners have explored different ways to help learners process written texts more efficiently. One of these instructional approaches is repeated reading, which involves reading the same text multiple times. Research has shown that repeated reading leads to faster reading rates, in both the L1 and the L2 (Chang & Millet, 2013; Therrien,

2004). Additionally, research in L1 reading using eye-tracking has shown that repeated reading has a facilitation effect in processing efficiency, as manifested through shorter fixations, fewer fixations, and longer saccades (Hyönä & Niemi, 1990; Raney & Rayner, 1995).

While the examination of eye movements can also inform about changes in L2 reading behavior and processing efficiency (Conklin, Pellicer-Sánchez, & Carrol, 2018), no previous study has explored how reading patterns change when rereading the same text multiple times. Eye-tracking research in L2 reading has mostly focused on processing of unknown vocabulary, rather than processing of whole texts. The evidence coming from this line of research suggests that the processing of repeated exposures to novel words becomes less effortful as repetitions increase (Elgort, Brysbaert, Stevens, & Van Assche, 2018; Godfroid et al., 2018; Pellicer-Sánchez, 2016) and that the amount of attention given to novel items seems to be related to vocabulary gains (e.g., Godfroid, Boers, & Housen, 2013; 2018; Pellicer-Sánchez, 2016). It remains to be known whether these patterns are also found in the context of repeated reading.

Apart from helping learners read faster and more efficiently, L2 repeated reading has also been found to foster incidental vocabulary learning (e.g., Liu & Todd, 2016). Considering that repeated exposure to novel words is necessary for incidental learning to take place (Nation, 2013; Uchihara, Webb, & Yanagisawa, 2019), reading the same text several times guarantees that learners will encounter the same words multiple times.

When analyzing any type of repeated L2 practice, it is important to consider inter-repetition spacing, as it could affect both learning processes and products (Koval, 2022; Suzuki, Nakata, & DeKeyser, 2019). Research in cognitive psychology as well as in SLA has shown that massed repetitions are processed more easily because of recency effects; however, this lack of processing effort can have negative consequences in terms of learning outcomes (Koval, 2019). Although there are some studies that have analyzed vocabulary learning through repeated reading over several sessions under differently spaced schedules (Serrano & Huang, 2018, 2023), no previous studies have examined how reading patterns change when L2 learners reread the same text repeatedly in one versus several sessions, or how reading patterns affect vocabulary learning. Exploring how different inter-repetition spacing affects reading speed and incidental vocabulary learning in the context of repeated reading can provide interesting insights on the role of input spacing in L2 practice. Additionally, findings from such research can also have relevant pedagogical implications concerning the implementation of repeated reading in the L2 class.

Considering these gaps, the goal of the present exploratory study is to throw more light on the practice of repeated reading in an L2 by examining both processing as well

as learning outcomes when inter-repetition spacing is massed as opposed to spaced. In terms of processing, the study will use eye-tracking to examine changes in online processing of a text as a whole as well as a set of novel words included in it. Regarding learning products, the study will examine how different spacing conditions (massed vs. spaced) affect the degree of incidental vocabulary learning through repeated reading.

2. Literature review

2.1. Repeated Reading for Fluency and Vocabulary Learning

Repeated reading, which involves reading short texts multiple times, is a technique used in schools to promote L1 reading fluency and comprehension, for children with or without reading disabilities (Therrien, 2004). Reading involves the interaction of diverse complex processes including word decoding and meaning comprehension at the word, sentence and discourse level. According to La Berge and Samuels (1974), and Samuels (2004), repetitions make decoding easier and also help create associations between words, which contributes towards reading fluency. Similarly, repeated reading allows for greater familiarity with the content of the text, which also leads to faster reading (Samuels, 2004).

In the L2 literature, the studies by Gorsuch and Taguchi (2008) and Chang and Millet (2013) show that repeated-reading interventions (five repetitions in one single session) with audio support helped Vietnamese and Taiwanese adult EFL learners significantly improve their reading fluency and comprehension. The faster reading rates attained through repeated reading show that the reading process is becoming less effortful for the readers, and, as Grabe and Stoller (2013) suggest, these faster reading rates might be a sign of automatization of reading processes (DeKeyser, 2007).

A few studies in the L1 context have used eye-tracking to examine the effect of repeated reading on online reading behavior. These studies have shown that repeated reading generates a facilitation effect that is reflected in fewer and shorter fixations, longer saccades (Hyönä & Niemi, 1990; Inhoff, Topolski, Vitu, & O'Regan, 1993; Raney & Rayner, 1995) and a reduction in the proportion of regressions (Schnitzer & Kowler, 2006), presumably because repeated reading decreases attention demands (Inhoff et al., 1993). This facilitation effect concerns both surface features related to visuographic information, as well as higher-level comprehension processes (Hyönä & Niemi, 1990). Crucially, no previous studies have examined the effect of repeated reading on L2 reading behavior.

Research also shows that repeated reading can promote incidental vocabulary learning (e.g., Han & Chen, 2010; Horst & Meara, 1999; Liu & Todd, 2016; Llanes

& Tragant, 2021; Webb & Chang, 2012). The number of repetitions examined in previous research varies from a minimum of two up to eight. The fact that repeated reading guarantees multiple exposures to novel words is probably one of the features that contributes to vocabulary learning in this type of reading practice. However, some authors have also raised concerns about the fact that repeated readings of the same text can have a detrimental effect on learners' attention and motivation (Nichols, Rupley, & Rasinski, 2008). Concerning L2 learners' attention while reading, eye-tracking research has shown that encountering the same words several times in a text leads to a decrease in reading times across repetitions (Godfroid et al., 2018; Elgort et al, 2018; Pellicer-Sánchez, 2016). Moreover, eye-tracking studies have also provided some evidence that longer processing times on target vocabulary are related to better performance in vocabulary tests (e.g., Godfroid, Boers, & Housen, 2013; Pellicer-Sánchez, 2016). However, these questions have not been investigated in the context of repeated reading. Exploring how reading times change across repeated exposures to novel words and how such processing is related to vocabulary gains through repeated readings of the same text can provide insights into the potential of repeated reading to promote incidental vocabulary learning. Additionally, empirical studies should be conducted to explore whether different implementations of repeated reading could reduce the effect of boredom or lack of attention. One alternative could be spacing the repetitions of same-text re-readings over several sessions instead of just one (Horst & Meara, 1999; Llanes & Tragant, 2021).

2.2. Spacing and Repeated Reading

The discussion of spacing effects is particularly relevant for the application of repeated reading, as for any type of L2 practice that involves repetition. Studies in cognitive psychology have provided evidence for the spacing effect, which suggests that including spacing between repetitions of target items is more effective for learning than massed sequences, in which repetitions appear immediately (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). In addition, when considering spaced learning, it has been suggested that longer lags are more beneficial than shorter lags, although this lag effect is not as uniform or robust as the spacing effect (Toppino & Gerbier, 2014). In fact, if the lags are too widely spaced, they might prevent retrieval of previous presentations from memory, which might eliminate the benefits of repeated practice (Thios & D'Agostino, 1976; Toppino & Bloom, 2002).

Most of the evidence for the spacing effect in cognitive psychology comes from verbal learning through paired-associates. Studies in SLA using a similar paradigm have also confirmed the spacing effect for vocabulary learning in one single session (Koval, 2022; Nakata, 2015; Nakata & Suzuki, 2019). There are different theories that try to account for the spacing effect, including the deficient processing, encoding

variability or study-phase retrieval theories (see Edmonds, Gerbier, Palasis, & Whyte, 2021; Serrano, 2012; Toppino & Gerbier, 2014). The desirable difficulties theory proposed by Bjork (Bjork, 1994; 2018) has received recent attention in the SLA literature (Rogers & Leow, 2020; Suzuki et al., 2019). This theory suggests that spacing makes processing more effortful, but this increased processing effort translates into better learning outcomes.

As several authors have emphasized (Suzuki et al., 2019; Koval, 2022), most SLA studies on the spacing and lag effects have focused on learning products rather than online processing during the learning phase. The few existing studies, however, confirm that processing is more effortful when repetitions are spaced. In the study by Koval (2019), a group of English L1 speakers was asked to infer the meaning of 24 Finnish words which appeared in consecutive sentences (massed) or in sentences that were spaced. The results of the eye-tracking measures showed longer reading times for spaced items, which were better learned than massed. Koval (2022) compared learning and processing of Finnish-English paired-associates repeated in massed, short-spaced and long-spaced sequences, and found that participants were significantly faster in retrieving the L1 form after seeing the L2 cue in massed rather than in spaced presentations. However, the results of the vocabulary posttests showed a significant advantage for spaced repetitions, which were processed more effortfully during training. While Koval provided evidence for the processing and learning of novel vocabulary through repeated exposures under massed vs. spaced conditions in one session, little is known about how learning and processing of vocabulary changes when repetitions occur in one session as opposed to several sessions, or when they occur in the context of repeated reading.

Although there are no studies focusing on how spacing in repeated reading affects reading processing or reading speed, there is some research in this direction in the area of oral fluency which provides some interesting insights on how fluency might be affected by inter-repetition spacing of L2 practice. In the case of Japanese EFL learners, Suzuki (2021) found that three blocked (massed) repetitions of an oral task (i.e., AAA, BBB, CCC) led to more fluent speech processing than interleaved practice, in which learners repeated the same tasks over three different days (ABC, ABC, ABC). This increased fluency achieved in the blocked condition also transferred to other tasks. Similarly, Bui, Ahmadian, and Hunter (2019) found that immediate repetitions of the same task led to significantly more fluent oral production than when repetitions were distributed over several days. These findings also suggest that processing under more concentrated schedules becomes less effortful (i.e., more fluent) than when repetitions are widely spaced.

To the authors' knowledge, there is only one study that has examined the role of spacing on incidental learning of vocabulary through repeated reading of the same

text. In the context of a Taiwanese high school, Serrano and Huang (2018) investigated how different schedules of repeated reading practice facilitated the learning of 36 English words. The study used a meaning recognition vocabulary test and found that the short-spaced schedule (repeated reading everyday over five consecutive days) encouraged significantly more learning at the end of the treatment than the long-spaced condition (repeated reading sessions over five weeks with a 7-day intersession interval, ISI). When analyzing long-term gains, however, the study found that there was a statistically significant loss for the short-spaced group, but not for the long-spaced group. This result supports the prediction of the desirable difficulty framework in the sense that more spacing might have created a desirable difficulty during the learning sessions, which was beneficial for long-term retention. However, this study does not offer any insights on how spacing affected reading fluency and reading processes.

The current paper draws on previous research on repeated reading and also on spacing/lag effects in order to learn about how repeated reading of the same text affects L2 reading behavior, L2 incidental vocabulary learning, as well as the relationship between the two, in massed versus spaced repeated reading. Crucially, despite the claims made about different spacing schedules leading to different amounts of online processing during the learning phase and the potential that this has to explain spacing effects (Koval, 2019, 2022), no previous studies have examined how spacing affects processing in the context of L2 repeated reading, which could have valuable theoretical and practical implications.

3. Research Questions

This study aims to address the gaps identified in the review of the literature and examine how L2 learners process repetitions of the same text and target vocabulary under a massed (subsequent repetitions) versus spaced (7-day ISI) distribution. Another aim of the study is to analyze vocabulary learning under the two spacing conditions, as well as how online processing of the target words relates to vocabulary learning. The following questions were addressed:

1. Are there any differences in the online processing of the text and target vocabulary between massed and spaced repeated reading, as examined by readers' eye movements?

2. Are there any differences in incidental vocabulary gains between massed and spaced repeated reading conditions?

3. Are vocabulary gains in massed and spaced repeated reading related to online processing of the target words?

Based on the existing literature on repeated reading as well as eye-tracking research, we hypothesize that repeated reading will facilitate the development of reading fluency, evidenced by a decrease in number and duration of fixations (Hyona & Niemi, 1990; Raney & Rayner, 1995), with a more rapid decrease under the massed condition (Koval, 2019), in line with findings on L2 fluency (Bui et al., 2019; Suzuki, 2021). This more significant decrease under the massed condition is expected to be reflected in lower vocabulary gains (Koval, 2019). These hypotheses are in line with the predictions of the desirable difficulties framework, according to which, more difficulties (i.e., more processing effort) during training would contribute to more solid learning gains (Suzuki et al., 2019). Finally, processing time on novel words is expected to be positively related to vocabulary gains, regardless of the spacing schedule (Godfroid et al., 2018; Pellicer-Sanchez, 2016).

4. Methods

4.1. Participants

The participants were 44 Spanish/Catalan bilingual undergraduate students from the same university taking English Studies (females = 34, mean age = 21.3). The participants were recruited from three different groups on the same course. On average, they had a vocabulary size of 6,937 words (SD = 1,050), according to the V_YesNo Test (Meara & Miralpeix, 2017), and advanced English proficiency (C1, Common European Framework of Reference, CEFR), as assessed by the Oxford Quick Placement Test (UCLES, 2001) (M = 48/60, SD = 5.6). All the participants obtained course credit for their voluntary participation. Half of the participants (n = 22) were randomly assigned to the massed condition and the other half (n = 22) to the spaced (7-day ISI). After discarding outliers, students missing one or more tests, and students whose eye movements were not accurately recorded, the final number of students whose data was included in the study were the following: massed = 18, spaced = 19 in the vocabulary analyses; and massed = 19, spaced = 19 in the analysis of eye movements.

There were no significant differences between the two participant groups in the Oxford Quick Placement Test (massed: M = 46.1, SD = 5.7; spaced: M = 48.5, SD = 5.7; t(35) = -1.09, p = .285, d = -0.37), or in the V_YesNo Test (massed: M = 6845.9, SD = 1010.8; spaced: M = 6973.0, SD = 1148.2; t(35) = -.357, p = .724, d = -0.17).

