Bilingualism: Language and Cognition

cambridge.org/bil

Adolescent learners' L2 English vocabulary knowledge and contact with extramural English: Longitudinal development and relationships between L2 vocabulary and extramural English

Vanessa De Wilde

Department of Translation, Interpreting and Communication, Ghent University, Ghent, Belgium

Research Article

Cite this article: De Wilde, V. (2025). Adolescent learners' L2 English vocabulary knowledge and contact with extramural English: Longitudinal development and relationships between L2 vocabulary and extramural English. *Bilingualism: Language and Cognition* 1–12. https://doi.org/10.1017/S1366728925100540

Received: 20 March 2025
Revised: 1 July 2025
Accepted: 26 August 2025

Keywords:
adolescent learners; vocabulary; extramural English; longitudinal study; network analysis; Complex Dynamic Systems Theory

Corresponding author:
Vanessa De Wilde;
Email: vanessa.dewilde@ugent.be

This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

Abstract

This longitudinal study investigates the development and interrelation of adolescent learners' L2 English vocabulary knowledge and extramural English (EE) input. The study examines the longitudinal development of L2 English receptive vocabulary knowledge, EE input and the dynamics between L2 proficiency and EE input. Data were collected at four time points by administering vocabulary tests and questionnaires on EE activities. Generalized additive mixed models and growth curve models indicated significant vocabulary growth, particularly in the early years of secondary school, which slowed down toward the end of the study. EE activities such as gaming, social media and reading positively predicted vocabulary development, while watching television with L1 subtitles had a negative effect. Temporal network analysis revealed reciprocal relationships, suggesting that L2 proficiency influences EE input and vice versa. The findings underscore the importance of EE in L2 vocabulary development and highlight the dynamic interplay between language learning and extramural activities.

1. Introduction

The importance of meaningful input in language acquisition is uncontested and is often seen as the foundation for second language (L2) learning (Gass, 2018). Usage-based, emergentist perspectives of language learning have claimed that exposure to the L2 drives language learning and is necessary to make progress when learning a second or foreign language (Bybee & Hopper, 2001; Ellis, 2002). As language gains often remain modest when input outside the classroom is limited (Peters et al., 2019), researchers have looked into the affordances of language learning opportunities that take place outside the classroom context (e.g., through extensive reading). This type of learning, in which the main focus is not on language learning but on an activity which might lead to language learning, is often referred to as incidental or contextual language learning (Elgort et al., 2018; Hulstijn, 2012). Whereas incidental language learning points to learning by chance and is often considered the opposite of intentional learning, contextual language learning is an 'intentionality agnostic' term and may refer both to simple intake and to deliberate attempts to discover form and meaning when encountering new words (Elgort et al., 2018).

As English is the lingua franca in large parts of the world (Dewey, 2007), there are ample opportunities for L2 English learners to engage with English outside the classroom context. This engagement with English outside the classroom often happens on the learners' own initiative and is not necessarily stimulated by the language teacher. It is often referred to as extramural English (EE; Sundqvist, 2009). The value of input outside the classroom for L2 learning has been shown in various studies (e.g., Lindgren & Muñoz, 2013).

In the present longitudinal study, I aim to investigate how L2 vocabulary knowledge and EE activities develop over a period of six years. I will further investigate which EE activities impact L2 vocabulary learning at various time points and over time. Finally, I will explore how L2 vocabulary knowledge and EE are interrelated and whether and how L2 proficiency and EE input might influence each other over time.

2. L2 learning and extramural English

Sundqvist (2024) considers extramural English (EE) to be an important individual difference variable that can have a large impact on L2 learning. She argues 'that EE has replaced classroom activities in school as the starting point and foundation for learning English' (Sundqvist, 2024, p. 10). In the past 10–15 years, various studies have investigated the role of EE in language learning. The studies have looked at various types of L2 knowledge and skills, such as vocabulary knowledge (e.g., De Wilde et al., 2022; Peters, 2018; Puimège & Peters, 2019), grammar knowledge (e.g., Muñoz & Cadierno, 2021), receptive skills (e.g., Brevik, 2016; Lindgren & Muñoz, 2013) and

productive skills (e.g., De Wilde et al., 2020). All studies have shown the impact of EE on L2 English learning. In the present study, I will focus on how learners' L2 English receptive vocabulary knowledge is impacted by EE throughout secondary school. Findings of previous studies which explored EE input and vocabulary learning will be discussed below.

2.1. L2 vocabulary knowledge and extramural English

Studies which have investigated EE input learners receive and its impact on L2 English learning have often been done with children and adolescents. One of the first studies to investigate the role of EE in L2 English learning was done by Sundqvist (2009). The author investigated how learners' out-of-school exposure to English impacted 16-year-old Swedish learners' vocabulary knowledge. Vocabulary knowledge was measured with a productive and receptive vocabulary test and EE was measured with a questionnaire and two one-week language diaries. The author found that there was a strong correlation between EE activities and vocabulary knowledge. However, not all types of EE predicted learners' vocabulary knowledge to a similar extent. Productive activities (e.g., gaming) were better predictors than more passive activities such as listening to music or watching television and movies. Other studies done by Sundqvist and colleagues (e.g., Hannibal Jensen, 2017; Sundqvist, 2019; Sylvén & Sundqvist, 2012) which focused on gaming and vocabulary learning confirmed the positive effect of this type of EE on L2 English vocabulary learning.

Puimège and Peters (2019) did a study in Flanders, the Dutch-speaking part of Belgium, which is the context in which the present study took place. The authors looked into EE and vocabulary learning prior to the start of formal classroom instruction. In Flanders, formal L2 English instruction only starts in secondary school. The study explored how much English primary school children had picked up before the start of English classes. The participants belonged to three age groups (10, 11 and 12 years old) representing learners from the three final years of primary school. They all did a meaning recognition and meaning recall test and filled in a questionnaire about their engagement with EE. The authors found that the learners' out-of-school exposure to English increased with age as did their L2 English vocabulary knowledge. The authors further found a significant relationship between gaming and streaming and scores on the meaning recall test. They also found a significant effect of passive exposure (e.g., watching television and listening) and a significant interaction for gaming and streaming and age in the model predicting the score for the meaning recognition test.

Peters (2018) and Peters et al. (2019) investigated vocabulary knowledge and EE engagement with adolescent learners. Peters (2018) investigated differences in the amount of EE between 16-year-old secondary school learners and 19-year-old university learners using a receptive vocabulary test and a questionnaire. The author found that both groups had frequent contact with English outside the classroom context. Secondary school learners gamed more than university students, and university students watched more television without subtitles than the younger group. The relationship between EE and vocabulary knowledge was inspected through correlations. It was found that watching television without subtitles, reading books and magazines and using the Internet were positively related to learners' L2 English vocabulary knowledge. Contrary to what was found in studies with younger learners, there was no effect of gaming in this study. The study by Peters et al. (2019) investigated language learning and out-of-school exposure in two languages, English and French, and three age groups, participants from the second and fourth year of secondary school and participants in the first year of university. Again, the participants did a receptive vocabulary test and filled in a questionnaire. Structural equation modelling was used to explore which variables predicted learners' vocabulary knowledge. The results for English showed that vocabulary knowledge increased with age and that participants who engaged more with online activities had a higher vocabulary knowledge.

2.2. Interrelatedness of L2 vocabulary knowledge and extramural English

The previous studies have consistently shown that receptive and productive L2 vocabulary knowledge is partly predicted by EE input. Only little is known, however, about the longitudinal effects of EE on language learning, about possible reciprocal relationships between vocabulary knowledge and EE (as was suggested by Puimège & Peters, 2019) and about how the effectiveness of certain activities differs with age (Peters, 2022). Even though various studies have pointed out the added value of longitudinal studies (e.g., Busby, 2024; Kusyk et al., 2025; Peters, 2022), not many longitudinal studies have been done which investigated the role of EE on L2 learning. In one such study, Verspoor et al. (2011) investigated how out-of-school exposure impacted vocabulary knowledge (and writing) during a semi-longitudinal study. Two separate groups of Dutch adolescent learners, who were in the first and third year of secondary school, respectively, all completed a questionnaire about their contact with English at the start of the school year and did three receptive vocabulary tests in one year. The authors found that learners who hardly had any media input next to classroom input developed more slowly. The study results also suggested that input is not a static variable in the sense that it remains stable over time, but that it interacts with proficiency. EE input had a larger effect for learners with a higher proficiency. A study by Busby (2024) tried to take into account university students' previous EE experiences by doing a survey in which she asked about learners' past experiences with EE and to identify the earliest EE activity which they felt contributed to their L2 English learning. This measure significantly predicted learners' vocabulary size.

