ARTICLE

Study Protocol: STarting AGe and Extramural English: Learning English in and Outside of School in Norway and Flanders (STAGE)

Pia Sundqvist [1]* [iD] , *Elke Peters* [2] [iD] , *Gustaf B. Skar* [3] [iD] , *Henrik Gyllstad* [4] [iD]

[1] *Department of Teacher Education and School Research, University of Oslo, 0317 Oslo, Norway*

[2] *Department of Linguistics, KU Leuven, 2000 Antwerp, Belgium*

[3] *Department of Teacher Education, Norwegian University of Science and Technology, 7012 Trondheim, Norway*

[4] *Centre for Languages and Literature, Lund University, 22100 Lund, Sweden*

ABSTRACT

This Study Protocol article outlines the aims, theory, and methodology of a 4-year project that primarily seeks to enhance our understanding of the role of starting age and Extramural English. The project unravels the language proficiency of individuals studying English as a foreign or second language (L2) who encounter substantial exposure to English outside school (extramural English, EE), but begin formal English instruction at varying ages. The focus is on two regions: Norway, where formal English education commences early (grade 1, ages 6–7), and Flanders, where it begins later (grades 7 or 8, ages 12–14). The overarching project aim is to evaluate whether an early introduction to formal English instruction influences learners' language proficiency and development within contexts abundant in EE exposure, hence the chosen regions. The article is structured into four sections. First, the introduction presents the main objective, provides a theoretical background and previous research, and ends by suggesting that the project will make a substantial and innovative contribution to theories of L2 proficiency, and also contribute with scientific knowledge that will have societal impact. The second section provides an account of the research questions and our predictions, the material and methods, and the measures connected with each of the five research questions that guide the project. Section three offers a discussion that addresses the cross-sectional design and sample size (aim: $N = 900$), as well as potential limitations. Section four addresses ethics and project dissemination. The article ends with information about funding and acknowledgments.

*CORRESPONDING AUTHOR:

Pia Sundqvist, Department of Teacher Education and School Research, University of Oslo, 0317 Oslo, Norway; Email: pia.sundqvist@ils.uio.no

1. Introduction

The project *STarting AGe and Extramural English: Learning English in and outside of school in Norway and Flanders (STAGE)* primarily seeks to enhance our understanding of the role of starting age for English language instruction, out-of-school language contact (extramural English, EE), and English language proficiency. The project focuses on learners studying English as a foreign or second language (L2) who encounter substantial exposure to English outside school (extramural English), but begin formal English instruction at varying ages. The focus is on two regions: Norway, where formal English education commences early (grade 1, ages 6–7), and Flanders, where it begins later (grades 7 or 8, ages 12–14). This project's objective is to evaluate whether an early introduction to formal English instruction influences learners' language proficiency within contexts abundant in EE exposure.

Learners of L2 English engage with English through classroom-based formal instruction but also outside the school setting, a phenomenon often designated as *Extramural English (EE)* [1]. In regions where foreign language television programs are subtitled in the region's majority language—typically the first language (L1) for most children—L2 English learners are exposed to substantial amounts of EE from an early age, due to the predominance of English in popular culture [2]. We label these learning settings as input-rich contexts. As a result, areas such as Norway and Flanders benefit from enhanced language learning, heightened language awareness, and increased motivation in comparison to regions where foreign programs are dubbed [3].

Regarding the initiation age for formal L2 instruction, commencing L2 instruction at a young age does not seem to inherently lead to superior language proficiency when the input and instructional time are limited to two to four 50-minute classes per week [4–6]. Indeed, it is frequently observed that learners who begin at a later age quickly reach the proficiency levels of early starters, because older learners may benefit from more developed cognitive skills, which can support more effective learning strategies (see studies [4, 6] for recent extensive studies conducted in Germany and Switzerland). Research has shown that the quantity and quality of language input are more important for language proficiency than starting age [7]. This is different from the scenario observed among immigrant L2 learners in naturalistic settings [8], where early starters develop higher proficiency levels, especially in terms of pronunciation. Nevertheless, investigations into early L2 instruction have predominantly been conducted in regions such as Germany, Spain, and Switzerland, where learners' main contact with English is the classroom and where learners have limited EE.

There is a paucity of research regarding the effects of initiating L2 instruction early in input-rich environments where learners have extensive exposure to EE, such as in Norway. Furthermore, the limited results available is not definitive, with studies finding positive effects [9] and others finding no effect for an early start [2, 10]. It is important to note that participants in previous studies [2, 9, 10] were not assessed following an extended period of instruction, a factor crucial for distinguishing between the rate of learning and ultimate language attainment [5, 6, 8]. In addition to mixed findings, some studies suggest that EE may be a more significant predictor of language proficiency than formal education [11]. A pertinent investigation in this context is the *European Survey on Language Competences*, ESLC [12], a large-scale study examining reading, writing, and listening proficiency across learners in 14 countries. Speaking proficiency was not assessed in the ESLC. Although the ESLC did not specifically aim to assess the impact of starting age, the results indicated that 15-year-old L2 English learners in Sweden—a nation with educational practices akin to Norway—and those in Flanders scored similarly in reading and listening, despite Sweden's earlier starting age. However, the difference between regions such as Flanders and Norway/Sweden remains to be investigated empirically; there is a gap in comparative research targeting input-rich settings.