4.2. Instruments

4.2.1. Reading Passage

The text was a short narrative from the book *New Proficiency Reading* (Stephens 2006), which provides practice for the Cambridge Proficiency Exam (equivalent to C2 in the CEFR). The original passage was modified to ensure that the words in the text would not pose any difficulty to the participants. Considering their advanced proficiency and that their overall vocabulary size was 6,937 words on average, it seemed safe to assume that they would be familiar with the first 4,000 most frequent words in English. The modified text was 751 words long and knowledge of the first 4,000 most frequent words provided a lexical coverage of 98%. The participants were asked to read the text naturally in order to understand the main ideas. A set of 22 *True/False* comprehension questions, including general questions about main idea units, was devised in order to make sure the students were reading the text for meaning. The comprehension questions were in English and did not require processing of the target vocabulary. In general, the students answered these questions without difficulty and there were no significant differences between the two conditions (massed: M =17.3, SD = 1.9; spaced: M = 17.3, SD = 1.8; $t(35) = -.065$, $p = .949$, $d = 0.02$).

4.2.2. Vocabulary Tests

Twelve target words (six nouns and six adjectives), which were likely to be unknown by participants were selected from the reading passage. The initial selection was based on the researchers' evaluation of the difficulty of the words and confirmed with a small group of learners of similar proficiency to the participants in the study. The knowledge of these words was pretested using a *Yes/No* checklist vocabulary test containing these words and 48 distractors: 12 high frequency nouns, 12 high frequency adjectives, 6 low frequency nouns, 6 low frequency adjectives and 6 pseudo-words (see Appendix A for the list of words and Appendix B for the *Yes/No* checklist vocabulary test). The participants were asked to select the words for which they knew both the form and the meaning. While this format has been subject to criticism due to the possibility of the participants overestimating their knowledge and for lacking verification of knowledge (Pellicer-Sánchez & Schmitt, 2012), it allows for the measurement of a relatively large number of items in a short time. This format was adopted for practical reasons. Due to time constraints, the pretest had to take place on the same day as the first reading, and we wanted to prevent excessive focus on the target words through testing. The large number of distractors allowed us to avoid making the target items particularly salient before the treatment started. It is possible that, without having to demonstrate their knowledge of the selected items, participants overestimated their knowledge of some words. However, while conservative, this allowed us to control even for initial, partial

knowledge of the target items. The reliability of this pretest was acceptable: Cronbach alpha = .757.

In order to capture vocabulary learning after the repeated reading sessions (posttest), a multiple-choice test was created including the 12 target words exclusively. The multiple-choice format was preferred since it involves a demonstration of knowledge. A meaning-recognition test was adopted because it typically captures more gains than recall tests (Waring & Takaki, 2003) and, considering that the words only appeared three times (once in each repetition of the text), it would have been challenging for students to demonstrate productive knowledge. In the multiple-choice test, the students had to choose the best definition for each target word, out of four possible definitions. The non-target definitions (from the Macmillan online dictionary) referred to words from the same linguistic category as the target words and were related to the content of the story. There was an "I don't know" option to minimize guessing (see Appendix C for the multiple-choice test). All the words in the definitions in the multiple-choice test were either among the first 3,000 most frequent words or were cognates, which ensured that the language could be understood by the target participants. Cronbach's alpha = .703 shows that the reliability of this test was acceptable.

4.3. Procedure

The data collection for this study was part of a larger project that also included other tasks related to students' L2 writing skills. For the present study, all the participants completed a pretest, treatment (reading the same text three times), an immediate posttest (after the last reading), and two delayed posttests (one and five weeks after the immediate posttest). Two weeks before the first reading session, the students completed the V_YesNo vocabulary size test. Having information about the participants' vocabulary size prior to the study was crucial to ensure the appropriate modification of the reading material. The test was performed during class time in a computer lab and typically took around 10 minutes to complete.

The rest of the sessions took place individually in a computer lab. In the second session, the students in the massed group firstly did the target vocabulary pretest (5 minutes), followed by the Oxford Quick Placement Test (20 minutes). The students then read the passage three times on a computer while their eye movements were recorded with a Tobii T120 eye tracker (Tobii, www.tobii.com). This is a remote, desktop eye-tracker, with the camera and infrared light integrated within the monitor. It has a sampling rate of 120 Hz, a typical accuracy of 0.5° and 0.2° resolution. There was a distance of approximately 64 cm between the participants and the screen, as suggested by the eye tracker manual. The text was displayed over 10 pages in bold

Courier New 14 font. The sentences were double-spaced and there were 8-10 lines on each page. On average, there were 74 words per page (range: 60-90). Participants had to press the space bar on the computer's keyboard in order to progress from one page to the next and they could not go back to previous pages. A 5-point calibration was performed at the beginning of the experiment and before each repetition of the text. The participants read at their own pace, with each reading typically lasting 3-5 minutes. After the final reading, the students performed a pen-and-paper comprehension test (5 minutes) and a writing task related to the larger project (20 minutes). It was important to introduce a task between the final reading and the vocabulary test so that previous exposure to the target words was not so recent. Therefore, for both the massed and spaced groups, there was a lag of approximately 25 minutes between the end of the last reading and the immediate vocabulary test. Participants were asked to come back to the lab 7 and 35 days after the immediate posttest, which is when the two delayed posttests were administered, in line with other studies on time distribution of L2 practice (Suzuki & DeKeyser, 2017). Participants were not told that they would be tested again on the same vocabulary and were expecting other tasks related to the larger project. The first retention interval (RI) was expected to be advantageous for the massed condition while the second was hypothesized to be more beneficial for the spaced (Rohrer & Pashler, 2007).

The procedure for the students in the spaced group was the same, with the exception that, instead of reading the passage three times in the second session, they only read it once and came back for two more sessions (with a 7-day ISI between them). Additionally, the learners in the spaced group were asked one general comprehension question orally after readings 1 and 2 to give some purpose to the reading activity, as they did not receive the comprehension test until they finished reading 3. Altogether, the data for the learners in the massed group was collected over four sessions and for the spaced group over six sessions (see Figure 1 for a summary of the procedure).

Figure 1: Procedure followed for the massed and spaced repeated reading groups

MASSED	ISI Days	**SPACED**	ISI Days
SESSION 1 V_YesNo Test		**SESSION 1** V_YesNo Test	
	15		15

SESSION 2 Pretest Oxford Quick Placement test Reading 1 Reading 2 Reading 3 Comprehension questions Writing task Immediate posttest	
	7
SESSION 3 7-day RI posttest	
	28
SESSION 4 35-day RI posttest	

SESSION 2 Pretest Oxford Quick Placement test Reading 1 1 general comprehension question	
	7
SESSION 3 Reading 2 1 general comprehension question	
	7
SESSION 4 Reading 3 Comprehension questions Writing task Immediate posttest	
	7
SESSION 5 7-day RI posttest	
	28
SESSION 6 35-day RI posttest	

5. Analyses

Eye-tracking measures of two different areas of interest were examined: the whole text and each of the target words. Eye-tracking studies on repeated reading in the L1 context have examined global eye-tracking measures based on entire sentences (e.g., Hyönä & Niemi, 1990) and entire passages (Raney & Rayner, 1995). In line with Raney and Rayner (1995), we examined measures at both the passage and the target word levels. By looking at both eye movements to the text and to specific target words, we were able to examine the effect that spacing had on reading the specific target items, as well as on overall reading behavior. Three eye-movement measures were examined: total reading time, fixation count, and average fixation duration. Total reading time and fixation count are late eye-movement measures that reflect lexical integration and are affected not only by lexical factors, but also by contextual, syntactic, and discourse-level properties of what is being read (Conklin et al., 2018). Average fixation duration was examined as it has been claimed to be useful when exploring how eye movements

unfold over time (Conklin et al., 2018). At the text level, it was expected that repeated reading would increase text familiarity, which would be reflected in shorter total reading time, shorter fixation durations, and fewer fixations on the text (Hyönä & Niemi, 1990; Raney & Rayner, 1995). Average fixation duration is sensitive to text difficulty and, therefore, the reduced difficulty of a repeated text might be reflected in shorter average fixation durations (Hyönä & Niemi, 1990). Similarly, at the target word level, it was expected that repeated encounters with the text would lead to fewer and shorter fixations on the target words (Godfroid et al., 2013, 2018; Pellicer-Sánchez, 2016).

In order to analyze vocabulary gains under the two conditions, we computed relative gains by item, applying a formula that has often been used in vocabulary research (e.g., Peters & Webb, 2018): relative gains = (number of learned words/ (number of target items – number of known words)) x 100. "Learned words" are those that were incorrect in the pretest and correct in the posttest; and "known words" are those that were correct both in pretest and posttest. The above formula can correct possible overestimations of students' knowledge in the *Yes/No* pretest, since it considers as "known" only the words that were correct at both testing times. We computed immediate relative gains (with pretest and posttest scores), RI-7 gains scores (with scores in the pretest and in the 7-day RI delayed posttest) and RI-35 gains (with scores in the pretest and in the 35-day RI delayed posttest).

The SPSS 27 program (IBM, 2020) was used to perform the statistical analyses. In order to examine eye movements (RQ1), a series of linear mixed models (LMM) with repeated measures were performed: condition (massed vs. spaced), reading time (1, 2, 3) and their interaction were included as fixed effects. Estimated marginal means were obtained for each factor, and p-values of pairwise comparisons were corrected using Bonferroni's method. The residuals were saved and analyzed for normality of distribution. The Kolmogorov-Smirnov test confirmed the normality of the residuals in all cases ($p > .05$) except for total reading time and fixation count for the target words. In that case, a logarithmic transformation was performed, after which the distribution of the residuals was normal ($p = .200$). In order to examine vocabulary gains (RQ2), we firstly performed a preliminary analysis to compare the two conditions at the pretest level. The t-test revealed that the massed group knew significantly fewer words ($M = 2.22$, $SD = 1.73$) than the spaced ($M = 3.48$, $SD = 1.54$); $t(35) = -2.32$, $p = .026$, $d = -0.76$. Because of this difference, the pretest scores were always added in the statistical models that were used to examine vocabulary gains. Another LMM with repeated measures was performed to examine vocabulary gains, with condition (massed vs. spaced), time (immediate, RI-7 and RI-35) and pretest scores as fixed effects. The interaction between condition and time was also explored. Finally, in order to address the third question, partial correlations were performed between the eye-tracking measures on the target

words and the vocabulary gains. Effect sizes were manually calculated using Cohen's d. The SDs for the estimated means were calculated by using the standard error (SE) applying the following formula: SD = sqrt(n) * SE. Effect sizes were interpreted using the benchmarks proposed by Plonsky and Oswald (2014).

6. Results

6.1. *Reading Patterns in Massed vs. Spaced Repeated Reading*

Reading patterns for the text were analyzed first. The descriptive statistics of these text measures at each reading time appear in Table 1.

Table 1: Descriptive statistics for eye-movement measures to the text across the three repeated readings (T1, T2, T3) by condition. Mean values and standard deviations (in parentheses) calculated per page of text. Fixation durations in milliseconds

	T1		T2		T3	
	Massed (n = 19)	Spaced (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)
Total Reading Time	200,588 (53,770)	169,228 (61,224)	168,453 (40,323)	161,727 (55,237)	154,106 (40,089)	160,642 (48,003)
Fixation Count	97.74 (23.67)	84.16 (25.40)	84.35 (16.94)	83.82 (26.69)	77.59 (17.39)	80.79 (18.80)
Average Fix. Duration	201.5 (28.8)	193.8 (26.4)	197.7 (26.0)	186.8 (19.8)	194.3 (27.9)	191.5 (26.0)

The results of the LMMs are presented in Table 2. In the case of **total reading time**, the effect of condition was not significant (p = .451, d = 0.17), while there was a main effect of time: significant decrease time 1-2 (p = .019, d = 0.28), and time 1-3 (p = .001, d = 0.39). There was a significant condition*time interaction (p = .038). When examining each condition separately, it was found that the massed group significantly decreased their total reading time between time 1-2 (p = .007, d = 0.45) and 1-3 (p = .001, d = 0.65), although the effect sizes were small. No significant changes were registered between time 2-3. The total reading time on the text in the case of the spaced group did not significantly change across time.

The results for **fixation count** showed a similar pattern: no effect of condition (p = .552, d = 0.14), and significant effect of time (p = .002) and condition*time interaction (p = .027). Under the massed condition, participants decreased the number of fixations significantly between time 1 and 2 (p = .009, d = 0.44), and time 1 and

3 (p <.001, d = 0.65), and, in both cases, the effect size was small. The spaced group showed no significant changes. The results for **average fixation duration** showed no significant effect of condition (p = .343), time (p = .185), or the interaction between the two (p = .386).

Table 2: Results of the LMMs with eye-tracking measures to the text

Target	Source	F	df1	df2	p	Condition	Time	
TRT	Condition	.572	1	107	.451	d = 0.17		
	Time	7.35	2	107	.001		1-2: 2.63, p = .019, d = 0.28 2-3: 3.72, p = .296, d = 0.11 1-3: 3.72, p = .001, d = 0.39	
	Condition *Time	3.37	2	107	.038	1: 1.91, p = .058, d =0.44 2: .421, p = .674, d =0.10 3: -.399, p = .691, d=-0.09	Massed	Spaced
							1-2: 2.99, p = .007, d =0.45 2-3: 1.37, p = .174, d =0.20 1-3: 4.44, p < .001. d =0.65	1-2: .717, p = 1.00, d =0.10 2-3: .104, p = 1.00, d =0.02 1-3: .821, p = 1.00, d =0.12
FC	Condition	.356	1	107	.552	d = 0.14		
	Time	6.74	2	107	.002		1-2: 2.12, p = .072, d = 0.22 2-3: 1.50, p = .137, d = 0.16 1-3: 3.65, p = .001, d = 0.37	
	Condition *Time	3.74	2	107	.027	1: 1.91, p = .059, d = 0.44 2: .066, p = .947, d = 0.02 3: .450, p = .653, d = -0.10	Massed	Spaced
							1-2: 2.90, p = .009, d = 0.44 2-3: 144, p = .151, d = 0.22 1-3: 4.42, p < .001, d = 0.65	1-2: .076, p = 1.00, d = 0.01 2-3: .665, p = 1.00, d = 0.10 1-3: .741, p = 1.00, d = 0.11
AFD	Condition	.908	1	107	.343	d = 0.24		
	Time	1.71	2	107	.185			
	Condition *Time	.961	2	107	.386			

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

Eye movements to the target words were then analyzed. The descriptive statistics appear in Table 3. The results were similar to those reported for the whole text: a decrease in reading times across time, with a more marked decrease in the massed group (see Figure 2 for a visual representation of results).