The findings from longitudinal studies or studies which include a temporal perspective suggest that not only current engagement with EE has an impact on L2 proficiency but that the input learners received in the past has a lasting, or at least a long-term effect. Furthermore, findings from Verspoor et al. (2011) suggested a differential effect of EE dependent on learners' proficiency.

2.3. Related studies

Two previous studies, De Wilde et al. (2020) and (2021), were done with the participants who also took part in the present study. De Wilde et al. (2020) tested how EE impacted learners' receptive vocabulary, listening, reading, writing and speaking skills when learners were in the final year of primary school, before the start of formal classroom instruction. In this study, receptive vocabulary knowledge was measured with a picture-based meaning recognition test, and EE was measured with a questionnaire. The authors found strong correlations between the various language measures (Pearson's r between .68 and .77). A principal components analysis showed that all English tests were captured in one component which the authors named 'overall proficiency'. The authors built regression models looking into the role of EE for all language tests

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

and for 'overall proficiency'. The regression models showed that EE explained 16 to 23% of the variance in the various L2 measures. The authors found that gaming, using social media and speaking positively predicted receptive vocabulary knowledge whereas listening to English music had a negative effect on learners' vocabulary scores. These findings again showed that activities which entailed an element of production were beneficial for language learning. De Wilde et al. (2020) linked these findings to the production effect in memory research (Macleod et al., 2010) which entails that active use of a language is more effective for learning than passive perception.

De Wilde et al. (2021) further explored the role of prior L2 English vocabulary knowledge (measured two years earlier) and various types of EE, together with other individual difference variables (e.g., analytic reasoning ability), on L2 receptive vocabulary knowledge and speaking skills. The authors tested L2 learners at two time points with an interval of two years. They found that learners' receptive vocabulary knowledge had strongly increased at time 2. The use of social media and gaming was positively related to vocabulary knowledge at time 2. However, once prior vocabulary knowledge measured at time 1 was added to the model, the effect of gaming became much smaller and the effect of using social media was no longer significant. As the authors had demonstrated in their previous study (De Wilde et al., 2020), this prior L2 knowledge resulted from out-of-school learning.

3. A process-oriented approach

If we want to get a better grasp of how EE and proficiency develop and how they are interrelated, a process-oriented approach is warranted. A theory of second language development which has foregrounded the learning process is complex dynamic systems theory (CDST). This theory has an eye for individual development over time and stresses the interrelatedness of both language-related (e.g., vocabulary knowledge) and individual difference (e.g., EE) variables (De Bot et al., 2007; Larsen-Freeman, 1997).

When focusing on the learning trajectory and modeling development, CDST-researchers have often used generalized additive mixed models (GAMMs) as these models can take into account possible non-linear development and differences between individuals (Pfenninger, 2020; Winter & Wieling, 2016). To model relations between various individual difference variables, Freeborn et al. (2023) proposed psychological network analysis. The authors highlight some characteristics of the method which make it suitable for CDST-inspired research such as how different existing (rather than latent) variables mutually influence each other within a system and the possibility to explore patterns of multi-causality and interconnectedness in a data-driven manner (cf. also Van Dijk et al., 2024). Freeborn et al. (2023) explored this method by analyzing two existing datasets and showed that psychological network analysis can be used to explore and better understand relationships between variables. Psychological network analysis explores patterns and structures of relationships in multivariate data (Borsboom et al., 2021). In a visual representation of a network, variables are represented as circles (called *nodes*) and associations between variables are represented by lines (called *edges*). Edges are typically undirected which reflects that variables are intertwined, and relationships are multicausal. Network models can also be used with longitudinal and panel data (Borsboom et al., 2021; Epskamp et al., 2018). A temporal network is a directed network that shows changes in variables within individuals over time. It shows how variables predict one another at the next time point. In the present study, a temporal network could give information about how L2 vocabulary knowledge and EE activities are related over time. It could thus be used to uncover longitudinal effects and possible reciprocal relationships between variables.

4. The present study

In the present study, I aim to contribute to research on the role of EE in language learning by investigating the development of L2 vocabulary learning and EE input over time and by looking into the relationships between L2 vocabulary and different types of EE both cross-sectionally and longitudinally.

The study builds on and uses data from two previously published studies (De Wilde et al., 2020, 2021). In De Wilde et al. (2020), I reported on a cross-sectional study (Time 1) in which I investigated how much English can be learned before the start of formal instruction and which type of extramural English activities contributed to language learning. I calculated descriptive statistics and correlations and did multivariate regression analysis. In De Wilde et al. (2021), I further investigated the role of out-of-school exposure and other individual differences on English development. In this study, data from two time points (Time 1 and Time 2) were used to be able to account for learners' prior knowledge in the analyses. The research questions were again answered by looking at descriptive statistics, correlations and running multivariate regression analyses.

The present study reports on a panel study with four data collection points, each with a two-year gap. The data from time points 1 and 2 have been used in the two studies mentioned above. Data from time points three and four have not been used before. Data was collected at times 3 and 4 to be able to get an idea of adolescent learners' vocabulary development over a longer period (six years, the entire period in which Flemish adolescents receive compulsory English instruction in school). In the 2021 study, I saw that a lot of learning took place after the introduction of classroom instruction, but it remained unclear how the learning trajectory would further evolve. One of the goals in the present study was to investigate this learning trajectory, both for the group but also for the individual learner. I also wanted to look into how learners might engage with extramural English in different manners over time, hence the network approach in the present study. Finally, the study focuses on the dynamics between L2 vocabulary knowledge and extramural English activities and the possible reciprocity of the relationship, a topic which has been touched upon in previous studies, but which has not yet been investigated in depth. The study is guided by the following research questions:

RQ1: How does adolescent learners' L2 English receptive vocabulary knowledge develop over time?

RQ2: How does adolescent learners' exposure to various types of EE develop over time?

RQ3: How do L2 English proficiency and L2 exposure to English interrelate at each separate time point and over time?

Based on previous studies which have looked into L2 vocabulary knowledge of young and adolescent learners in a context in which there is easy access to English (e.g., Peters, 2018; Peters et al., 2019), I hypothesize that learners' vocabulary knowledge will increase throughout the study.

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

The second research question is concerned with EE input over time. Puimège and Peters (2019) found that engagement with EE increased throughout the three final years of primary school. In an earlier study with the participants from this study (De Wilde et al., 2021), I found that there was an increase in the amount of out-of-school exposure to English when children transitioned from primary to secondary school. Peters' study with adolescent learners (2018) showed that exposure to English was not necessarily higher when 16-year-olds were compared to 19-year-olds but the types of activities they engaged with were different. Younger learners spent more time playing games while older learners spent more time watching television and movies without subtitles. I thus hypothesize that learners will engage differently with EE at times points 3 and 4 compared to time points 1 and 2, both with regards to the amount of exposure to EE and the type of EE activity.

With the final research question, I aim to explore the dynamics between L2 vocabulary learning and EE input both at various time points and longitudinally. Based on earlier studies that have investigated the role of EE input in L2 vocabulary learning (e.g., Sundqvist, 2009), I hypothesize there will be a link between various types of EE and L2 vocabulary learning and links among various types of EE as was demonstrated by the principal component analysis in Puimège and Peters (2019). I further assume a stronger relationship between L2 learning and EE activities that entail an element of production (De Wilde et al., 2020; Sundqvist, 2009). Whether EE and L2 vocabulary knowledge are related in similar ways at various time points remains to be seen. Various studies have called for more longitudinal research to investigate the role of EE over time (e.g., Peters, 2018). Other researchers have hinted at possible reciprocal relationships between L2 learning and EE (De Wilde et al., 2021; Puimège & Peters, 2019). To the best of our knowledge these dynamics have not been investigated before. Based on results of earlier, cross-sectional studies, I hypothesize that reciprocal relationships, i.e., EE input impacts L2 knowledge and L2 knowledge impacts EE input, will be present.

5. Method

5.1. Context

The study took place in Flanders, the northern, Dutch-speaking part of Belgium. In Belgium there are three official languages. Dutch, which is spoken in Flanders; French, which is spoken in Wallonia and German, which is spoken in a small area that borders on Germany. The capital region Brussels is bilingual (Dutch–French). Because of the presence of multiple official languages, the Flemish government has decreed that French is the first foreign language to be taught in Flanders. This means that the only foreign language taught in most Flemish primary schools is French. English classes typically start in secondary school when learners are 12 to 13 years old. This is rather late compared to other European countries (De Wilde et al., 2020). In secondary school most learners receive two to three 50-minute lessons per week. The learners who are in an academic track are expected to have a CEFR B1$^+$ level by the end of secondary school, after six years. Expectations for learners in technical and vocational tracks are lower. Students in these tracks are expected to be proficient at the CEFR A2 to B1 levels, depending on the specific track (e.g., with or without a focus on languages).

5.2. Participants

The first data collection took place from October 2016 until January 2017, when the participants were in the final year of primary school. 780 children from 38 primary schools participated in the first data collection (402 boys and 378 girls). The schools were selected through a stratified random sampling method, ensuring geographical diversity and diversity of different school types. The Flemish educational system is organized through three networks: state schools, subsidized public schools and subsidized free schools. All three networks are represented in the sample.

The participants were 10–12 years old at the start of the data collection. The second wave of data was collected two years later (April–June 2019) when the learners were at the end of the second year of secondary school. All participants who had given consent to be contacted for a follow-up study received an e-mail and were asked if they were willing to participate. The second data collection concerned 114 learners (59 boys and 55 girls) who were then in 55 different secondary schools. A third wave of data was collected when the learners were in the fourth year of secondary school (April–May 2021). Due to measures related to the COVID-pandemic, the data collection was done in an online session via Microsoft Teams. The participants and the researcher were both online in the Teams session during the entire data collection and participants could ask questions or signal problems throughout the session. All learners were used to online teaching and use of technology as this was common practice for them during the COVID-pandemic. 81 students agreed to participate in the third wave. Five participants were 14 years old, 46 participants were 15 years old, 29 participants were 16 years old and one participant was 17 years old at the time of the third data collection. The sample consisted of 38 boys and 43 girls. The final data collection was done in April and May 2023 after six years of secondary school. In this wave 63 learners participated. Four participants were 16 years old, 42 participants were 17 years old and 17 participants were 18 years old at the time of the fourth data collection. The sample consisted of 32 boys and 31 girls. Most participants joined an online session; seven participants did the session in school in the presence of a researcher. To answer the research questions that include time as a variable, I included data from 81 learners who participated at least three times. 48 learners participated in all four waves; 33 learners participated three times. All learners participated at time 1 (n = 81); 73, 77 and 60 learners participated at times 2, 3 and 4, respectively. An overview of the participants at each time point can be found in Table 1. The study was approved by the ethics committee of the university. Both the participants and their parents gave consent to take part in the study for each wave of the data collection.

5.3. Instruments

Receptive vocabulary knowledge was measured with the Peabody Picture Vocabulary Test 4 (PPVT 4; Dunn & Dunn, 2007). In this test, participants are presented with a spoken word and four pictures. Participants had to select one of the pictures which corresponded to the meaning of the word. The tests were administered in the classroom (at times 1 and 2) or via the online tool Limesurvey (at times 3 and 4). All items were recorded by a research

Table 1. Number of participants for each wave of the data collection

	T1 (2016–2017)	T2 (2019)	T3 (2021)	T4 (2023)
Full sample	n = 780	n = 114	n = 81	n = 63
Longitudinal sample	n = 81	n = 73	n = 77	n = 60

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

assistant who was an advanced L2 English speaker, and the recordings were presented to the participants. They were allowed to listen to the recording multiple times if necessary. At time 1, the participants were presented with the first 120 items; at time 2, the first 144 items were tested. As all learners knew the words from the first three sets of 12 items at time 2 and I had to keep an eye on the timing and the length of the vocabulary test, these 36 words were no longer tested at times 3 and 4. At time 3, more challenging sets were added, and learners were again presented with 144 items (starting from items 37–item 180). To avoid a ceiling effect, one extra set of 12 items was added at time 4. Learners were presented with 156 items (item 37–item 192). The items were scored dichotomously receiving either a score of 0, for a wrong answer, or 1, for a correct answer.

Out-of-school exposure to English was measured with a questionnaire (cf. De Wilde et al., 2020). In the questionnaire, participants reported how often they engaged with various types of exposure per day: watching English television (with subtitles in the L1, with English subtitles, without subtitles), listening to English music, reading in English, using social media in English, gaming in English and speaking English outside the school context. Learners indicated the amount spent on each of these types of exposure. Possible answers were: 'I do not do this', '0–30 minutes', '30 minutes–1 hour', '1 hour–1 hour and 30 minutes', '1 hour and 30 minutes–2 hours' and 'more than 2 hours'. The answers were transformed to a numerical score (from 0 to 5) for further analysis.

5.4. Analysis

To answer research questions 1 and 2, I first calculated descriptive statistics and plotted the scores for the vocabulary test and the different types of EE both for the group and the individual learners. Plots were made with the ggplot2-package (Wickham, 2011) in R version 4.4.1. (R Core Team, 2024). I then modeled whether and how time predicted changes in the variables of interest. As most plots showed a non-linear trend, I used generalized additive mixed models (GAMM) to also be able to model non-linear trajectories using the mgcv-package (Wood & Wood, 2015). To answer the third research question and explore relationships between the variables, I used network analysis. First four cross-sectional networks, one for each time point, were modeled using the bootnet (Epskamp & Fried, 2015) and qgraph-packages (Epskamp et al., 2012) in R. The stability of the networks was estimated using bootstrapping, and the four networks were compared using the network comparison test, which I ran with the NetworkComparisonTest R-package (van Borkulo et al., 2015). I then modelled GAMMs to look into the relationship between vocabulary knowledge and EE over time. To do this, I added time and the EE activities to the model. As the GAMM showed either linear or quadratic relationships, I opted for a simpler growth curve model (GCM) as this model is easier to interpret, especially when there are multiple fixed effects (Winter & Wieling, 2016). GAMMs were modeled using the mgcv-package and GCMs with the lmer-package (Kuznetsova et al., 2017) in R. Finally, a temporal network model was estimated using the psychonetrics-package (Epskamp et al., 2020). One of the assumptions for temporal network analysis is stationarity (i.e., data characteristics remain constant over time). As expected, the variable which measured L2 vocabulary knowledge was not stationary. I ran the model with the original data and with the detrended data but the model with the original data proved more informative as the detrended model resulted in a much sparser network with only two autoregressive effects and no other associations between nodes. The model with the detrended variable can be found in the Supplementary Figure S4. All data and code are available on https://osf.io/7ns4k/.

6. Results

6.1. L2 English vocabulary development

Descriptive statistics for the vocabulary tests at the different time points can be found in Table 2. I report the scores for the 81 participants who participated in the six-year longitudinal study. Descriptive statistics showed that there is a large spread in the results of the receptive vocabulary test. A boxplot showing the results for the vocabulary test at each time point and a plot showing the mean development of the vocabulary score throughout the study are shown in Figure 1. Supplementary Figure S1 shows individual learners' vocabulary development. Results showed an increase in learners' L2 receptive vocabulary knowledge throughout the study but also showed that large differences between the learners remain over time. As I wanted to take non-linearity into account in the analysis of how time predicted learners vocabulary knowledge, I modeled a GAMM in R according to the following model specifications:

$$\begin{aligned}\text{model} <- \text{bam}(&\text{ppvt} \sim \text{s}(\text{Time}, \text{k} = 4) \\ &+ \text{s}(\text{ID}, \text{k} = 10, \text{bs} = 'fs', \text{m} = 1), \text{data} = \text{mydata}).\end{aligned}$$

Smooths are linear when the edf-value (effective degrees of freedom) is 1 and non-linear when the edf-value is higher than 1. I added a random effect which allowed individual variability. The effect of time was non-linear (edf = 2.88) and highly significant ($p < .001$) indicating that time impacted learners' vocabulary knowledge. Figure 1 suggests that L2 vocabulary growth slowed down toward time 4. The individual trajectories in Supplementary Figure S1 showed that for some learners the growth was linear, whereas for other learners the growth slowed down. There was an upward trajectory for all learners.