Table 3: Descriptive statistics for eye-tracking measures to the target words across the three repeated readings (T1, T2, T3) by condition. Mean values and standard deviations (in parentheses) calculated per target item. Fixation durations in milliseconds

	T1		T2		T3	
	Massed (n = 19)	Spaced (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)
TRT	789.0 (433.3)	646.1 (269.30)	522.8 (202.3)	570.9 (342.77)	483.3 (149.58)	477.3 (163.83)
TRT Log transform	-.155 (.223)	-.235 (.222)	-.318 (.195)	-.305 (235)	-.335 (.136)	-.346 (.156)
TRT Log transform	3.35 (1.14)	2.87 (1.06)	2.31 (0.63)	2.72 (1.53)	2.17 (0.55)	2.30 (0.58)
FC Log transform	.501 (.151)	.429 (.169)	.345 (.138)	.386 (.205)	.325 (.112)	.346 (.130)
AFD	207.6 (59.12)	208.8 (64.9)	201.5 (59.0)	188.8 (55.1)	199.6 (50.9)	185.5 (42.9)

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

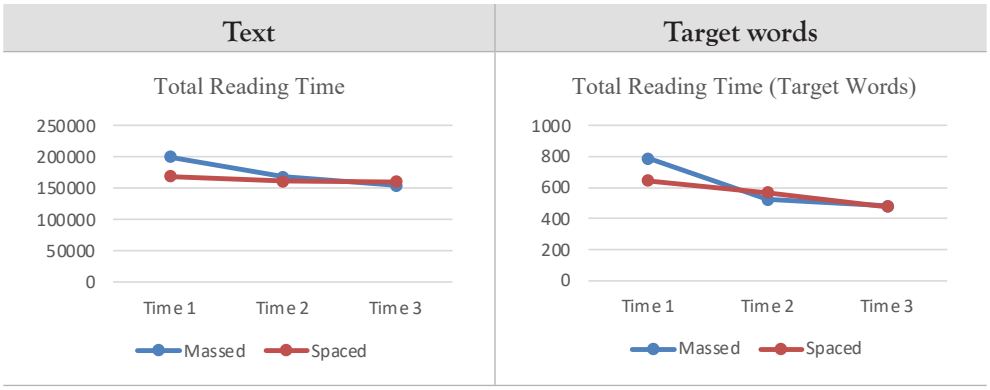
The results of the LMMs (see full report in Table 4) show that, for **total reading time**, there was a main effect of time ($p < .001$), with fixation durations significantly decreasing between time 1-2 ($p = .002$, $d = 0.43$), and time 1-3 ($p < .001$, $d = 0.52$). The effects of condition ($p = .608$, $d = 0.12$) and condition*time interaction ($p = .357$) were not significant. Similar results were found for **fixation count** to the target words, with no main effect of condition ($p = .910$, $d = 0.03$), no significant condition*time interaction ($p = .094$), and a significant main effect of time ($p < .001$), with the number of fixations decreasing significantly between time 1-2 ($p = .001$, $d = 0.45$) and time 1-3 ($p < .001$, $d = 0.60$). Finally, with respect to **average fixation duration** to the target words, the results were similar to those of the same measure for the whole text: no significant effect of condition ($p = .584$), time ($p = .174$), or condition*time interaction ($p = .637$).

Table 4: Results of LMM with eye-tracking measures to the target words

Target	Source	F	df1	df2	p	Condition	Time
TRT	Condition	.265	1	107	.608	d =0.12	
	Time	10.56	2	107	<.001		1-2: 3.44, p = .002, d =0.43 2-3: .868, p = .387, d =0.10 1-3: 4.35, p <.001, d =0.52
	Condition* Time	1.03	2	107	.357		
FC	Condition	.013	1	107	.910	d = 0.03	
	Time	12.67	2	107	<.001		1-2: 3.60, p = .001, d =0.45 2-3: 1.19, p = .236, d =0.15 1-3: 4.84, p < .001, d =0.60
	Condition* Time	2.42	2	107	.094		
AFD	Condition	.301	1	107	.584	d = 0.03	
	Time	1.78	2	107	.174		
	Conditio* Time	.453	1	107	.637		

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

Figure 2: Results of the eye-movement measures to the text and target words



6.2. Relative Vocabulary Gains in Massed versus Spaced Repeated Reading

Table 5 shows the mean percentage of immediate and delayed relative vocabulary gains for each condition after controlling for pretest scores. The descriptive statistics show that the spaced group made more gains than the massed group, especially in the long term.

Table 5: Descriptive statistics vocabulary gains (in percentages) by condition and testing times, including estimated means and standard error in parentheses

Testing time	Experimental group	
	Massed (n = 18)	Spaced (n = 19)
Immediate gains (Pretest-posttest)	19.83 (4.17)	23.28 (4.05)
Gains RI-7 (Pre-delayed 1)	20.66 (4.24)	40.16 (4.09)
Gains RI-35 (Pre-delayed 2)	22.73 (4.16)	40.95 (3.98)

Continuous predictors are fixed at the following values: Pretest=2.934

The results of the LMM (see Table 6 for full report) showed that there were significant main effects of condition (p = .008, d = 0.65) and time (p <.001) with small effect sizes. The condition*time interaction was also significant (p = .006). When

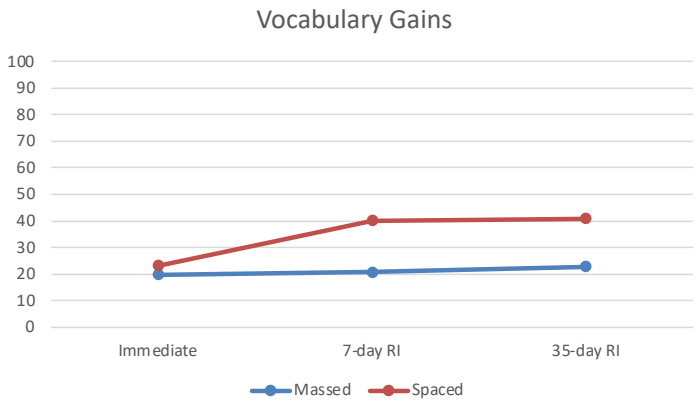
comparing conditions for immediate gains, no differences were observed between the two groups (p = .516, d = -0.16), while for long-term gains (both RI-7 and RI-35 gains), the spaced group significantly outperformed the massed group (p = .002, d = -0.78; p = .003, d = -0.73 respectively) and the effect size of this difference was medium. When examining the two conditions separately, the performance of the students in the massed group did not change significantly after the immediate posttest. The performance of the spaced group, on the other hand, significantly improved after the immediate posttest: immediate versus 7-day RI gains (p < .001, d = -0.69); immediate versus 35-day RI gains (p <.001, d = 0.73). Figure 3 illustrates how the performance differed between the two groups.

Table 6: Results of the LMMs with vocabulary gains

Source	F	df1	df2	p	Condition	Time	
Condition	7.25	1	115	.008	d = 0.17		
Time	9.45	2	115	< .001		1-2: -3.41, p = .002, d = -0.37 2-3: -5.39, p = .591, d = -0.06 1-3: -4.02, p < .001, d = -0.44	
Condition* Time	5.41	2	115	.006	1: -652, p = .516, d = -0.16 2: -3.22, p = .002, d = -0.78 3: -3.07, p = .003, d = -0.73	Massed	Spaced
						1-2: -.335 p = 1.00, d = -0.05 2-3: -.542 p = 1.00, d = -0.08 1-3: -.892 p = 1.00, d = -0.13	1-2: -4.53, p <.001, d = -0.69 2-3: -.215, p = .830, d = -0.03 1-3: -4.86, p <.001, d = -0.73

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

Figure 3: Vocabulary gains (%) in massed and spaced repeated reading



These results suggest that the long-term gains experienced by the spaced group were probably not a direct product of the treatment, since more significant improvement took place after the treatment than during it.

6.3. *Relationship between Vocabulary Gains and Online Processing*

Next, we examined whether online processing of the target words was differentially related to vocabulary gains in the two conditions through partial correlations, controlling for pretest vocabulary scores. The results of the correlations indicate that cumulative reading times (each including time 1 + time 2 + time 3) were not related to vocabulary gains from repeated reading under the massed condition. Interestingly, the correlations were significant under the spaced condition, but only in the case of immediate vocabulary gains: total reading time (r = .582, p = .01), fixation count (r = .466, p = .05) and average fixation duration (r = .600, p = .009) (see Table 7).

Table 7: Results of the correlations between eye-tracking measures (T1 + T2 + T3) on the target words and vocabulary gains

		Adjusted Immediate			Adjusted RI-7			Adjusted RI-35		
		All (n = 37)	Massed (n = 18)	Spaced (n = 19)	All (n = 37)	Massed (n = 18)	Spaced (n = 19)	All (n = 37)	Massed (n = 18)	Spaced (n = 19)
TRT	**Corr.**	.257	-.139	.582	.097	.060	.145	.154	.025	.267
	Sig.	.131	.594	.01	.573	.819	.566	.370	.923	.283
FC	**Corr.**	.268	-.283	.466	.320	-.099	.100	.306	-.120	.184
	Sig.	.114	.271	.05	.057	.706	.692	.070	.646	.465
AFD	**Corr.**	.176	-.179	.600	.049	.139	.388	.90	.175	.393
	Sig.	.304	.492	.009	.776	.594	.111	.601	.502	.106

TRT = Total fixation duration; FC = Fixation count; AFD = Average fixation duration

7. Discussion

In this exploratory study, we analyzed the development of L2 reading fluency through two different implementations of repeated reading, examining online processing of text and target words through an eye-tracker. Another goal of the study was to analyze incidental vocabulary learning in massed versus spaced repeated reading as well as its relationship with online reading processing of the target words.

In response to RQ1, while the effect of the spacing condition on overall textual processing was not significant, a significant interaction was observed between time and spacing condition for two out of the three measures examined, namely total reading time and fixation count. Only the participants under the massed condition experienced a significant decrease across repetitions after reading the text the first time, which is in line with the facilitation effect found for repeated reading in the

L1 in previous eye-tracking research (Hyönä & Niemi, 1990; Raney & Rayner, 1995; Inhoff et al., 1993). This finding is also consistent with the facilitation effect of L2 repeated reading observed through reading rates (Chang & Millet, 2013; Gorsuch & Taguchi, 2008). In all these studies, repeated reading was done in one session, as was the case for the massed group in the present study. It can be assumed that faster reading rates and shorter processing times across repetitions happen because of the high activation of previous presentations of the same words and structures in the text. This increased familiarity with the materials makes the text easier to read and leads to a decrease in attentional demands. Efficient reading processing depends on the proceduralization and later automatization of lower-level processing skills, such as word recognition, syntactic parsing or propositional encoding (Grabe, 2009). High activation of the representations obtained from previous readings, as well as increased familiarization with the content of the text, made it easier for the learners under the massed condition to attain faster processing. This more efficient processing might be indicative of proceduralization, which, according to some studies, benefits from short-spaced/massed repetitions (Li & DeKeyser, 2019; Suzuki, 2020). It must be emphasized, however, that the effect sizes for the differences in processing were small, which may be due to the low number of repetitions, in comparison with other repeated reading studies (e.g., Chang & Millet, 2013; Gorsuch & Taguchi, 2008).

On the other hand, the fact that the students in the spaced group did not show differences in processing patterns across repetitions may suggest that they were processing the text each time as if it were a new text. A 7-day lag might have been a long interval for learners to be able to remember previous presentations of the text. In fact, study-phase retrieval theories of the spacing effect suggest that spacing is beneficial for learning as long as it allows for retrieval of previous presentations (Toppino & Bloom, 2002). Under the desirable difficulties framework, it can be said that, in this case, spacing added a difficulty that was not desirable for the development of automatic reading processes. This finding is also in line with previous findings for speech production, which show that more concentrated repetitions of the same task led to increased oral fluency in comparison to spaced repetitions (Bui et al., 2019; Suzuki, 2021).

The lack of interaction between reading time and condition for the eye-tracking measures on the target words suggests that there were no significant differences between the conditions with respect to the processing of the target vocabulary, despite the fact that the decrease in reading times is clearer for the massed than for the spaced condition. In this case, even though the learners under the massed condition had more recent exposure to previous presentations of the target words and their orthographic representations were more active, processing these words was probably still challenging because their meaning was unknown and probably difficult to guess.

This result is not in line with Koval's (2019), who found that target words in massed sequences were given less processing time than those in spaced sequences. This conflicting finding may be explained by methodological differences between the studies. The time between repetitions of the target items under the massed condition was shorter in Koval's research than in the current study, as the target words were introduced in short sentences in the participants' L1 that appeared subsequently, as opposed to a long text in the L2. Additionally, Koval's study examined intentional learning conditions, which might have affected how learners approached the reading task.

In relation to RQ2, the findings reported in this paper suggest that massed and spaced repeated reading did not lead to significantly different immediate vocabulary gains (19% and 23%, respectively). These gains are statistically significant, in line with other studies that have examined vocabulary learning through repeated reading in one session (e.g., Liu & Todd, 2016) or in more than one session (e.g., Llanes & Tragant, 2021). The results concerning long-term gains cannot be discussed in relation to the treatment, as intentional vocabulary learning possibly happened for some students in the spaced group, who improved their vocabulary scores after the treatment more significantly than during the treatment.

The fact that no significant differences were observed for incidental vocabulary learning between the two conditions at the immediate testing stage (short RI) is in line with other studies of distributed-practice effects for intentional learning both in cognitive psychology (Pavlik & Anderson, 2005) and SLA (Bird, 2010) which, despite showing a significant advantage for spaced conditions in the long term, failed to show an equivalent advantage at shorter RIs. According to Pavlik and Anderson (2005), at short RIs, the increased activation of massed items due to recency can make it possible for these items to be recalled as well as or even better than spaced items.

Our results are different from the studies by Koval (2019), and Nakata and Elgort (2021), which showed that spaced exposure significantly promoted more vocabulary gains than massed. This divergence in results can be explained by the fact that these two studies examined within-session spacing, whereas the current one investigated between-session spacing. Another reason that can explain the lack of differences is that incidental learning of vocabulary through just three repeated exposures to the target words was too demanding for all the students, regardless of the spacing of repetitions. While the immediate vocabulary gains registered as a product of repeated reading were in line with other previous studies on incidental vocabulary learning from reading (see Pellicer-Sánchez, 2016), they were quite low (about 20% of the words).