6.2. EE input over time

Descriptive statistics for the various EE activities at the different time points can be found in the Supplementary Table S1. Figure 2 shows the boxplots with the learners' EE input at the various time points. GAMMs were modeled to investigate how time predicted EE input according to the following model specifications:

$$\begin{aligned}\text{model} <- \text{bam}(&\text{EE} \sim \text{s}(\text{Time}, \text{k} = 4) + \text{s}(\text{ID}, \text{k} = 10, \text{bs} = 'fs', \text{m} = 1), \\ &\text{data} = \text{mydata}).\end{aligned}$$

Table 2. Descriptive statistics receptive vocabulary tests (PPVT = Peabody Picture Vocabulary Test)

		Min	Max	Median	Mean	SD
1.	PPVT T1 (max = 120, *n* = 81)	36	118	83	82.84	17.62
2.	PPVT T2 (max = 144, *n* = 73)	71	141	117	116.8	15.23
3.	PPVT T3 (max = 180, *n* = 77)	101	174	148	146	18.35
4.	PPVT T4 (max = 192, *n* = 60)	115	188	163	161.6	18.02

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

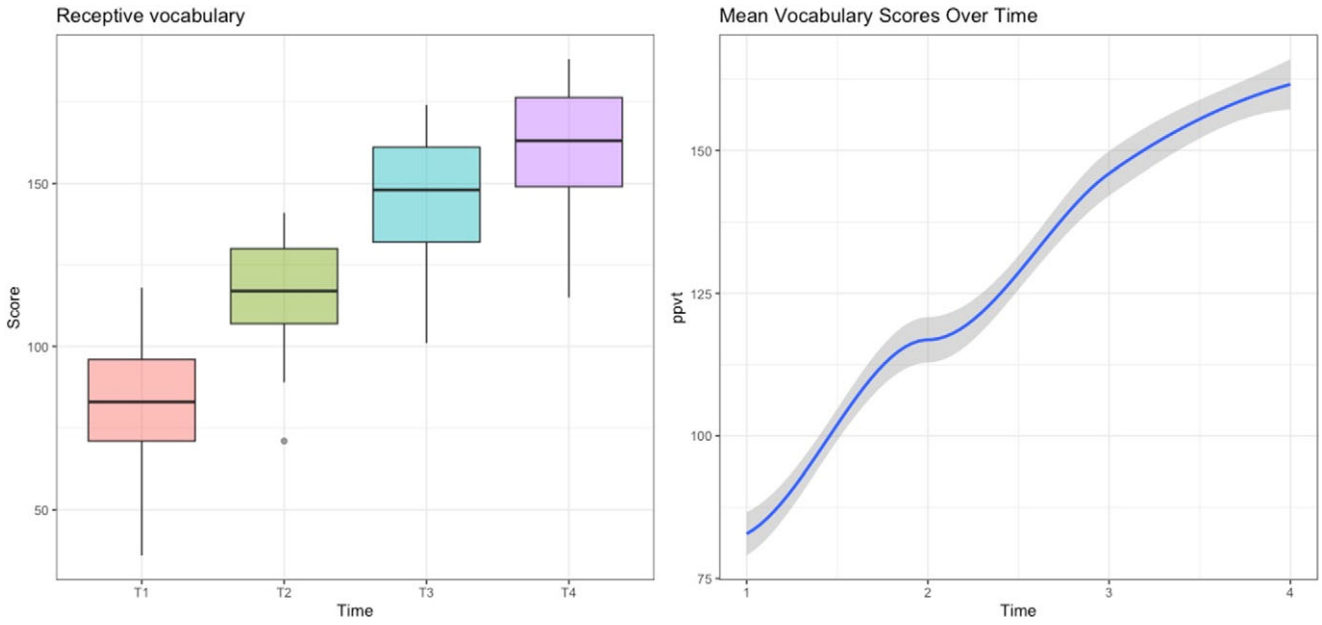

Figure 1. Spread of the vocabulary test scores at each time point and mean vocabulary scores over time.

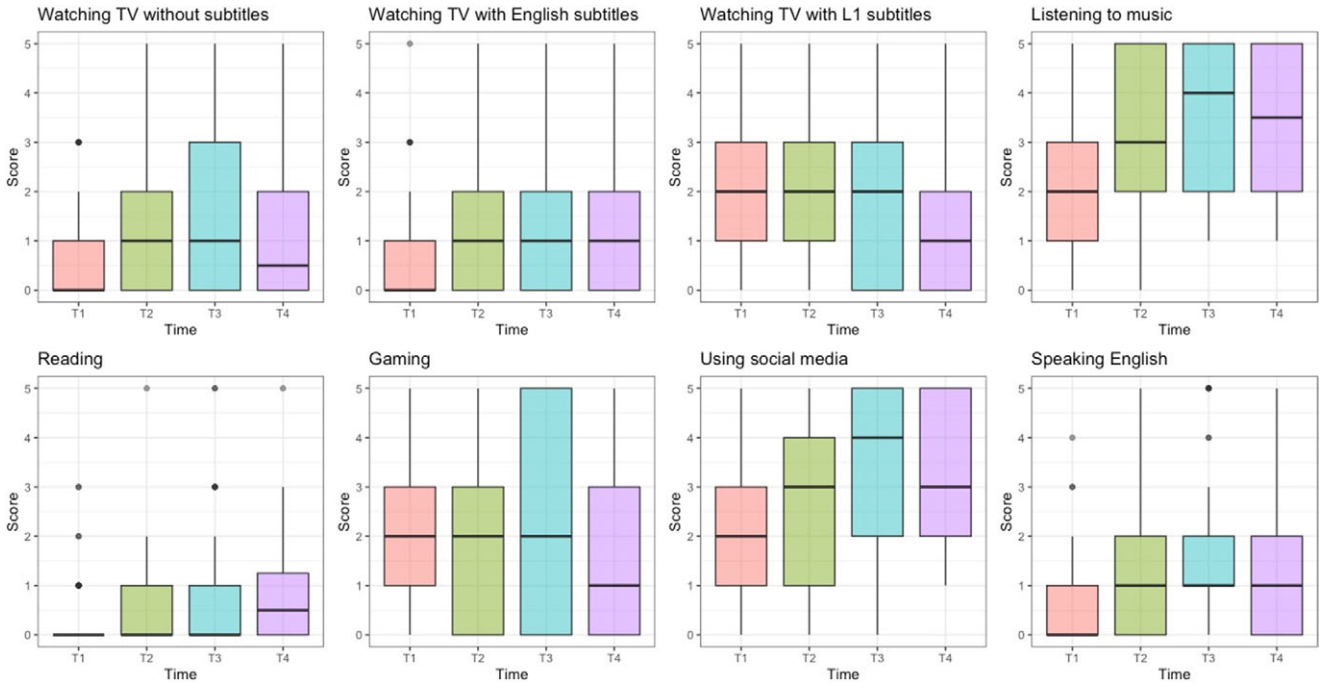

Figure 2. Boxplots showing the spread of EE engagement at each time point.

The effect of time was highly significant for all types of EE activities and non-linear for most activities: watching television without subtitles, edf = 2.49; watching television with English subtitles, edf = 2.20; watching television with L1 subtitles, edf = 2.12; listening to music, edf = 2.39; gaming, edf = 2.70; using social media, edf = 2.45; speaking in English, edf = 2.25. The only exception is reading in English which was not done frequently but significantly increased over time ($p < .001$) in a linear manner (edf = 1). A visualization of the mean scores over time is shown in Figure 3. The plots show that all types of EE activities increased between times 1 and 2. Between times 3 and 4, most activities were done less frequently, except for reading in English and watching television with English subtitles. The decrease was most pronounced for gaming and watching television with L1 subtitles. Supplementary Figure S2 shows individual trajectories for all types of EE activities. EE input changed across different time points for many learners, sometimes remaining stable, sometimes increasing or decreasing. The overall trend I observed for the group thus did not correspond with findings for individual learners.

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

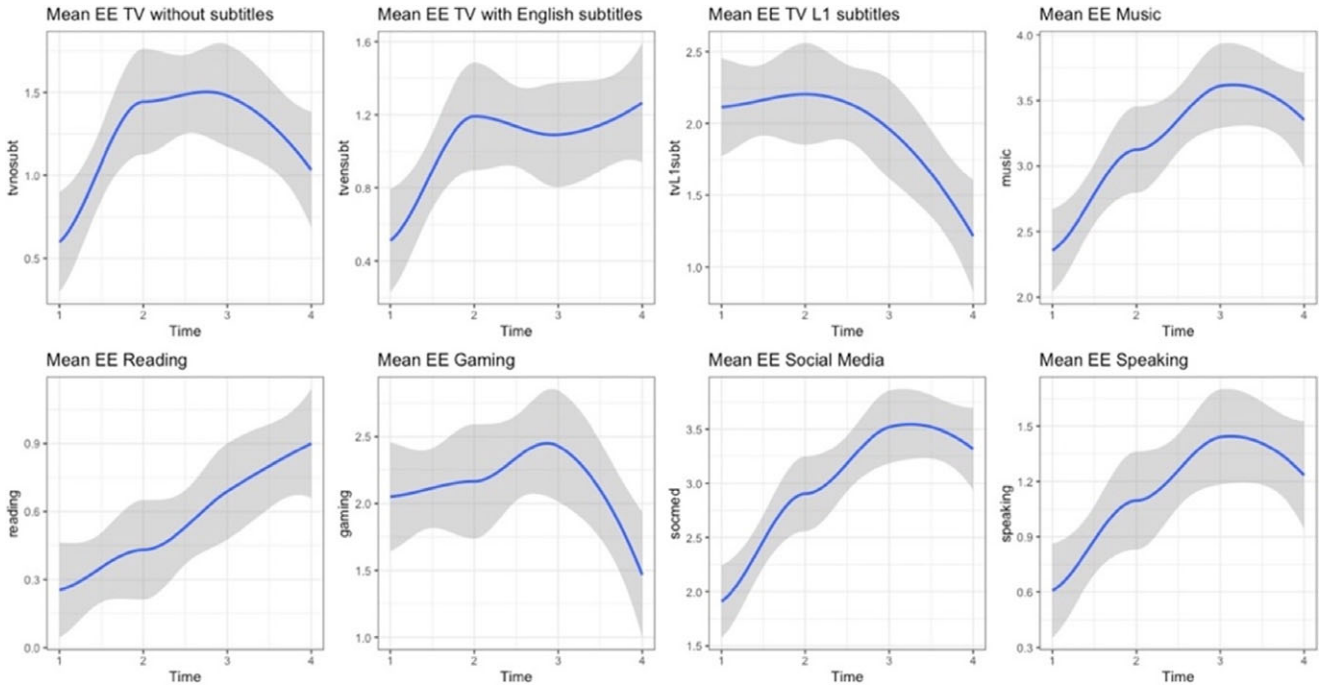

Figure 3. Mean engagement with various types of EE over time.

6.3. Interrelatedness between L2 English proficiency and L2 exposure to English

To explore the relationships between various types of EE and L2 vocabulary at each separate time point, I first modeled four cross-sectional networks. The networks were modeled using all the data that were available from that time point. A visual representation of the models can be found in Figure 4. Non-parametric bootstrapping was used to assess model stability (cf. Supplementary Figure S3). The analysis showed that the first model was more stable than the three other models which were not very stable and should be interpreted with care. This is unsurprising as the sample size of the first data collection was far larger than at the other time points. Finally, I used network comparison tests to assess whether the networks differed at the various time points. This was done with the NCT-function from the NetworkComparisinTest-package in R (van Borkulo et al., 2015). The network invariance test computes whether the overall structure between networks is different. The global strength invariance tests compute whether the total connectivity of the networks is similar. Finally, the edge invariance test calculates whether edges between two nodes in two networks are significantly different. Pairwise comparisons between networks showed that there were no significant differences in terms of overall network structure or global strength between the four networks. There were a few significant differences for individual edges. There was a significant difference in the edge between watching television with English subtitles and listening to music (p = .01) for the networks at times 1 and 2. When comparing networks at times 1 and 3, a significant difference was found for the edges between receptive vocabulary and watching television with L1 subtitles (p = .02), watching television with English subtitles and gaming (p = .04) and using social media and watching tv with L1 subtitles (p = .03). The comparison between networks at times 1 and 4 showed differences between the edges for reading and gaming (p = .01) and speaking and gaming (p = .04). When comparing networks at times 2 and 3, edges were different for watching television with English subtitles and listening to music (p = .04) and watching television with L1 subtitles and using social media (p = .02). The edges in the networks at times 2 and 4 only showed significant differences for the edges between watching television with English subtitles and listening to music (p = .04) and reading and gaming (p = .04). Finally, for the networks at times 3 and 4, a significant difference was found for the edge between reading and gaming (p = .01).

To investigate the relationship between L2 English receptive vocabulary and EE input over time, I first modelled a GAMM. Results of the best model can be found in the Supplementary Table S2. The results of the full model showed that the effect of time was non-linear (edf = 2.82) and the effect of various types of EE on L2 vocabulary knowledge was linear. Based on the results of the GAMM and the trend in the vocabulary scores I observed in Figure 1, I then modeled a growth curve model (GCM) as this model is easier to interpret than a GAMM. I added a random intercept for ID and a random slope for time. In the fixed effects, a quadratic term for time was added next to linear terms for time and the various types of EE. The GCM model with the best fit can be found in Table 3. Time was the strongest predictor of vocabulary knowledge. The quadratic term showed that the growth in vocabulary knowledge slowed down toward the end of the study. Next to the effect of time, four types of EE activities significantly predicted L2 vocabulary knowledge: watching television with subtitles in the L1, reading, gaming and using social media in English. One variable, watching television with subtitles in the L1, had a negative effect on vocabulary scores; the other variables had a positive effect. The conditional R^2 was .96, suggesting that both the fixed and random effects together accounted for 96% of the variance. The model had a marginal R^2 of .79, indicating that the fixed effects alone explained 79% of the variability in the scores.

Because the reciprocity of variables is also of interest in this study, I then modeled a temporal network using the psychonetrics package in R. As the dataset is rather small and I wanted the network to be as stable as possible, I decided not to add all the types of EE. All variables