Finally, with respect to RQ3, about the relationship between processing and vocabulary learning, our results suggest that cumulative reading times and number of

fixations to the target words were related to immediate incidental vocabulary gains but only for the spaced condition. The fact that the correlations became non-significant for long-term gains in this condition might again point towards the fact that the scores in the delayed posttest were not a direct product of the treatment.

Previous eye-tracking studies have shown a positive relationship between reading times and vocabulary gains (Godfroid et al., 2013; Koval, 2019; Pellicer-Sánchez, 2016), suggesting that longer reading times might be evidence of more attention to language, or "noticing" of target features. Our results show that the students under the spaced condition who paid more attention to the target items were able to make more immediate vocabulary gains. This finding suggests that spacing might have imposed a desirable difficulty for those students who noticed the target words and devoted more time to processing them than for others who might have been more focused on understanding the meaning of the whole text. According to these results, it was especially important for readers under that condition, which was the most challenging, to spend time processing the target words in order for them to be able to identify their meaning later on.

8. Limitations

As explained in the methodology, for practical reasons, different vocabulary tests were performed for the pretest and posttest. However, the less challenging test was used as the pretest and, thus, a conservative approach was followed in the calculation of gains. Any effect that this may have would be expected to be similar under the two conditions. Another limitation is that the results of long-term learning could not be analyzed as a direct product of the treatment due to the significant gains experienced by the spaced group after the treatment, which were likely due to intentional learning and not repeated reading. Although this might be an interesting, positive consequence of spaced repeated reading, which promoted more engagement with vocabulary learning after the end of the treatment for that group (with interesting pedagogical implications), the analysis of this phenomenon is beyond the scope of this study. Future research could address this issue by performing between-participant testing, in which participants only complete one of the two delayed posttests. Alternatively, using pseudowords would also prevent intentional learning outside the treatment. Despite these limitations, the present study constitutes a valuable contribution to the literature, as it is one of the first to examine processing differences in massed versus spaced reading and, to the best of the authors' knowledge, the first to analyze processing differences when learners are only asked to read for meaning comprehension and not vocabulary learning. More studies should be conducted on this topic with more participants in different contexts in order to confirm the findings from this initial exploratory study.

9. Conclusion and Pedagogical Implications

The present study showed that spacing in repeated reading led to interesting processing differences, with participants under the massed condition experiencing a more significant decrease in number and duration of fixations to the text than those under the spaced condition. Target words were processed in a similar way regardless of the condition, indicating that the processing fluency created by being repeatedly exposed to novel items in reading seems to hold when the repetitions are differently spaced. This also indicates that attentional demands on novel words were the same regardless of the spacing schedule.

One possible pedagogical implication is that, when the goal of repeated reading is to develop faster reading speed, it is probably better to do massed repeated reading, as is normally done in repeated reading techniques implemented in schools to promote L1 literacy but also L2 reading fluency.

Furthermore, our results suggest that, unlike some previous findings, the time distribution of repeated reading episodes does not have any effect on subsequent short-term incidental vocabulary learning, when comparing massed and spaced exposures. The pedagogical implication that can be derived from this finding is that it is not so important whether repeated reading happens in one session or over several sessions if one of the goals is incidental vocabulary learning, but, considering the benefit of massed repetitions for fluency, teachers might be advised to promote repeated readings of the same text in one rather than over several sessions. However, more research should be performed before this recommendation can be made. First, it is only immediate gains that were analyzed in this study, and L2 teaching should be concerned with more durable gains, which could not be successfully examined in the present paper. Second, depending on the number of repetitions, massed practice can become counter-productive (Suzuki & Hanzawa, 2022).

Our findings also suggest a relationship between processing time on target words and vocabulary gains under the spaced condition, supporting the use of teaching techniques that attract learners' attention to target vocabulary, such as input enhancement (e.g., Barcroft, 2003; Kim, 2006), especially when readers are not likely to encounter repetitions of novel words within a short time span, and repeated reading happens on different days.

Finally, our study is one of the few that have examined how spacing affects L2 processing under repeated reading conditions; therefore, the results are not intended to provide conclusive evidence on the topic. To obtain more generalizable findings, more research should be conducted in this direction in order to understand the effect of spacing on reading processes and vocabulary learning through repeated reading.

Acknowledgments

This work was supported by Grant PID2019-110536GB-I00 of the Spanish Science Ministry.

References

Barcroft, J. (2003). Distinctiveness and bidirectional effects in input enhancement for vocabulary learning. *Applied Language Learning, 13*(2), 47–73.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics, 31*(4), 635–650. https://doi.org/10.1017/S0142716410000172

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A. (2018). Being suspicious of the sense of ease and undeterred by the sense of difficulty: Looking back at Schmidt and Bjork (1992). *Perspectives on Psychological Science, 13*(2), 146–148. https://doi.org/10.1177/1745691617690642

Bui, G., Ahmadian, M. J., & Hunter, A-M. (2019). Spacing effects on repeated L2 task performance. *System, 81*, 1–13. https://doi.org/10.1016/j.system.2018.12.006

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 19*(2), 389-396. https://doi.org/10.1037/0278-7393.19.2.389

Chang, A. C-S., & Millet, S. (2013). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language, 25*(2), 126-148. https://doi.org/10125/66864

Cobb, T. (n.d.). Lextutor. Vocabprofile [computer program]. Accessed at https://www.lextutor.ca/vp/comp/

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research.* Cambridge: Cambridge University Press.

DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). New York and London: Routledge.

Edmonds, A., Gerbier, E., Palasis, K., & Whyte, S. (2021). Understanding the

distributed practice effect and its relevance for the teaching and learning of L2 vocabulary. *Lexis: Journal in English Lexicology, 18*, 1-23. https://doi.org/10.4000/lexis.5652

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Elgort, I., Brysbaert, M., & Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition, 40*(2), 341–366. https://doi.org/10.1017/S0272263117000109

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning, 64*(2), 365-414. https://doi.org/10.1111/lang.12052

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye tracking. S*tudies in Second Language Acquisition, 35*(5), 483–517. https://doi.org/10.1017/S0272263113000119

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., et al. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition, 21*(3), 563–84. https://doi.org/10.1017/S1366728917000219

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. https://doi.org/10.1016/j.system.2007.09.009

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* Cambridge: Cambridge University Press.

Han, Z., & Chen, C. 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. https://doi.org/10125/66838

Horst, M. & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *The Canadian Modern Language Review, 56*(2), 308-328. https://doi.org/10.3138/cmlr.56.2.308

Hyönä, J., & Niemi , P. (1990). Eye movements in repeated movements of a text. *Acta Psychologica, 73*(3), 259 – 280. https://doi.org/10.1016/0001-6918(90)90026-C

IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp.

Inhoff, A. W., Topolski, R., Vitu, F., & O'Regan, J. K. (1993) Attention demands during reading and the occurrence of brief (express) fixations. *Perception & Psychophysics, 54*(6), 814–823. https://doi.org/10.3758/BF03211806

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through

reading by Korean learners of English as a foreign language. *TESOL Quarterly, 10*(2), 341–373. https://doi.org/10.2307/40264526

Koval, N. G. (2019). Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking. *Applied Psycholinguistics, 40*(1), 1103-1139. https://doi.org/10.1017/S0142716421000370

Koval, N. (2022). Testing the reminding account of the lag effect in L2 vocabulary learning. *Applied Psycholinguistics, 43*(1), 1-40. https://doi.org/10.1017/S0142716421000370

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science, 42*(3), 373–388. https://doi.org/10.1007/s11251-013-9285-2

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Laufer, B. (2003) Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*(4), 567–587. https://doi.org/10.3138/cmlr.59.4.567

Li, M., & DeKeyser, R. (2019). Distribution of practice effects in the acquisition and retention of L2 Mandarin tonal word production. *The Modern Language Journal, 103*(3), 607–628. https://doi.org/10.1111/modl.12580

Liu, Y. T., & Todd, A. G. (2016). Implementation of assisted repeated reading techniques for the incidental acquisition of novel foreign vocabulary. *Language Teaching Research, 20*(1), 53-74. https://doi.org/10.1177/1362168814559802

Llanes, À., & Tragant, E. (2021, online). Measuring the effects of repeated exposure to children's graded readers. *Language Teaching Research.* https://doi.org/10.1177/13621688211036609

Macmillan (n.d.). *Macmillan online dictionary.* Retrieved from https://www.macmillandictionary.com/

Meara, P. M., (1992). *EFL Vocabulary Tests.* Swansea: Centre for Applied Language Studies.

Meara, P. M. & Miralpeix, I. (2017). *Tools for researching vocabulary*. Clevedon: Multilingual Matters.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition, 37*(4), 677–711. https://doi.org/10.1017/S0272263114000825

Nakata, T. (2020). Learning words with flash cards and word cards. In S. Webb (Ed.), *The Routledge hanbook of vocabulary studies* (pp. 304-319). New York and London: Routledge.

Nakata, T., & Elgort, I. (2021). Effects of spacing on contextual vocabulary learning: Spacing facilitates the acquisition of explicit, but not tacit, vocabulary knowledge. *Second Language Research, 37*(2), 233-260. https://doi.org/10.1177/0267658320927764

Nakata, T., & Suzuki, Y. (2019a). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition, 41*(2), 287–311. https://doi.org/10.1017/s0272263118000219

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition, 38*(3), 523-552. https://doi.org/10.1017/S0272263115000236

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston, MA: Heinle and Heinle.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I. S. P. (2013). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary.* Boston, MA: Heinle and Heinle.

Nichols, W. D., Rupley, W. H., & Rasinski, T. (2008). Fluency in Learning to read for meaning: Going beyond repeated readings. *Literacy Research and Instruction, 48*(1), 1–13. https://doi.org/10.1080/19388070802161906

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29*(4), 559-586. https://doi.org/10.1207/s15516709cog0000_14

Pellicer-Sánchez, A. (2016). Incidental vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A. (2020). Expanding English vocabulary knowledge through reading: Insights from eye-tracking studies. *RELC Journal, 51*(1), 134-146.

Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring *Yes-No* vocabulary tests: Reaction time vs. nonword approaches. *Language Testing, 29*(4), 489–509. https://doi.org/10.1177/0265532212438053

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. S*tudies in Second language Acquisition, 40*(3), 551-577. https://doi.org/10.1017/S0272263117000407

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements

during two readings of a text. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 49*(2), 151–173. https://doi.org/10.1037/1196-1961.49.2.151

Rogers J., & Leow, R. (2020). Toward greater empirical feasibility of the theoretical framework for systematic and deliberate L2 practice: Comments on Suzuki, Nakata, & DeKeyser (2019). *Modern Language Journal, 104*(1), 309-312. https://doi.org/10.1111/modl.12621

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Psychological Science, 16*(4), 183–186. https://doi.org/10.1111/j.1467-8721.2007.00500.x

Samuels, S. J. (2004). Toward a theory of automatic information processing in reading revisited. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 1127–1148). Newark, DE: International Reading Association.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55 - 88. https://doi.org/10.1177/026553220101800103

Schnitzer, B. S., & Kowler, E. (2006). Eye movements during multiple readings of the same text. *Vision Research, 46*(10), 1611-1632. https://doi.org/10.1016/j.visres.2005.09.023

Serrano, R. (2012). Is intensive learning effective? In C. Muñoz (Ed.), I*ntensive exposure experiences in second language learning* (pp. 3-22). Clevedon: Multilingual Matters.

Serrano, R., & Huang, H. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly, 52*(4), 971-994. https://doi.org/10.1002/tesq.445

Serrano, R., & Huang, H. (2023). Time distribution and intentional vocabulary learning through repeated reading: a partial replication and extension. *Language Awareness, 32*(1), 1-18. https://doi.org/10.1080/09658416.2021.1894162

Stephens, M. (2006). *New proficiency reading.* Essex: Pearson Education Limited.

Suzuki, Y. (2021). Optimizing fluency training for speaking skills transfer: Comparing the effects of blocked and interleaved task repetition. *Language Learning, 71*(2), 285-325. https://doi.org/10.1111/lang.12433

Suzuki, Y., & Hanzawa, K. (2022). Massed task repetition is a double-edged sword for fluency development. An EFL classroom study. S*tudies in Second Language Acquitision, 44*(2), 536-561. https://doi.org/10.1017/S0272263121000358

Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the proceduralization of morphology. L*anguage Teaching Research, 21*(2), 166-188. https://doi.org/10.1177/1362168815617334

Suzuki, Y., Nakata, T., & DeKeyser, R. M. (2019). The desirable difficulty framework

as a theoretical foundation for optimizing and researching second language practice. *Modern Language Journal, 103*(3), 713-720. https://doi.org/10.1111/modl.12585

Suzuki, Y., Nakata, T., & DeKeyser, R. M. (2020). Empirical feasibility of the desirable difficulty framework: Toward more systematic research on L2 practice for broader pedagogical implications. *Modern Language Journal, 104*(1), 313-319. https://doi.org/10.1111/modl.12625

Therrien, W. (2004). Fluency and comprehension gains as a result of repeated reading: A metaanalysis. *Remedial and Special Education, 25*(4), 252–261. https://doi.org/10.1177/07419325040250040801

Thios S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15*(5), 529-536. https://doi.org/10.1016/0022-5371(76)90047-5

Toppino T. C., & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(3), 437-444. https://doi.org/10.1037/0278-7393.28.3.437

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *The psychology of learning and motivation: Volume 60: The psychology of learning and motivation* (pp. 113-189). San Diego, CA: Elsevier Academic Press.