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

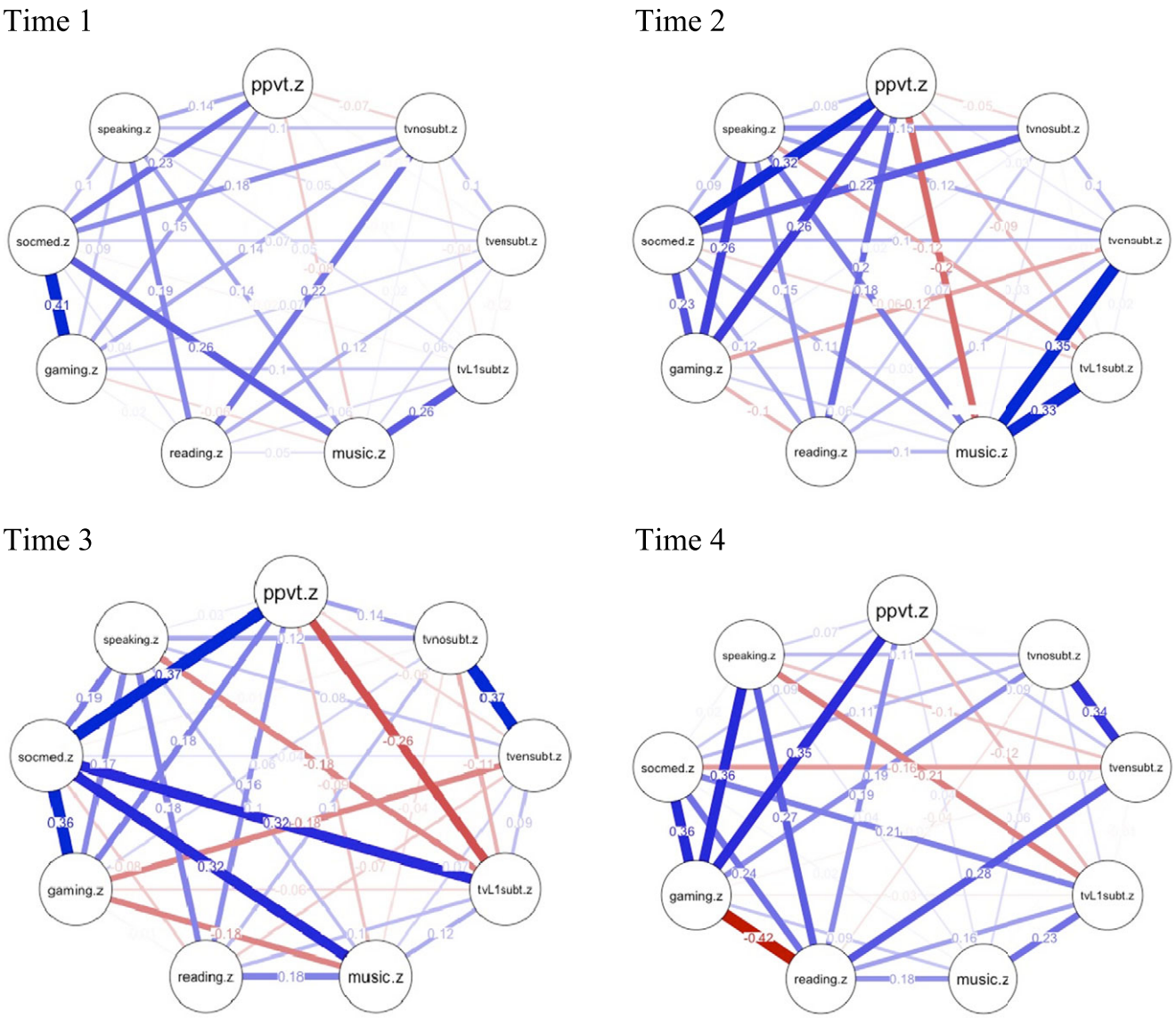

Figure 4. Cross-sectional networks for each time point.
(ppvt.z = receptive vocabulary knowledge, tvnosubt.z = watching English television without subtitles, tvensubt.z = watching television with English subtitles, tvL1subt.z = watching television with L1 subtitles, music.z = listening to English music, reading.z = reading in English, gaming.z = gaming in English, socmed.z = use of social media, speaking.z = speaking English).

that were present in the best GCM were included in the temporal network. The network model assumes stationarity (i.e., data characteristics remain constant over time) which can be problematic when studying L2 development. Detrending can address this, but it may also remove key developmental trends. In Figure 5, the network is shown without detrending; the network with detrended variables can be found in Supplementary Figure S4. The temporal network in Figure 5 showed autoregressive effects for all variables except watching television with L1 subtitles, which means that for these variables the value at t – 1 was predictive for the value of that same variable at the next time point. Certain types of EE (watching television with L1 subtitles, using social media and gaming) positively predicted vocabulary scores at the next time point. The model also showed that learners with higher vocabulary scores tended to spend less time gaming and watching television with L1 subtitles at the next time point. Finally, more frequent reading at t – 1 negatively impacted vocabulary scores at the next time point. However, there is very little variability in reading scores and little time spent on reading (cf. Figure 3), so this finding should be interpreted with care.

7. Discussion

First, I investigated the development of learners' L2 English receptive vocabulary knowledge throughout secondary school. The results were in line with the hypothesis and showed an increase in vocabulary knowledge throughout secondary school. Similar results were found in cross-sectional studies with various age groups (Peters, 2018; Peters et al., 2019; Puimège & Peters, 2019) and longitudinal studies (Webb & Chang, 2012). A previous longitudinal study with these participants showed there was a strong increase in learners' vocabulary knowledge in the first years of secondary school, which coincided with the start of formal L2 English instruction and suggested complementary effects of extramural English and schooling (De Wilde et al., 2021). This complementarity was also investigated in a qualitative study by Rød and Calafato (2025) who found that EE exposure was beneficial for general vocabulary knowledge whereas classroom instruction benefited the development of academic and specialized vocabulary knowledge. The present study further showed that the growth in

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

Table 3. GCM predicting scores on the vocabulary test

Random effects			Variance	*SD*
ID	(Intercept)		176.78	13.296
	Time		13.59	3.687
Residual			41.89	6.472
Fixed effects	**Estimate**	***SE***	**t-value**	***p***
(Intercept)	41.05	2.90	14.14	<.001***
Time	47.76	2.25	21.22	<.001***
Time^2	−4.52	0.44	−10.32	<.001***
TV L1 subtitles	−1.37	0.42	−3.28	.001**
Reading	1.51	0.68	2.24	.027*
Gaming	1.20	0.46	2.63	.009**
Using social media	1.45	0.46	3.17	.002**

*p < .05.
**p < .01.
***p < .001.

study. Some learners' vocabulary knowledge increased linearly, whereas others exhibited non-linear growth. This is in line with the premises of CDST, which states that the learning process is highly individual and dependent on many different factors (De Bot et al., 2007).

Next, I looked into learners' EE input over time. The results showed an increase in EE input in the first years of the study. This is in line with findings from other studies with younger learners (e.g., Puimège & Peters, 2019). In the final years of the study, engagement with some types of EE such as watching television with English subtitles, listening to English music, using social media and speaking English outside the classroom seemed to stabilize. Two types of activities were done less frequently, watching television without subtitles and gaming, a finding which is in line with Peters (2018). One type of EE increased throughout the study, reading English. However, reading in English was the least frequent activity throughout the entire study. This finding is also in line with what was found in earlier studies (e.g., Peters, 2018). The present study showed that findings concerning EE input that were found in cross-sectional studies were also present in a longitudinal study and that EE habits changed over time within learners. Even though certain

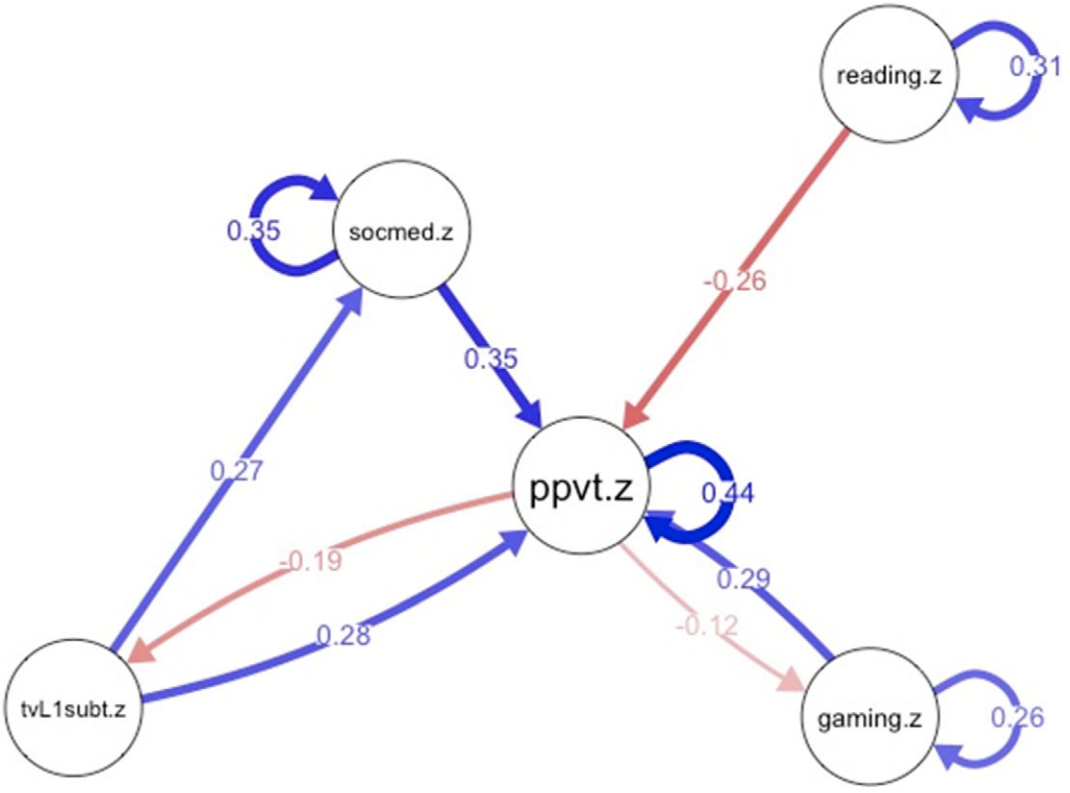

Figure 5. Temporal network showing the relationships between various types of EE and L2 vocabulary knowledge over time.
(ppvt.z = receptive vocabulary knowledge, tvL1subt.z = watching television with L1 subtitles, reading.z = reading in English, gaming.z = gaming in English, socmed.z = use of social media).

learners' vocabulary knowledge slowed down in the final two years of secondary school. When I considered the individual learning trajectories, I observed an increase in L2 vocabulary knowledge for all learners. At the same time, the shape of the learning trajectory was different across learners. Some learners showed a steeper vocabulary growth than others. This could be due to the fact that there were large individual differences between the learners throughout the study and possibly more added value of classroom instruction for learners with lower proficiency throughout the

trends could be observed over time to describe learners' input through various types of EE, the overall trend did not reflect individual learners' EE activities. Individual learners' EE activities were sometimes stable across time points but sometimes also showed large differences and changes from one time point to the next (cf. Supplementary Figure S2). This finding is also in line with CDST-studies which have posited that individual difference variables can change over time (Dörnyei & Ryan, 2015) and have further shown that findings for the group do not necessarily reflect

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

findings for individual learners, i.e., the ergodicity problem (De Wilde & Lowie, 2024; Lowie & Verspoor, 2019).

Finally, I investigated how L2 vocabulary development and engagement with EE are interrelated. I modeled cross-sectional networks between L2 vocabulary knowledge for data which was available at each time point. The results showed that, even though networks at times 2, 3 and 4 should be interpreted with care, there were many links between EE variables and L2 vocabulary knowledge and between various types of EE. The global strength and connectivity were similar at all time points. Most edges also stayed similar across networks modeled at different time points. Even though the results showed that vocabulary growth slowed down and engagement with various types of EE changed over time, the patterns and relationships among various types of EE and vocabulary and EE remained largely the same in the cross-sectional networks. There was a positive relationship between L2 receptive vocabulary knowledge, gaming, using social media and, to a lesser extent, speaking at each time point. These confirmed findings from previous studies which showed that types of EE which contain an element of production had a positive impact on L2 knowledge (e.g., Sundqvist, 2009). I found no or only a weak relationship between watching television with English or no subtitles and L2 vocabulary knowledge, but these types of EE were quite strongly related, indicating that participants who watched English television without subtitles also tended to watch television with English subtitles. There was no significant relationship between reading and L2 vocabulary knowledge at time one, but a positive relationship emerged at time 2 and was found at all other time points. This is in line with other work investigating the impact of L2 reading on vocabulary knowledge (Pellicer-Sanchez & Schmitt, 2010; Webb, 2007). It should be noted, however, that reading was not done frequently compared to the other EE activities. Two other types of EE showed different patterns at different time points. For listening to music there was a negative relationship with L2 knowledge at the first three time points but a positive relationship at the fourth time point. For watching television with L1 subtitles, I found no significant relationship with L2 vocabulary knowledge at time 1 but a negative relationship at all other time points. Different findings for music and watching L1 television were also observed in other studies (Lindgren & Muñoz, 2013; Puimège & Peters, 2019).

After exploring cross-sectional relationships between L2 vocabulary knowledge and EE at the different time points, I investigated the development of vocabulary knowledge and the impact of EE on L2 vocabulary knowledge over time. The results showed large learning gains over time, with a decrease in the speed of vocabulary growth toward the end of the study. Three types of EE were positive predictors of vocabulary scores: reading, gaming and social media. All three types expect active engagement from the learner and have been shown to impact L2 (vocabulary) learning (e.g., Puimège & Peters, 2019; Sundqvist, 2009; Sylvén & Sundqvist, 2012). Watching television with L1 subtitles affected vocabulary scores in a negative manner. De Wilde et al. (2021) explained a similar finding by indicating that L2 input that is supported by the L1 might be preferred by learners with a lower L2 English proficiency, rather than that contextual language learning does not happen when watching television with L1 subtitles. To further explore reciprocal relationships and changes in the relationships between variables over time, I then modeled a temporal network with the variables that were present in the best growth curve model. This network thus consisted of five variables: L2 vocabulary knowledge, gaming, using social media, reading and watching television with L1 subtitles. The network showed that for all variables except for watching television with L1 subtitles, the score at the previous time point was highly predictive of the score at the next time point. This means that adolescents who gamed, used social media and read frequently tended to continue to do this throughout secondary school. For reading this needs to be interpreted with care as overall this activity was rather uncommon (cf. Supplementary Figure S2). This effect was also clearly present for L2 vocabulary knowledge. Learners with a high vocabulary knowledge at one time point also had a higher vocabulary score at the next time point. This is in line with findings from other studies which have considered the impact of prior L2 knowledge (e.g., De Wilde et al., 2021; Jaekel et al., 2017). The model also clearly showed reciprocal relationships between L2 vocabulary knowledge and EE activities. For gaming and watching television with L1 subtitles, the results showed that engagement with EE positively impacted L2 vocabulary knowledge at the next time point, but learners with higher L2 vocabulary scores at one time point tended to watch television with L1 subtitles and game less frequently at the next time point. These findings confirm hypotheses that were formulated in earlier studies (e.g., De Wilde et al., 2021; Puimège & Peters, 2019) but that had not yet been investigated.

As mentioned earlier, the dataset is relatively small to model networks. The cross-sectional analyses showed that the network at time 1, with a substantially larger dataset, was more stable than the other networks. The results should thus be considered as a first exploration to investigate reciprocal relationships and should be interpreted with care. The same goes for the temporal network, in which I added only the EE variables which were present in the best GCM to avoid adding too many variables in the network. In the temporal network, the assumption of stationarity was violated and a network with detrended variables was added in the Supplementary Materials, but this network proved less informative. The networks did, however, enable us to explore relationships between L2 English vocabulary knowledge and various types of EE at different time points and across time points. The results of the analyses showed how these variables are interrelated and how both EE activities contributed to L2 knowledge and how L2 knowledge influenced learners' choice of EE activities. The focus of this study was on the development of vocabulary knowledge as this aspect of L2 English proficiency has been frequently investigated and I could thus build on previous studies to investigate reciprocal relationships. Future research should ideally also explore other aspects of L2 English proficiency. Finally, the questionnaire only considered EE exposure, whereas the way learners engage with EE input can also influence their learning gains (Calafato & Clausen, 2024; De Wilde et al., 2020). However, a more in-depth analysis of learners' engagement was beyond the scope of this study.

The findings of the present study showed the lasting impact of extramural English input on L2 vocabulary development. The study further revealed that the relationship between EE input and L2 English learning is not unidirectional but rather reciprocal. These findings also have implications for the classroom as they confirmed that EE input can be beneficial for L2 English learning and EE can be an important source of input for L2 English learners. Various types of input could benefit learners at different proficiency levels. The temporal network analysis, for example, showed that watching television with L1 subtitles was beneficial for learning, but it might not be the preferred activity for more proficient learners. The study also showed that large individual differences between children remained over time, something which should be considered by

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press

EFL teachers. Finally, the study also showed that despite these differences there was a clear and positive development of L2 vocabulary knowledge for all learners after the start of formal English instruction. This finding indicated that, even though differences between learners remained, formal English instruction was also important for L2 English learning.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728925100540.

Data availability statement. The data that support the findings of this study are openly available in OSF at https://osf.io/7ns4k/.

Funding statement. This research was supported by the FWO (Research Foundation Flanders, fellowship 1203923 N).

Competing interests. The author(s) declare none.

References

Brevik, L. M. (2016). The gaming outliers: Does out-of-school gaming improve boys' reading skills in English as a second language? In *Educational technology and polycontextual bridging* (pp. 39–61). SensePublishers.