UCLES (University of Cambridge Local Examination Syndicate) (2001). *Quick Placement Test.* Oxford: Oxford University Press.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130–163. https://doi.org/10125/66776

Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review, 68*(3), 267–290. https://doi.org/10.3138/cmlr.1204.1

Appendix A: Target words and distractors

Item	Type	Frequency band	Source
smock	target noun	11K	BNC-COCA 1-25K (Lextutor)
libel	target noun	7K	BNC-COCA 1-25K (Lextutor)
smock	target noun	11K	BNC-COCA 1-25K (Lextutor)
libel	target noun	7K	BNC-COCA 1-25K (Lextutor)
yam	target noun	11K	BNC-COCA 1-25K (Lextutor)
throng	target noun	7K	BNC-COCA 1-25K (Lextutor)
hype	target noun	6K	BNC-COCA 1-25K (Lextutor)
measles	target noun	10K	BNC-COCA 1-25K (Lextutor)
wacky	target adjective	10K	BNC-COCA 1-25K (Lextutor)
appalled	target adjective	4K	BNC-COCA 1-25K (Lextutor)
batty	target adjective	13K	BNC-COCA 1-25K (Lextutor)
shrewd	target adjective	6K	BNC-COCA 1-25K (Lextutor)
floppy	target adjective	5K	BNC-COCA 1-25K (Lextutor)
creased	target adjective	6K	BNC-COCA 1-25K (Lextutor)
twose	Pseudoword		Meara (1992)
scother	Pseudoword		Elgort (2011)
buttle	Pseudoword		Meara (1992)
wray	Pseudoword		Meara (1992)
preachet	Pseudoword		Elgort (2011)
obsolate	Pseudoword		Elgort (2011)
adair	Pseudoword		Meara (1992)
regrain	Pseudoword		Elgort (2011)
galpin	Pseudoword		Meara (1992)
mundy	Pseudoword		Meara (1992)
bance	Pseudoword		Meara (1992)
stace	Pseudoword		Meara (1992)
whim	low freq noun	9K	VST (Nation & Beglar, 2007)

rouble	low freq noun	13K	VST (Nation & Beglar, 2007)
skylark	low freq noun	13K	VST (Nation & Beglar, 2007)
blaspheme	low freq noun	10K	VLT (Nation, 1990)
beagle	low freq noun	13K	VST (Nation & Beglar, 2007)
peasantry	low freq noun	10K	VST (Nation & Beglar, 2007)
gauche	low freq adjective	14K	VST (Nation & Beglar, 2007)
canonical	low freq adjective	14K	VST (Nation & Beglar, 2007)
jovial	low freq adjective	13K	VST (Nation & Beglar, 2007)
limpid	low freq adjective	14K	VST (Nation & Beglar, 2007)
bawdy	low freq adjective	14K	VST (Nation & Beglar, 2007)
upbeat	low freq adjective	9K	VST (Nation & Beglar, 2007)
basket	high freq noun	2K	VLT (Nation, 1990)
standard	high freq noun	1K	VST (Nation & Beglar, 2007)
birth	high freq noun	2K	VLT (Nation, 1990)
patience	high freq noun	2K	VST (Nation & Beglar, 2007)
dinosaur	high freq noun	3K	VST (Nation & Beglar, 2007)
soldier	high freq noun	3K	VST (Nation & Beglar, 2007)
flesh	high freq noun	2K	VLT (Nation, 1990)
drawer	high freq noun	2K	VST (Nation & Beglar, 2007)
victory	high freq noun	2K	VLT (Nation, 1990)
stone	high freq noun	2K	VST (Nation & Beglar, 2007)
time	high freq noun	1K	VST (Nation & Beglar, 2007)
debt	high freq noun	2K	VLT (Nation, 1990)
holy	high freq adjective	2K	VLT (Schmitt, Schmitt & Clapham, 2001)
private	high freq adjective	2K	VLT (Nation, 1990)
total	high freq adjective	2K	VLT (Nation, 1990)
candid	high freq adjective	4K	VST (Nation & Beglar, 2007)
annual	high freq adjective	4K	VLT (Nation, 1990)
naked	high freq adjective	4K	VLT (Schmitt, Schmitt & Clapham, 2001)
upset	high freq adjective	2K	VST (Nation & Beglar, 2007)

ancient	high freq adjective	2K	VLT (Schmitt, Schmitt & Clapham, 2001)
aware	high freq adjective	3K	VLT (Schmitt, Schmitt & Clapham, 2001)
original	high freq adjective	2K	VLT (Nation, 1990)
lovely	high freq adjective	2K	VLT (Schmitt, Schmitt & Clapham, 2001)
poor	high freq adjective	1K	VST (Nation & Beglar, 2007)

Appendix B: Vocabulary pretest

Please mark "YES" if **you know the meaning** of the following words, and mark "NO" if you don't know the meaning (no guessing). Keep in mind that there are non-words (for which you'll have to mark "NO", obviously), so please be honest and mark "NO" when applicable.

	YES	NO
twose		
holy		
smock		
wacky		
private		
total		
basket		
scother		
appalled		
buttle		
wray		
batty		
preachet		
shrewd		
whim		
candid		
standard		
floppy		
rouble		
skylark		
creased		
libel		
birth		
yam		
blaspheme		
annual		
throng		
canonical		
patience		
dinosaur		

	YES	NO
beagle		
naked		
obsolate		
peasantry		
soldier		
upset		
ancient		
flesh		
hype		
aware		
adair		
drawer		
jovial		
victory		
limpid		
stone		
original		
lovely		
regrain		
galpin		
gauche		
time		
poor		
mundy		
bance		
stace		
measles		
bawdy		
upbeat		
debt		

Appendix B: Vocabulary immediate and delayed posttest

Multiple choice test. Please choose the meaning of the following words:

1. smock
   a) a long loose shirt
   b) false appearance
   c) a type of facial cream
   d) determination
   e) I don't know
2. libel
   a) a corrupt plan, especially for getting money
   b) the illegal act of writing things about someone that are not true
   c) holding an opinion and sharing it without careful thought
   d) the act of stealing money that people trust you to look after as part of your work
   e) I don't know
3. hype
   a) something you say that is not true
   b) the act of emphasizing what you're saying
   c) sounds made by voices or instruments
   d) the use of a lot of publicity to influence people
   e) I don't know
4. throng
   a) a large number of people crowded or assembled together
   b) someone whose job is to report the news for a newspaper
   c) a small group of people who have a lot of advantages
   d) someone who belongs to an organization
   e) I don't know
5. yam
   a) a sweet sticky food made from boiled fruit and sugar
   b) an African tree known for its sweet, yellow fruit
   c) a root vegetable that looks like a long white potato
   d) large grey bird, originally from Africa
   e) I don't know
6. measles
   a) an extended shortage of a basic nutrient
   b) large groups of insects flying or moving together
   c) an uncontrolled increase in the numbers of an insect
   d) an infectious disease characterized by red spots all over the body
   e) I don't know

7. creased
    a) full of lines
    b) not pale in colour
    c) completely even
    d) full of brown spots
    e) I don't know
8. appalled
    a) obvious or very easily noticed
    b) hurt by hitting or kicking
    c) offended or shocked
    d) with skin that is lighter than usual
    e) I don't know
9. floppy
    a) done in a very careless way
    b) spending or costing a lot of money
    c) lacking power or influence
    d) soft and hanging down in a loose way
    e) I don't know
10. batty
    a) careful and using good judgment
    b) not able to see clearly
    c) slightly crazy
    d) refusing to listen to other opinions
    e) I don't know
11. shrewd
    a) clever and able to make good judgments
    b) behaving in a way that is not polite
    c) morally bad, dangerous, or frightening
    d) caring about other people more than about oneself
    e) I don't know
12. wacky
    a) famous or successful
    b) morally wrong
    c) funny or silly
    d) not strong or definite
    e) I don't know

The relationship between extramural English and learners' listening comprehension, reading comprehension, motivation, and anxiety

Michiel Wouters
Department of Linguistics
KU Leuven
michiel-wouters@outlook.com

Lieven Bollansée
Department of Linguistics
KU Leuven
lieven.bollansee@gmail.com

Elien Prophète
Department of Linguistics
KU Leuven
elien.prophete@kuleuven.be

Elke Peters
Department of Linguistics
KU Leuven
elke.peters@kuleuven.be

Abstract

Recently, researchers have shown an increased interest in the relationship between extramural English and learners' language proficiency. However, the majority of the studies have focused on vocabulary knowledge. Less is known about extramural English and learners' motivation, reading and listening proficiency. Further, few studies have focused on English-as-a-foreign language learners attending technical education. The main objective of this study is to investigate whether extramural English is related to the motivation, language anxiety, listening and reading proficiency of three distinct groups of participants: grade 6 (age 11-12), grade 8 (age 13-14) and grade 10 (age 15-16) learners. Data were collected from 108 learners. All participants were administered a questionnaire and a listening and reading comprehension test. The findings suggest that learners are frequently exposed to English in their spare time. The results also indicate that grade 6 learners were already capable of performing listening tasks at

the A2 level of the Common European Framework of Reference for Languages, even though they had not had any English lessons in school yet. Watching non-subtitled TV in English appeared to be positively related to both listening and reading proficiency. Finally, relationships were found between various extramural English activities and factors concerning motivation and anxiety.

Keywords: extramural English; listening; reading; motivation; language anxiety.

Résumé

Récemment, les chercheurs ont montré un intérêt accru pour la relation entre l'anglais extrascolaire et les compétences linguistiques des apprenants. Cependant, la majorité des études se sont concentrées sur la connaissance du vocabulaire. On dispose de moins d'informations sur l'anglais extrascolaire et la motivation des apprenants, ainsi que sur leurs compétences en lecture et en écoute. En plus, il y a peu d'études qui se sont concentrées sur les apprenants d'anglais en tant que langue étrangère dans le cadre de l'enseignement technique. Cette étude a pour principal objectif de déterminer si l'anglais extrascolaire est lié à la motivation, à l'anxiété linguistique et aux compétences en lecture et en écoute de trois groupes distincts de participants : la sixième année (11-12 ans), la huitième année (13-14 ans) et la dixième année (15-16 ans). Les collectes de données ont été effectuées auprès de 108 apprenants. Tous les participants ont été soumis à un questionnaire et à un test de compréhension orale et écrite. Les résultats indiquent que les apprenants sont fréquemment exposés à l'anglais pendant leur temps libre. Les résultats révèlent également que les apprenants de 6e année étaient déjà capables de passer des tests d'écoute au niveau A2 du Cadre européen commun de référence pour les langues, même s'ils n'avaient pas encore suivi de cours d'anglais à l'école. Le fait de regarder des émissions télévisées sous-titrées en anglais semblait être positivement lié aux compétences en matière de compréhension orale et écrite. Enfin, on a constaté des relations entre diverses activités d'anglais extrascolaire et des facteurs concernant la motivation et l'anxiété.

Mots clés : anglais extrascolaire ; écoute ; lecture ; motivation ; anxiété linguistique.

1. Introduction

In recent years, there has been an increasing interest in the relationship between learners' extramural English and their language proficiency (e.g., Bollansée et al., 2021; De Wilde et al., 2020; Hannibal Jensen, 2017; Peters, 2018; Sundqvist, 2019). Extramural English is a term coined by Sundqvist (2019) and refers to different types of engagement with English through out-of-school activities such as viewing TV, reading, playing video games, and using the internet. Research has shown that extramural

English may foster English-as-a-foreign language (EFL) learners' language proficiency (Busby, 2020; Lindgren & Muñoz, 2013; Peters, 2018; Sundqvist, 2019). Additionally, several studies have demonstrated that language learning through extramural English can occur in children prior to formal instruction (De Wilde et al., 2020; Kuppens, 2010; Puimège & Peters, 2019). However, most studies investigating the relation between extramural English and second language (L2) proficiency have focused on vocabulary knowledge, whereas fewer studies have examined listening comprehension, reading comprehension, writing, and speaking (see De Wilde et al., 2020; Lindgren & Muñoz, 2013; Muñoz & Cadierno, 2021; Sylvén & Sundqvist, 2012, for exceptions). Additionally, few studies have explored the relationship between motivation and L2 competence. An exception is Leona et al. (2021). The same holds true for language anxiety. Both motivation and anxiety are learner-related variables that are believed to influence second language acquisition (SLA; Bernaus & Gardner, 2008; Dörnyei, 2005). Finally, the field of SLA has been found to suffer from biased sampling, with most studies collecting data from university students. Younger learners tend to be underrepresented.

The present study aimed to fill these gaps by investigating whether extramural English was related to learners' English listening and reading proficiency, motivation, and anxiety. Further, the present study was conducted with young learners and learners from the technical strand of secondary education in Flanders (the Dutch-speaking region in Belgium) in order to increase the generalizability of findings (Andringa & Godfroid, 2020).

The current research article opens with a literature review, discussing various extramural English activities as well as motivation and language anxiety. The next sections present the research questions, methodology and results. Finally, the paper ends with a discussion of the findings and a conclusion, including the limitations and pedagogical implications.

2. Background

2.1. Extramural English

2.1.1. TV viewing

Webb (2015) argued that extensive TV viewing might be an effective method to acquire L2 vocabulary and listening skills. He stated that watching L2 television extensively "could help to fill the need for greater L2 input" (p. 159) that is typically lacking in many foreign language learning contexts. Webb defines extensive viewing as "regular silent uninterrupted viewing of L2 television inside and outside of the classroom" (p. 159).

Recent studies have indicated that EFL learners of different ages frequently engage in watching English TV programs and movies outside of the classroom (Bollansée et al., 2021; De Wilde et al., 2020; Kuppens, 2010; Lindgren & Muñoz, 2013; Peters, 2018). Most studies have focused on the relation between watching TV and EFL learners' vocabulary knowledge. For instance, in her study conducted with adolescent EFL learners (16 and 19 years old), Peters (2018) found that viewing non-subtitled TV programs and movies correlated positively with learners' vocabulary knowledge. This finding was corroborated by Warnby (2022), who collected data from Swedish EFL learners aged 16 and 18. Studies have also shown that watching TV and movies may also have a positive effect on learners' knowledge of phrasal verbs (Schmitt & Redwood, 2011) and collocations (González-Fernández & Schmitt, 2015).

Multiple studies have also indicated that watching English TV may be beneficial to the vocabulary knowledge of young learners prior to formal English instruction. For instance, in Kuppens' (2010) study, it was found that watching subtitled television or movies was an important predictor of the Dutch-to-English and English-to-Dutch translation skills of Flemish children in their last year of primary education (grade 6). Puimège and Peters' study (2019), with 560 pupils in grades 4, 5 and 6 (age 10-12), also claimed passive exposure (e.g., watching TV with Dutch subtitles, listening to songs) to be beneficial to vocabulary knowledge. The study by Bollansée et al. (2021) found a positive relationship between extramural English and young Flemish learners' productive word knowledge prior to formal instruction. Watching TV without subtitles appeared to be the most beneficial, whereas watching TV with subtitles in the L1 was negatively correlated with the young learners' word knowledge.