Borsboom, D., **Deserno, M. K.**, **Rhemtulla, M.**, **Epskamp, S.**, **Fried, E. I.**, **McNally, R. J.**, **Robinaugh, D. J.**, **Perugini, M.**, **Dalege, J.**, **Costantini, G.**, **Isvoranu, A.-M.**, **Wysocki, A. C.**, **Van Borkulo, C. D.**, **Van Bork, R.**, & **Waldorp, L. J.** (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, **1**(1), 58. https://doi.org/10.1038/s43586-021-00055-w.

Busby, N. L. (2024). Exploring the effects of early extramural English exposure on the vocabulary size of university students. *Education Sciences*, **14**(4), 372. https://doi.org/10.3390/educsci14040372.

Bybee, J. L., & **Hopper, P. J.** (Eds.) (2001). *Frequency and the emergence of linguistic structure* (Vol. **45**). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.

Calafato, R., & **Clausen, T.** (2024). Vocabulary learning strategies in extramural English gaming and their relationship with vocabulary knowledge. *Computer Assisted Language Learning*, 1–19. https://doi.org/10.1080/09588221.2024.2328023

De Bot, K., **Lowie, W.**, & **Verspoor, M.** (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, **10**(01), 7. https://doi.org/10.1017/S1366728906002732.

Dewey, M. (2007). English as a lingua franca and globalization: An interconnected perspective. *International Journal of Applied Linguistics*, **17**(3), 332–354. https://doi.org/10.1111/j.1473-4192.2007.00177.x.

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.

De Wilde, V., **Brysbaert, M.**, & **Eyckmans, J.** (2021). Young learners' L2 English after the onset of instruction: Longitudinal development of L2 proficiency and the role of individual differences. *Bilingualism: Language and Cognition*, **24**(3), 439–453. https://doi.org/10.1017/S1366728920000747.

De Wilde, V., **Brysbaert, M.**, & **Eyckmans, J.** (2022). Formal versus informal L2 learning: How do individual differences and word-related variables influence French and English L2 vocabulary learning in Dutch-speaking children? *Studies in Second Language Acquisition*, **44**(1), 87–111. https://doi.org/10.1017/S0272263121000097.

De Wilde, V., & **Lowie, W.** (2024). The Forest and the trees: Investigating groups and individuals in longitudinal second language English speaking development. *Language Learning*, lang.**12698**. https://doi.org/10.1111/lang.12698

Dörnyei, Z., & **Ryan, S.** (2015). *The psychology of the language learner revisited*. Routledge, Taylor & Francis Group.

Dunn, L., & **Dunn, L.** (2007). *Peabody picture vocabulary test* (4th ed). American Guidance Service.

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.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, **24**(02). https://doi.org/10.1017/S0272263102002024.

Epskamp, S., **Cramer, A. O.**, **Waldorp, L. J.**, **Schmittmann, V. D.**, & **Borsboom, D.** (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, **48**, 1–18.

Epskamp, S., **Epskamp, M. S.**, & **Rcpp, L.** (2020). *Package 'psychonetrics'*. Computer Software.

Epskamp, S., & **Fried, E. I.** (2015). Package 'bootnet'. *Bootstrap Methods for Various Network Estimation Routines*, **5**, 0–1.

Epskamp, S., **Waldorp, L. J.**, **Mõttus, R.**, & **Borsboom, D.** (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, **53**(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823.

Freeborn, L., **Andringa, S.**, **Lunansky, G.**, & **Rispens, J.** (2023). Network analysis for modeling complex systems in SLA research. *Studies in Second Language Acquisition*, **45**(2), 526–557. https://doi.org/10.1017/S0272263122000407.

Gass, S. (2018). *Input, interaction and the second language learner*. Routledge.

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.

Hulstijn, J. H. (2012). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The Encyclopedia of applied linguistics*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0530.

Jaekel, N., **Schurig, M.**, **Florian, M.**, & **Ritter, M.** (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school: From early starters to late finishers? *Language Learning*, **67**(3), 631–664. https://doi.org/10.1111/lang.12242.

Kusyk, M., **Arndt, H. L.**, **Schwarz, M.**, **Yibokou, K. S.**, **Dressman, M.**, **Sockett, G.**, & **Toffoli, D.** (2025). A scoping review of studies in informal second language learning: Trends in research published between 2000 and 2020. *System*, **130**, 103541. https://doi.org/10.1016/j.system.2024.103541.

Kuznetsova, A., **Brockhoff, P. B.**, & **Christensen, R. H.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**, 1–26.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, **18**(2), 141–165. https://doi.org/10.1093/applin/18.2.141.

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.

Lowie, W. M., & **Verspoor, M. H.** (2019). Individual differences and the Ergodicity problem: Individual differences and Ergodicity. *Language Learning*, **69**, 184–206. https://doi.org/10.1111/lang.12324.

MacLeod, C. M., **Gopie, N.**, **Hourihan, K. L.**, **Neary, K. R.**, & **Ozubko, J. D.** (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36**(3), 671–685. https://doi.org/10.1037/a0018785.

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.

Pellicer-Sanchez, A., & **Schmitt, N.** (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart. *Reading in a Foreign Language*, **1**(22), 31–55.

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. (2022). The Golden age of foreign language learning. In H. Reinders, C. Lai, & P. *Sundqvist,* the Routledge handbook of language learning and teaching beyond the classroom (1st ed., pp. 112–125). Routledge. https://doi.org/10.4324/9781003048169-10

Peters, E., **Noreillie, A.**, **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.

Pfenninger, S. E. (2020). The dynamic multicausality of age of first bilingual language exposure: Evidence from a longitudinal content and language integrated learning study with dense time serial measurements. *The Modern Language Journal*, **104**(3), 662–686. https://doi.org/10.1111/modl.12666.

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.

R Core Team. (2024). *R: A language and environment for statistical computing (Version R version 4.4. 1) [Computer software]*. R Foundation for Statistical Computing.

Rød, A. J., & **Calafato, R.** (2025). Bridging extramural English and formal instruction. *ELT Journal*, **ccaf026**. https://doi.org/10.1093/elt/ccaf026.

Sundqvist, P. (2009). *Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary* (Doctoral dissertation, 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.

Sundqvist, P. (2024). Extramural English as an individual difference variable in L2 research: Methodology matters. *Annual Review of Applied Linguistics*, 1–13. https://doi.org/10.1017/S0267190524000072

Sylvén, L. K., & **Sundqvist, P.** (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, **24**(03), 302–321. https://doi.org/10.1017/S095834401200016X.

van Borkulo, C., **Boschloo, L.**, **Borsboom, D.**, **Penninx, B. W. J. H.**, **Waldorp, L.**, & **Schoevers, R.** (2015). Package 'NetworkComparisonTest'. *JAMA Psychiatry*, **72**, 1219–1226.

Van Dijk, M., **Lowie, W.**, **Smit, N.**, **Verspoor, M.**, & **Van Geert, P.** (2024). Complex dynamic systems theory as a foundation for process-oriented research on second language development. *Second Language Research*, 02676583241246739. https://doi.org/10.1177/02676583241246739.

Verspoor, M., **De Bot, K.**, & **Van Rein, E.** (2011). English as a foreign language: The role of out-of-school language input. In A. De Houwer & A. Wilton (Eds.), *AILA applied linguistics series* (Vol. **8**, pp. 147–166). John Benjamins Publishing Company. https://doi.org/10.1075/aals.8.10ver.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, **28**(1), 46–65.

Webb, S. A., & **Chang, A. C.-S.** (2012). Second language vocabulary growth. *RELC Journal*, **43**(1), 113–126. https://doi.org/10.1177/0033688212439367.

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, **3**(2), 180–185.

Winter, B., & **Wieling, M.** (2016). How to analyze linguistic change using mixed models, growth curve analysis and generalized additive Modeling. *Journal of Language Evolution*, **1**(1), 7–18. https://doi.org/10.1093/jole/lzv003.

Wood, S., & **Wood, M. S.** (2015). Package 'mgcv'. *R package version*, **1**(29), 729.

https://doi.org/10.1017/S1366728925100540 Published online by Cambridge University Press