Fewer studies have investigated the relationship between extramural English and reading and listening proficiency (De Wilde et al., 2020; Lindgren & Muñoz, 2013; Muñoz & Cadierno, 2021). A study by Lindgren and Muñoz (2013), with 10- and 11-year-old EFL learners in their fourth year of L2 instruction, found that watching TV and movies in the L2 (with or without subtitles) was an important predictor of learners' listening and reading proficiency. Muñoz and Cadierno (2021) found a significant positive relationship between exposure to audiovisual input with English subtitles and the listening test scores of 14- and 15-year-old Spanish EFL learners. In contrast, De Wilde et al. (2020), who collected data from Flemish children aged 10 to 12 without English instruction, did not find any relationship between watching English-spoken TV with or without subtitles and reading and listening proficiency. However, it remains unclear whether these findings can be extended to older EFL learners in secondary education and EFL learners in technical education in particular.

2.1.2. Written input

Most studies with primary and secondary school EFL learners have shown that EFL learners do not read very often outside of school (e.g., De Wilde et al., 2020; Lindgren & Muñoz, 2013; Peters, 2018). However, some studies have found a positive relationship between extramural reading and vocabulary knowledge. Yet, most of these studies involved EFL learners in secondary education or at university. For instance, in her study with 16-year-old EFL learners and university students, Peters (2018) found a positive correlation between extramural reading and vocabulary knowledge. Similarly, Warnby (2022) showed that reading outside of school was an important predictor of Swedish EFL learners' (age 16 and 18) knowledge of academic vocabulary. A study by Busby (2020) also revealed a positive relationship between the extramural reading habits of Norwegian university students and their scores on a vocabulary test. Similar findings have been revealed for learners' knowledge of multiword items like collocations (González-Fernández & Schmitt, 2015) or phrasal verbs (Schmitt & Redwood, 2011). Another study (De Wilde et al., 2020) with Dutch-speaking children without English classroom instruction (aged 10-12) found a positive relationship between reading English books and listening, reading and writing skills. Yet, in the regression analyses, extramural reading was no longer a significant predictor. Lindgren and Muñoz (2013) explored the correlation between out-of-school exposure and the receptive language proficiency of 10- and 11-year-old European children. However, reading was too infrequent among the children and was therefore not included in their analysis.

2.1.3. Computer use

A third important type of extramural English for EFL learners is computer use, which entails multiple activities such as playing computer games, navigating the internet, and using social media (Sundqvist & Sylvén, 2016). Multiple studies have put forward playing computer games as a potentially effective method to enlarge EFL learners' vocabulary (Bollansée et al., 2021; Busby, 2020; De Wilde et al., 2020; Sylvén & Sundqvist, 2012).

Studies conducted with young learners (ages 6 to 12) all point to the positive relationship between gaming and English language proficiency. The findings of Sylvén and Sundqvist (2012) showed that frequent gamers, aged 11-12, knew significantly more English words and obtained higher scores on reading and listening comprehension than non-gamers. Similarly, Hannibal Jensen (2017) found a positive relationship between gaming with both spoken and written English input and the vocabulary scores of Danish English language learners aged 8 and 10. Several studies have focused on young learners who have not received English instruction yet. These studies corroborated the

positive relationship between gaming and receptive vocabulary knowledge (Puimège & Peters, 2019), productive vocabulary knowledge (Bollansée et al., 2021), and speaking and listening comprehension (De Wilde et al., 2020).

Gaming has also been found to be beneficial for adolescent EFL learners and university learners. Studies conducted by Sundqvist and colleagues (Sundqvist, 2019; Sundqvist & Wikström, 2015) consistently showed a positive effect of gaming on vocabulary knowledge and vocabulary use in writing. These findings were confirmed in Peters (2018) and Peters et al. (2019). In her study with university students, Busby (2020) found a strong positive correlation between their frequency of gameplay in English and their scores on a vocabulary test. In her study, the majority of students stated that they were gaming in English at least sometimes.

Compared to the number of studies on vocabulary, few studies have examined the relation between computer use and listening and reading comprehension (see De Wilde et al., 2020; Lindgren & Muñoz, 2013 for studies focusing on young learners). Little is known about the relationship between extramural English and adolescents' listening and reading comprehension. The present study aims to fill this research gap by focussing on three distinct groups of participants.

2.1.4. Music

A final extramural English activity is listening to music. Several studies have revealed that EFL learners are frequently exposed to English songs (Bollansée et al., 2021; De Wilde et al., 2020; Peters, 2018). However, it is not yet clear whether listening to English music can predict learners' language proficiency, since studies have yielded contradictory findings. De Wilde et al. (2020) found a negative relationship between listening to music and the receptive vocabulary, as well as the reading comprehension and writing of Dutch-speaking children aged 10 to 12 who had not received English lessons yet. Other studies have found a positive relationship between listening to English songs and language proficiency. A study by Lindgren and Muñoz (2013) with young learners in their fourth year of foreign language instruction revealed a positive correlation between listening to songs with English lyrics and reading and listening comprehension. With regard to secondary school learners and university students, Peters (2018) only found a negligible negative relationship between listening to English music and vocabulary knowledge. Therefore, it appears that the effect of English songs may be dependent on age, years of formal instruction, and type of language proficiency.

2.2. Motivation and anxiety

Motivation towards L2 learning concerns an individual's personal reasons, effort, and willingness to learn a foreign language (Dörnyei & Ushioda, 2011) and is

considered to be one of the most decisive factors influencing L2 proficiency (Gardner, 2006; Dörnyei, 2005). One of the dominant theoretical frameworks that emerged in the study of L2 acquisition is Dörnyei's (2005) L2 Motivational Self System. This model consists of three key components: the Ideal L2 Self, the Ought-to L2 Self, and the L2 Learning Experience. The Ideal L2 Self refers to the L2-specific dimension of the language learner's ideal self and comprises internal desires or wishes. The second component, the Ought-to L2 Self, consists of the learner's more extrinsic motives for learning an L2, for instance, meeting societal expectations and avoiding possible negative outcomes. The third component of the L2 Motivational Self System, L2 Learning Experience, is concerned with the L2 learner's learning environment and past and current experiences, such as interaction with the teacher, the peer group, the experiences of successes or failures, and the curriculum.

Few studies have looked at the relationship between motivation and extramural English. Leona et al. (2021) examined whether various motivational factors mediated the relationship between extramural English activities and the English vocabulary knowledge of Dutch primary school children with and without formal English instruction. Path analyses showed that for children learning English at school, the relationship between extramural English, through entertaining media and family members, and English vocabulary knowledge was mediated by the motivational factor of linguistic self-confidence. In her Ph.D. dissertation, Arndt (2019) studied the interaction between motivation and attitudes towards language learning, engagement in informal L2 practices, and general language proficiency of German secondary school students by means of qualitative and quantitative analyses. Her findings suggest that the language learners' Ideal L2 selves, their reasons for wanting to learn English, and their attitudes towards informal and formal language learning significantly predicted engagement in informal L2 practices. Furthermore, a reciprocal relationship was found between engagement in informal L2 practices and motivation. In an Indonesian university context, Lee and Drajati (2019) examined the relation between digital learning of English activities, L2 willingness to communicate and multiple affective variables (motivation, self-confidence, L2 speaking anxiety and grit). Their study found significant correlations between L2 willingness to communicate, and informal digital learning of English activities and affective variables. However, the only significant predictors of students' L2 willingness to communicate were found to be grit, self-confidence, motivation, and productive informal digital learning activities (e.g., chatting with others in English via social media). Lee and Lee (2020) studied whether the Ideal L2 Self, the Ought-to L2 Self and digital extramural English activities were related to the foreign language enjoyment of South Korean middle school, high school and university students. The Ideal L2 Self and digital language learning activities were found to be predictive of all groups' foreign language enjoyment.

Language anxiety, which according to Bernaus and Gardner (2008) can be divided into situation specific L2 use anxiety and L2 class anxiety, is believed to have a direct negative effect on L2 acquisition. Foreign language classroom anxiety can be defined as "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process" (Horwitz et al., 1986, p. 128)

To date, only a few studies have investigated the relationship between language anxiety and extramural English exposure. An exception is Leona et al. (2021), who found a positive correlation between using English in communication with friends and willingness to communicate with peers for pupils without formal English instruction. Furthermore, for pupils learning English formally, positive relationships were found between formal reading, familial extramural English exposure, and the use of English in interaction with friends, on the one hand, and willingness to communicate with peers on the other. Another exception is the aforementioned study of Lee and Drajati (2019).

3. The present study

Whereas the majority of previous studies have focused on EFL learners' vocabulary knowledge, to date, few studies have investigated the relation between extramural English and learners' listening and reading comprehension. Additionally, little research has been conducted on the relationship between extramural English and learners' motivation and anxiety. In spite of the increasing number of studies on extramural English, only a limited number of studies have focused on language learners in technical secondary education. Furthermore, studies with a cross-sectional design are scarce. To fill these research gaps, the present cross-sectional study includes an underrepresented group of EFL learners, that is, EFL learners from the technical strand of grade 10 (secondary education), in addition to grade 6 pupils without formal English instruction and grade 8 EFL learners who had just started formal English instruction.

The research questions addressed in this study are:

1. To what extent do learners in different grades differ in their engagement with extramural English?

2. What is the relationship between extramural English and EFL learners' listening and reading comprehension?

3. To what extent are motivation and language anxiety related to EFL learners' extramural English exposure?

4. Methods

4.1. Participants

We used convenience sampling to recruit a total of 108 participants: 36 pupils in grade 6 (age 11-12), 33 pupils in grade 8 (age 13-14), and 39 pupils in grade 10 (age 15-16). Each group of EFL learners was recruited from a different school in Flanders, Belgium. Grade 8 learners attended A-stream secondary education, a preparatory programme for general and technical secondary education. Grade 10 participants attended the technical strand of secondary education.

Participants in grade 6 had not received English classes, as English is not part of primary schools' curricula in Flanders. Participants in grade 8 had received approximately two hours of English classes per week over a period of five months, and participants in grade 10 had received approximately two hours of English classes per week for two years and five months. Table 1 gives an overview of the characteristics of the three groups of participants.

Among the 108 participants, 90 indicated that they mainly spoke Dutch at home, 10 reported speaking Dutch and one or more other languages at home. Finally, eight participants indicated that they did not usually speak Dutch with their parents or siblings. They spoke Portuguese, Polish or Spanish at home instead.

Table 1: Summary of characteristics of the three groups of participants

Education level	N	Male-female	Age
Grade 6	36	18 - 18	11-12
Grade 8	39	16 - 23	13-14
Grade 10	33	28 - 5	15-16

4.2. Test instruments

4.2.1. Listening test

Listening proficiency was tested by means of an extract of the European Survey on Language Competences (European Commission, 2012a) at the levels A1, A2, B1, and B2 of the Common European Framework of Reference for Languages (CEFR). Since we did not want to fatigue the young grade 6 pupils, we removed the level B2

task from their listening test. For the tasks at level A1 and A2, the participants were asked to listen to short conversations and to answer 4 and 5 questions respectively by selecting the correct pictures.

For the tasks at levels B1 and B2, the participants were required to listen to an interview with a singer-songwriter and a presentation of a former scout leader and answer six multiple-choice questions about each of them. Each question was accompanied by three answers. We did not provide "I don't know"-options. While the instructions and questions were provided in both English and Dutch, the answers were only made available in English.

4.2.2. Reading test

Similar to the listening test, the reading test was taken from the European Survey on Language Competences (European Commission, 2012a) and consisted of four tasks at CEFR levels A1, A2, B1, and B2. Again, task level B2 was removed from the reading test for grade 6 pupils. Each of the reading tasks was presented in the same format: a text box was placed at the top of the page with underneath four to six multiple-choice questions. Each question was accompanied by three answers to choose from. The multiple-choice questions and instructions were made available in both English and Dutch, while the text box and the three possible answers were only offered in English.

4.2.3. Questionnaire

A questionnaire was administered to assess the participants' current extramural English. The questionnaire was developed in Dutch and consisted of two parts. The first part comprised nine Likert-scale items, one *yes/no* item, and five open-ended questions. The following types of EE activities were addressed: TV and movies (with or without subtitles), video games, music, streaming videos (YouTube), written or spoken communication with family or friends, written texts (books, cartoons, newspapers). With regard to the Likert-scale items, the response categories were the following: never, sometimes, often, always.

One part of the questionnaire focused on the participants' behaviour when encountering a word whose meaning they did not understand. However, we do not report on this part of the questionnaire, as it is beyond the scope of this study.

The participants in grades 8 and 10 were asked to answer six questions regarding motivation and language anxiety. Grade 6 participants did not have to answer these questions, as some questions focused on classroom instruction. Participants were asked to what degree they agreed with these statements on a four-point scale. Two questions

were related to learners' Ideal L2 Self, two questions to learners' Ought-to L2 Self (Dörnyei, 2005), and two questions to learners' foreign language anxiety while speaking.

4.2.4 Procedure

The data collection took place in three schools in the Flemish community of Belgium in January 2021. Because the study involved children and underage adolescents, the ethical guidelines developed by the university were strictly adhered to. The participation in the data collection was completely voluntary, and the requirements of the General Data Protection Regulation (GDPR) were met. Before collecting data, participants were informed about the scope of the study and were asked to complete written consent forms. Respondents were instructed to also ask their parents' consent before participating in the study by means of a second consent form.

First, the participants were given approximately 10 minutes to fill in the questionnaire. Due to time constraints, grade 6 participants were instructed to fill in the extramural English questionnaire at home. Second, the listening test was administered and completed within approximately 20 minutes. The participants were allowed to listen to each recording twice while completing the test. Third, they were given 20 minutes to complete the reading test autonomously. To reduce the influence of random guessing on the test results, we instructed the participants to leave blank any questions they could not answer. The tests did not include "I don't know"-options. All participants were able to complete the data collection within the allotted time frame of 50 minutes.

4.3. Data analysis

The listening and reading comprehension test were scored dichotomously: correct responses received a score of 1, incorrect responses received a score of 0. Missing responses were also coded as 0. A score of at least 80% of the maximum score on a task would be considered full mastery of English at the corresponding CEFR level (European Commission, 2012b).

To determine differences, i.e., whether the three grades differ in their extramural English, Kruskal-Wallis tests were run. To explore the relationship between learners' extramural English and their listening and reading proficiency, on the one hand, and their motivation/anxiety on the other, Spearman's rank correlations were computed. We used Bonferroni correction to control for multiple comparisons. In other words, the p-level of 0.05 was adapted to 0.006 (0.05/9). Multiple regression analyses were used to determine the variables that predicted learners' listening and reading comprehension.

5. Analyses

5.1. *Participants' engagement with extramural English*

The results of the extramural English questionnaire showed that Flemish primary and secondary school learners are regularly exposed to English outside the classroom (see Table 2). Listening to English-language music was the most popular extramural English activity for the three groups of participants. The second most popular activity was watching YouTube videos. With regard to watching English TV, a distinction was made between viewing TV programs or movies without subtitles, with Dutch subtitles, with English subtitles, and with subtitles in another language. Watching TV with Dutch subtitles was the most popular one. More than half of the participants of each group indicated that they watched English TV with Dutch subtitles "often or always". A high percentage of participants reported that they, at least "sometimes", spoke or wrote in English with family members or friends. Finally, among all groups, extramural reading was less popular: around half of the pupils in each group reported to "never" read English texts. In spite of some differences in frequency of the different extramural English activities, there were no significant differences between the three grades.

Table 2: Contingency tables extramural English

Frequency of watching English language TV without subtitles					
	n	Never	Sometimes	Often	Always
Grade 6	36	41.7%	30.6%	13.9%	13.9%
Grade 8	39	23.1%	23.1%	38.5%	15.4%
Grade 10	33	24.2%	24.2%	36.4%	15.2%
Frequency of watching English language TV with Dutch subtitles					
	n	Never	Sometimes	Often	Always
Grade 6	36	17.9%	28.2%	33.3%	20.5%
Grade 8	39	17.9%	28.2%	33.3%	20.5%
Grade 10	33	18.2%	18.2%	27.3%	36.4%
Frequency of watching English language TV with English subtitles					
	n	Never	Sometimes	Often	Always
Grade 6	36	61.1%	19.4%	19.4%	0.0%
Grade 8	39	41.0%	28.2%	20.5%	10.3%
Grade 10	33	42.4%	33.3%	21.2%	3.0%

Frequency of watching English language TV with English subtitles					
	n	Never	Sometimes	Often	Always
Grade 6	36	83.3%	13.9%	0.0%	2.8%
Grade 8	39	82.1%	10.3%	5.1%	2.6%
Grade 10	33	81.8%	15.2%	3.0%	0.0%
Frequency of playing video games in English					
	n	Never	Sometimes	Often	Always
Grade 6	36	8.3%	11.1%	41.7%	38.9%
Grade 8	39	12.8%	20.5%	33.3%	33.3%
Grade 10	33	12.1%	21.2%	36.4%	30.3%
Frequency of listening to English language music					
	n	Never	Sometimes	Often	Always
Grade 6	36	0.0%	5.6%	66.7%	27.8%
Grade 8	39	0.0%	12.8%	25.6%	61.5%
Grade 10	33	0.0%	9.1%	51.5%	39.4%
Frequency of watching streaming videos in English					
	n	Never	Sometimes	Often	Always
Grade 6	36	11.1%	27.8%	36.1%	25.0%
Grade 8	39	5.1%	20.5%	35.9%	38.5%
Grade 10	33	3.0%	27.3%	33.3%	36.4%
Frequency of reading English texts (newspapers, books, online texts, ...)					
	n	Never	Sometimes	Often	Always
Grade 6	36	55.6%	27.8%	11.1%	5.6%
Grade 8	39	43.6%	43.6%	12.8%	0.0%
Grade 10	33	54.5%	33.3%	12.1%	0.0%

5.2. Listening and reading comprehension test

The group of grade 10 participants achieved the highest average score on all test levels (A1, A2, B1, and B2), followed by grade 8 participants. Participants obtained higher scores for listening compared to reading. Table 3 and Table 4 give an overview

of the distribution of the participants' highest CEFR attained for listening and reading comprehension respectively, based on the 80% target, as well as the total mean score on the tasks at CEFR levels A1, A2 and B1.

Table 3: Distribution of participants' CEFR level and mean score for listening comprehension

	Pre-A1	A1	A2	B1	B2	Mean score (%) A1-B1
Grade 6 (n=36)	0	8	21	7		72.59%
Grade 8 (n=39)	2	5	19	12	1	77.95%
Grade 10 (n=33)	0	3	9	11	10	87.47%
Total (n=108)	2	16	49	30	11	79.07%

Table 4: Distribution of participants' CEFR level and mean score for reading comprehension

	Pre-A1	A1	A2	B1	B2	Mean score (%) A1-B1
Grade 6 (n=36)	2	20	7	7		53.97%
Grade 8 (n=39)	4	23	5	5	2	57.88%
Grade 10 (n=33)	1	12	4	9	7	72.94%
Total (n=108)	7	55	16	21	9	61.18%

5.2. Relationship between out-of-class exposure and learners' listening and reading proficiency

Four extramural English activities correlated with learners' listening comprehension (see All participants in Table 5). The strongest correlation was found for watching TV without subtitles, followed by communicating with family or friends, watching YouTube videos and reading. A multiple regression analysis in SPSS (version 28) was run to determine the predictors of listening. The analysis showed that watching TV without subtitles (β = 0.33, $p < 0.001$) and communicating with friends or families (β= 0.22, $p = 0.03$) significantly predicted learners' listening comprehension ($R^2 = 0.23$, $F(2,105) = 15.42$, $p < 0.001$).

Table 5: Summary of correlations (Spearman's rank) between types of extramural English and listening comprehension

	Listening			
	r_s (*p*)			
	All participants	Grade 6 (n = 36)	Grade 8 (n = 39)	Grade 10 (n = 33)
TV no subtitles	.44* (<.0001)	.50* (.002)	.43* (.006)	.32 (.07)
TV Dutch subtitles	-.21 (.01)	-.13 (.46)	-.24 (.14)	-.42 (.01)
TV English subtitles	.20 (.02)	.26 (.13)	.17 (.31)	.30 (.09)
TV subtitles in other language	-.04 (.36)	.13 (.46)	-.09 (.57)	-.16 (.38)
Video games	.17 (.047)	.36 (.03)	-.10 (.57)	.37 (.03)
Music	.17 (.04)	.13 (.44)	.20 (.22)	.25 (.16)
Streaming videos (YouTube)	.33* (<.0001)	.44 (.007)	.15 (.36)	.23 (.21)
Communicating with family or friends	.38* (<.0001)	.31 (.07)	.49* (.002)	.26 (.14)
Reading (books, cartoons, newspapers, internet, ...)	.31* (.001)	.45* (.006)	.29 (.07)	.27 (.13)

*Correlation is significant at the 0.006 level (2-tailed) (with Bonferroni correction for multiple comparisons).

As can be seen in Table 6, there were two (small to moderate) positive correlations: between reading comprehension and watching TV without subtitles and between reading comprehension and communicating with friends or family. The regression analysis showed that watching TV without subtitles (β = 0.32, p = 0.002) was the only predictor of reading comprehension (R^2 = 0.16, F(1, 106) = , p < 0.001).

Table 6: Summary of correlations (Spearman's rank) between types of extramural English and reading comprehension

	Reading			
	r_s (*p*)			
	All participants	Grade 6 (n = 36)	Grade 8 (n = 39)	Grade 10 (n = 33)
TV no subtitles	.39* (<.0001)	.47* (.004)	.10 (.53)	.54* (.001)
TV Dutch subtitles	-.19 (.03)	-.22 (.19)	-.02 (.92)	-.34 (.05)
TV English subtitles	.22 (.01)	.09 (.59)	.31 (.06)	.24 (.17)

TV subtitles in other language	-.05 (.29)	.10 (.56)	.02 (.91)	-.09 (.61)
Video games	.23 (.008)	.43* (.008)	-.05 (.77)	.41 (.02)
Music	.04 (.35)	-.13 (.46)	.11 (.49)	.09 (.63)
Streaming videos (YouTube)	.24 (.007)	.26 (.13)	.13 (.44)	.21 (.25)
Communicating with family or friends	.26** (.003)	.10 (.57)	.31 (.06)	.41 (.02)
Reading (books, cartoons, newspapers, internet, ...)	.21 (.016)	.24 (.17)	.24 (.14)	.30 (.10)

*Correlation is significant at the 0.006 level (2-tailed) (with Bonferroni correction for multiple comparisons).

5.4. Relationship between extramural English and motivation and anxiety

The results revealed that all participants believe that they will need English in the future. However, grade 8 participants agreed more with this statement than grade 10 pupils. The results regarding their future L2 self were more varied: about one in ten participants could not imagine themselves speaking English fluently in the future.

Furthermore, the majority of the participants indicated that they did not feel anxious when speaking English in class, while about one third of both the grade 8 and grade 10 learners agreed with the statement regarding in-class anxiety.

Fewer learners indicated that they felt anxiety when speaking English outside the classroom: 13% of the grade 8 and 24% of the grade 10 participants agreed with the statement concerning out-of-class anxiety.

Because the third research question is exploratory in nature, only correlation analyses are reported (see Table 7). With regard to the components of Dörnyei's (2005) L2 Motivational Self System, some significant positive relationships were found. For instance, exposure to English language TV and movies without subtitles and reading related positively to two of the four factors included in the questionnaire. Furthermore, positive correlations were found regarding speaking English with family and friends, and three of the four motivation factors, while listening to music was only related to the Ideal L2 Self Future Need.

We also found multiple significant correlations concerning in-class and out-of-class anxiety. Watching TV without subtitles and with English subtitles, watching streaming videos (YouTube), speaking English with family or friends and reading English texts, all of these extramural activities correlated negatively with the participants' anxiety levels. This means that learners who engaged more frequently with English outside of school tended to have lower anxiety levels.

Table 7: Summary of correlations (Spearman's rank) between extramural English and factors regarding motivation and anxiety (n = 72)

	Ideal L2 Self Future Need	Ideal L2 Self Future Image	Ought-to L2 Self Parents	Ought-to L2 Self Friends	Anxiety classroom	Anxiety use
	rs (*p*)	*rs* (*p*)	*rs* (*p*)	*rs* (*p*)	*rs* (*p*)	*rs* (*p*)
TV no subtitles	.29 (.01)	.46** (<.001)	-.03 (.82)	.39** (<.001)	-.39** (<.001)	-.42** (<.001)
TV Dutch subtitles	-.17 (.16)	-.26 (.03)	.21 (.07)	-.15 (.21)	.26 (.03)	.32 (.01)
TV English subtitles	.22 (.06)	.24 (.04)	-.05 (.70)	.06 (.64)	-.16 (.18)	-.36** (.002)
TV subtitles in other language	.08 (.52)	-.12 (.30)	-.07 (.55)	.15 (.21)	.08 (.53)	-.06 (.61)
Video games	.12 (.32)	.22 (.07)	-.03 (.82)	.15 (.20)	-.20 (.09)	-.30 (.01)
Music	.38* (<.001)	.31 (.009)	.12 (.31)	.39* (<.001)	-.12 (.33)	-.26 (.03)
Streaming videos (YouTube)	.21 (.08)	.32 (.01)	.03 (.82)	.19 (.11)	-.27 (.02)	-.45* (<.001)
Communicating with family or friends	.38* (.001)	.45* (<.001)	.15 (.21)	.39* (<.001)	-.42* (<.001)	-.56* (<.001)
Reading (books, cartoons, newspapers, internet, ...)	.20 (.09)	.34* (.004)	.06 (.61)	.28 (.02)	-.40* (<.001)	-.40* (<.001)

*Correlation is significant at the 0.006 level (2-tailed) (with Bonferroni correction for multiple comparisons).

6. Results

6.1. *Summary of findings*

The results of this study showed that the three groups of participants were regularly exposed to English in their spare time. We also found that they did not differ in their engagement with extramural English. The findings suggest that Flemish learners are mainly exposed to spoken English input: songs, streaming videos, video games, movies and TV, and communicating with family or friends. Reading English texts appeared to be the least popular extramural English activity among the three

groups of participants. Our findings are in agreement with previous research that has shown that EFL learners are frequently exposed to songs, video games and audiovisual input, but rarely engage in reading English texts (e.g., Bollansée et al., 2021; De Wilde et al., 2020; Lindgren & Muñoz, 2013; Puimège & Peters, 2019; Sundqvist, 2019)

This study suggests that a great number of primary school pupils had already achieved level A2, and some even level B1, for listening prior to formal English instruction. These findings are in line with those of De Wilde et al. (2020), who reported that a quarter of the grade 6 participants in their study could already perform listening tasks at the A2-level before the start of formal English instruction. Additionally, grade 8 learners, who attended a technical school, had already achieved a high level of listening proficiency at the start of formal English instruction. Finally, grade 10 participants had the highest scores and most obtained a B1 or B2 level, even though they had only received approximately two hours of English classes per week for two years and five months. This finding is also in line with Peters et al.'s (2020) results, which showed a ceiling effect (90%) for listening at B1 level in grade 10.

Compared to the listening comprehension scores, the reading comprehension scores were lower. This holds true for all three grades. Nevertheless, 20 learners in grade 6 obtained an A2 and seven learners even attained a B1 level without ever having had an English lesson. This finding is consistent with that of De Wilde et al. (2020), who reported that 10% of the participants obtained an A2-level for reading and writing, even though the results in the present study point to a slightly higher level.

6.2. Relationship between out-of-class exposure and learners' listening and reading proficiency

The present study shows that there were positive correlations between several types of extramural English activities and learners' listening or reading comprehension.

We found that watching TV without subtitles, communicating with friends and family, watching YouTube clips, and reading correlated positively with participants' listening comprehension. Our findings extend those of Peters (2018), who focused on learners' vocabulary knowledge in the academic strand of secondary education. The regression analysis showed that two activities were significant predictors of listening comprehension: watching TV and communicating. These findings provide further support for the positive effect of watching English TV on learners' L2 proficiency (Bollansée et al., 2021; Kuppens, 2010; Lindgren & Muñoz, 2013; Puimège & Peters, 2019). Furthermore, in line with Bollansée et al. (2021), the findings suggest that non-subtitled TV may be more beneficial than L1-subtitled TV. As has been shown in previous research (De Wilde et al., 2020), using English outside of school is an important predictor of learners' language proficiency.

We also found positive correlations between reading comprehension and watching TV without subtitles on the one hand, and between reading comprehension and communicating with friends or family, on the other. However, the correlations were less strong compared to the listening comprehension. A possible explanation for this might be that these extramural English activities involve very little textual input. The findings are in line with those of De Wilde et al. (2020), who tested reading and writing in an integrated fill-in-the-gap test. Watching TV without subtitles was the only significant predictor of reading comprehension. Our results confirm those of Lindgren and Muñoz (2013), who focused on younger language learners and found that TV was the most important predictor of the extramural activities they investigated.

6.3. Relationship between extramural English and motivation and anxiety

Spearman's rank analyses indicated that there were multiple relationships between extramural English activities and factors regarding motivation and anxiety, even though caution is warranted given the limited number of questions about motivation and anxiety. For instance, watching English TV without subtitles, listening to English music, using English with family or friends, and reading English texts were all positively associated with at least one aspect related to the Ideal L2 Self. In other words, participants who often engaged in these extramural activities were more likely to believe that they would speak English fluently in the future or that they would need English in the future.

Our results corroborate the findings of Leona et al. (2021), who found a positive relationship between exposure to entertaining media and the motivational factor Linguistic Self-Confidence for a group of primary school pupils with formal English instruction. Furthermore, using English in communication with family was positively related to Linguistic Self-Confidence for this English-at-school group. The findings of this study are also in accordance with those of Arndt (2019), who found a reciprocal relationship between L2 learners' motivation and their engagement in informal L2 practices.

With regard to Ought-to L2 Self, the aspect "parental attitude" towards English did not correlate with any of the extramural English activities. The factor perceived attitudes of friends towards English, in contrast, was positively related to watching English TV without subtitles, listening to English music, and using English in interaction with family or friends.

Finally, the results showed negative relationships between EFL learners' out-of-class anxiety and five extramural activities included in the analyses. Learners who engaged more frequently in watching TV without subtitles or with English subtitles, watching YouTube videos, communicating with family or friends, or reading English-language books showed lower anxiety levels. Our findings concerning language anxiety

are in agreement with those of Leona et al. (2021). In their study, a positive relationship was found between using English to interact with friends and the factor Willingness to Communicate with Peers for primary school pupils with and without formal English instructions. Willingness to Communicate with Peers may partially overlap with the factor out-of-class anxiety included in the analyses of the present study.

7. Conclusion

The purpose of this study was to explore learners' extramural English and whether extramural English was related to the listening and reading proficiency of Flemish primary and secondary school learners. One-hundred and eight participants from three different grades were administered a questionnaire and a listening and reading comprehension test.

The results indicated that a great number of the primary school learners could already perform listening tasks at the A2 level of the CEFR prior to formal English instruction. The results of the questionnaire revealed that the three groups were frequently exposed to English-language media. The most popular extramural English activity was listening to English music. Watching TV without subtitles outside of school predicted both learners' listening and reading comprehension scores. Listening comprehension was also affected by the frequency with which learners communicated in English outside of school. In general, the in-class speaking anxiety of secondary school pupils appeared to be higher than their out-of-class anxiety. Finally, multiple relationships were found between extramural English and factors concerning motivation. We also found lower anxiety levels among learners who engaged more frequently with English outside of school.

However, this study inevitably has a few limitations. First, the extramural English questionnaire focused on a limited number of extramural English activities. With regard to written input, no distinction was made between reading English books, newspapers, magazines, or websites. Although suggested by Kuppens (2010), the questionnaire did not differentiate between different types of video game environment. Future studies could treat these types of input as separate variables to explore the relationship between extramural English and L2 proficiency more extensively.

Second, the motivation and language anxiety questions aimed at assessing L2 Motivational Self System and L2 speaking anxiety, but they were limited in scope, as the questionnaire included only six items. The current findings are, thus, exploratory in nature. A more elaborate and standardized questionnaire, such as the Motivation and Attitudes Questionnaire by Arndt (2019), the English Learner Questionnaire (ELQ; Dörnyei & Taguchi, 2010), or the Attitude/Motivation Test Battery (AMTB;

Gardner, 2004), can be adopted in future research to study the effects of motivation and anxiety on L2 acquisition in depth.

Finally, although this study yielded some results regarding the listening and reading proficiency of three age groups prior to and with formal English instruction, future research could adopt a longitudinal approach to obtain a more comprehensive picture of the potential long-term effects of extramural English, motivation, and anxiety on L2 proficiency.

The findings of this study have implications for SLA and pedagogy. For instance, since the present study found a positive relationship between watching TV without subtitles outside of school and listening and reading comprehension, it is recommended that L2 teachers integrate this activity in in-class or out-of-class assignments. Furthermore, based on the findings of this study, L2 teachers should acknowledge the importance of motivational factors and language anxiety in language acquisition and could ask their pupils to reflect on their Ideal L2 Self Image, as well as potential factors causing language anxiety. L2 teachers may also try to create a positive learning environment in order to decrease their pupils' speaking anxiety.

8. References

Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in Applied Linguistics. *Annual Review of Applied Linguistics, 40*, 134–142. https://doi.org/10.1017/s0267190520000033

Arndt, H. L. (2019). *Informal second language learning: The role of engagement, proficiency, attitudes and motivation.* [Doctoral dissertation, Department of Education, St. Hilda's College, University of Oxford].

Bernaus, M., & Gardner, R. C. (2008). Teacher motivation strategies, student perceptions, student motivation, and English achievement. *The Modern Language Journal, 92*(3), 387-401. https://doi.org/10.1111/j.1540-4781.2008.00753.x

Bollansée, L., Puimège, E., & Peters, E. "Watch out! Behind you is the enemy!" An exploratory study into the relationship between extramural English and productive vocabulary knowledge. In V. Werner & F. Tegge (Eds.), *Pop Culture in Language Education: Theory, Research, Practice* (pp. 199-214). London: Routledge. https://doi.org/10.4324/9780367808334

Busby, N. L. (2020). Words from where? *ITL - International Journal of Applied Linguistics, 172*(1), 58–84. https://doi.org/10.1075/itl.19018.bus

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types

of input are important? *Bilingualism: Language and Cognition, 23*(1), 171-185. https://doi.org/10.1017/S1366728918001062

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. New York, NY: Routledge. https://doi.org/10.4324/9781410613349

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing* (Second Language Acquisition Research Series) (2nd ed.). New York, NY: Routledge. https://doi.org/10.4324/9780203864739

Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). London: Routledge. https://doi.org/10.4324/9781315833750

European Commission. (2012a). *First European survey on language competences: Final report*. Publications Office of the European Union. https://doi.org/10.2766/34160

European Commission. (2012b). *First European survey on language competences: Technical report*. Publications Office of the European Union. https://ec.europa.eu/assets/eac/languages/policy/strategic-framework/documents/language-survey-technical-report_en.pdf

Gardner, R. C. (2004). *Attitudes/Motivation Test Battery: International AMTB research project*. Canada: The University of Western Ontario. https://publish.uwo.ca/~gardner/docs/englishamtb.pdf

Gardner, R. C. (2006). The socio-educational model of Second Language Acquisition: A research paradigm. *EUROSLA Yearbook, 6*(1), 237-260. https://doi.org/10.1075/eurosla.6.14gar

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? *ITL - International Journal of Applied Linguistics, 166*(1), 94–126. https://doi.org/10.1075/itl.166.1.03fer

Hannibal Jensen, S. (2017). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal, 34*(1), 1–19. https://doi.org/10.1558/cj.29519

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology, 35*(1), 65-85. https://doi.org/10.1080/17439880903561876

Lee, J. S., & Drajati, N. A. (2019). Affective variables and informal digital learning of English: Keys to willingness to communicate in a second language. *Australasian Journal of Educational Technology, 35*(5), 168–182. https://doi.org/10.14742/ajet.5177

Lee, J. S., & Lee, K. (2020). The role of informal digital learning of English and

L2 motivational self system in foreign language enjoyment. *British Journal of Educational Technology, 52*(1), 358–373. https://doi.org/10.1111/bjet.12955

Leona, N. L., van Koert, M. J. H., van der Molen, M. W., Rispens, J. E., Tijms, J., & Snellings, P. (2021). Explaining individual differences in young English language learners' vocabulary knowledge: The role of Extramural English Exposure and motivation. *System, 96*, 1–13. https://doi.org/10.1016/j.system.2020.102402

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism, 10*(1), 105–129. https://doi.org/10.1080/14790718.2012.679275

Muñoz, C., & Cadierno, T. (2021). How do differences in exposure affect English language learning? A comparison of teenagers in two learning environments. *Studies in Second Language Learning and Teaching, 11*(2), 185–212. https://doi.org/10.14746/ssllt.2021.11.2.2

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL-International Journal of Applied Linguistics, 169*(1), 142-168. https://doi.org/10.1075/itl.00010.pet

Peters, E., Noreillie, A-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning, 69*(3), 747–782. https://doi.org/10.1111/lang.12351

Peters, E., Noreillie, A-S., Puimège, E. (2020). Frans eerste vreemde taal en Engels tweede vreemde taal? Of toch niet? Een blik op de Engelse en Franse taalvaardigheid van jongeren in Vlaanderen. In W. Coudenys (Ed.), *Taal en de wereld* (pp.53-70). Leuven: Leuven University Press.

Puimège, E., & Peters, E. (2019). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning, 69*(4), 943–977. https://doi.org/10.1111/lang.12364

Schmitt, N. & Redwood, S. (2011). Learner knowledge of phrasal verbs. A corpus-informed study. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 173–209). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scl.45.12sch

Sundqvist, P. (2009). *Extramural English matters. Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary.* [Doctoral dissertation, Faculty of Arts, Karlstad University].

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology, 23*(1), 87–113. https://doi.org/10125/44674

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in Teaching and Learning.* London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-46048-6

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System, 51*, 65–76. https://doi.org/10.1016/j.system.2015.04.001

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL, 24*(3), 302–321. https://doi.org/10.1017/s095834401200016x

Warnby, M. (2022). Receptive academic vocabulary knowledge and extramural English involvement – is there a correlation? *ITL - International Journal of Applied Linguistics, 173*(1), 120–152. https://doi.org/10.1075/itl.21021.war

Webb, S. (2015). Extensive viewing: language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). New York, NY: Routledge.

Appendix 1: Contingency tables Motivation and Anxiety

"I will need English in my future." (Ideal L2 Self)

	n	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
Grade 8	36	0.0%	0.0%	48.7%	51.3%
Grade 10	39	0.0%	0.0%	63.6%	36.4%

"When I look at my future, I imagine myself speaking English fluently." (Ideal L2 Self)

	n	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
Grade 8	36	2.6%	12.8%	53.8%	30.8%
Grade 10	39	0.0%	9.1%	48.5%	42.4%

"My parents think English is an important language." (Ought-to L2 Self)

	n	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
Grade 8	36	2.6%	23.1%	59.0%	15.4%
Grade 10	39	3.0%	18.2%	48.5%	30.3%

"My friends think English is an important language." (Ought-to L2 Self)

	n	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
Grade 8	36	0.0%	23.1%	59.0%	17.9%
Grade 10	39	3.0%	30.3%	54.5%	12.1%

"I find it scary to speak English during class." (L2 speaking anxiety)

	n	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
Grade 8	36	25.6%	46.2%	23.1%	5.1%
Grade 10	39	15.2%	51.5%	27.3%	6.1%

"I find it scary to speak English outside the classroom." (L2 speaking anxiety)

	n	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
Grade 8	41.0%	46.2%	7.7%	5.1%	51.3%
Grade 10	33.3%	42.4%	21.2%	3.0%	36.4%

Appendix 2: Questionnaire for secondary school pupils (Translation)

(For primary school pupils, sections [C] and [D] were eliminated).

Name: _____ Sex: boy / girl

Date of birth: _____ Class: _____

At home, I usually speak _____(Dutch, French, English, Turkish, Italian, ...)

Answer the questions as honestly as possible!

[A] How often...

1. How often do you watch English-language television programmes, films, cartoon films, documentaries (on TV, tablet, computer, mobile phone...)?	**Never**	**Sometimes**	**Often**	**Always**
Without subtitles				
With Dutch subtitles				
With English subtitles				
With other foreign-language subtitles				
Which programmes/films do prefer/do you watch most?				

2. How often do you play video games in English (at computer, tablet, PlayStation, Xbox, mobile phone, ...)?	**Never**	**Sometimes**	**Often**	**Always**
Without subtitles				
Which video games do you prefer?				

3. How often do you listen to English songs?	**Never**	**Sometimes**	**Often**	**Always**

4. How often do you watch or listen to English-language YouTube clips/videos?	**Never**	**Sometimes**	**Often**	**Always**
Which clips/videos do you prefer/do you watch most?				

5. How often do you speak or write in English with family, friends or online (to chat, on social media, on websites...)?	**Never**	**Sometimes**	**Often**	**Always**

6. How often do you read in English (books, strips, e-books, newspapers, magazines, internet...)?	**Never**	**Sometimes**	**Often**	**Always**
What do you read most?				

Are there some other ways in which you come in contact with English? (e.g. language holiday, internet (Twitch)...)? Yes / No

If yes, in what way?

[B] When you encounter an English word that you do not understand, how often do you do the following ...

	Never	Sometimes	Often	Always
I do nothing if I come across an English word I do not understand				
I try to think of the meaning of the English word myself				
I look up the English word in a dictionary or on the Internet				
I ask someone else to explain the English word				

Is there anything else you do when you hear or see an English word you do not understand? Yes / No

If yes, what do you do?

[C] Please indicate to what extent you agree or disagree with the following statements.

	Strongly disagree	Disagree	Agree	Strongly agree
I need English for my future life				
I could imagine myself speaking English fluently in the future.				
I find it scary to speak English during class.				
I find it scary to speak English outside the classroom.				
My parents think English is an important language.				
My friends think English is an important language.				

[D] Please indicate to what extent the teacher encourages contact with English.

	Never	Sometimes	Often	Always
The English teacher gives us reading assignments				
The English teacher gives us assignments that require us to watch an English movie, series or YouTube video.				
The English teacher uses authentic materials during the lesson (e.g. a Youtube video, music, series or movie, game ... in English)				

Thank you for your participation!

Contents:

Universidade Vigo