With regard to EE, learners' exposure to English often comes through media such as (subtitled) television, video

games, music, and social media, where children are exposed to authentic language in various contexts[2, 13, 14]. A rare large-scale study ($N = 3,048$) was carried out in Spain[15], showing age and gender differences in extramural English engagement amongst three groups of L2 English learners (younger adolescents aged 12–14; older adolescents aged 15–17; and adults aged 18–39). However, this study did not include any transnational comparison. The extant literature offers less data concerning young learners aged 7 to 11 years; notable exceptions include studies by Cadierno et al.[10] and Hannibal Jensen[2, 16] in Denmark, and Sylvén[17] in Sweden. Another exception is the comprehensive study on early L2 learning in Europe (Early Language Learning in Europe, ELLiE), in which data were collected from over 1,400 young learners in seven countries: Croatia, England, Italy, Netherlands, Poland, Spain, and Sweden[18]. These countries exhibit variances in demographics and geographical location in Europe, as well as in terms of linguistic characteristics, specifically, the Germanic, Romance, or Slavonic origins of the majority language in each country. Nonetheless, all countries implemented an early introduction of formal English instruction. A particular sub-study within the ELLiE project investigated the impact of out-of-school factors on children's listening and reading skills during their fourth year of English instruction ($N = 865$)[19]. The findings indicated that cognate linguistic distance was the strongest predictor of proficiency in both listening and reading, followed closely by exposure to the target language (in essence, EE). The educational levels of caregivers influenced reading scores only. Among various extramural activities, viewing films in English emerged as the most influential. The study further demonstrated a statistically significant correlation between caregivers' use of English in the workplace and their children's exposure to the language. The authors conclude that caregivers' influence may directly affect their own children's opportunities for target language exposure, which then would impact their learning outcomes[19]. With the exception of two studies[2, 14], there is a scarcity of research exploring how EE engagement evolves with age. In their comparative analysis of three age cohorts (10, 11, and 12 years old), Puimège and Peters[14] observed that 12-year-olds were more frequently involved in computer gaming and viewing English-language television than their younger counterparts. Nevertheless, a compre-

hensive understanding of how learners' engagement with English outside the formal educational setting changes as they age remains undeveloped.

Research investigating the relationship between EE and language development has shown that such informal learning experiences can enhance vocabulary acquisition[14, 20–22], listening comprehension[6, 13, 19, 23, 24], reading[19, 25, 26], writing[21, 22, 27], grammaticality judgement[23], and even speaking skills[1, 13, 28], as learners engage with real-world language. The advantages of EE have been found both for learners receiving formal instruction[19, 21, 24, 29] and for English language learners who have not yet undertaken formal English classes[13, 14, 30]. Puimège and Peters[14] demonstrated that Flemish sixth-grade students without prior English instruction possessed knowledge of approximately 3,000 word families, and De Wilde et al.[13] reported that 25% of these students attained an A2 proficiency level in listening comprehension as per the *Common European Framework of Reference for Languages*, CEFR[31]. It should be noted that not all EE modalities benefit language proficiency equally. Research shows that gaming and watching non-subtitled TV can improve productive language skills and that watching subtitled television can improve receptive vocabulary knowledge[14]. However, watching TV with L1 subtitles does not have the same effect[20]. However, to date, there is a notable absence of controlled, large-scale studies that systematically compare the association between various types of EE activities (e.g., playing computer games, watching television) and multiple L2 proficiency measures (e.g., speaking, reading, vocabulary, and grammar) across different age cohorts and instructional contexts. Furthermore, with the exception of the work by Puimège and Peters[14], few studies have accounted for learners' linguistic backgrounds—whether monolingual or multilingual (see below)—when investigating the advantages conferred by EE.

To our knowledge, the focus of the STAGE project is the first project (of size) to investigate the impact of an early start in an input-rich context by comparing early (Norwegian) and late (Flemish) starters' L2 English proficiency at three levels in school: in school years/grades 1, 6, and 10, respectively. Given that second language acquisition (SLA) research typically involves university students, our project will make a unique contribution by focusing on young(er)

learners. Further, both monolingual and multilingual learners will be recruited to avoid a monolingual bias[6]. Two data collection sites will be involved: Norway and Flanders. Our focus will be on three language proficiency measures: *vocabulary knowledge*, *reading*, and *speaking*. As part of *speaking*, will also take a closer look at *interactional competence*. It should be mentioned that *listening*, *writing*, *grammaticality judgements*, and *vocabulary* is the focus of a sister project in Flanders funded by the Flemish Research Council, FWO (Principal Investigator: Elke Peters; host university: KU Leuven). We anticipate that the STAGE project can have major academic and societal impact. STAGE will produce scientific knowledge of the relative benefits of formal and informal out-of-school English for three language proficiency measures (vocabulary, reading, speaking), which may affect curricular prioritization on policy level (role of starting age in language curricula), content in teacher education (age-appropriate EE activities for fostering English language proficiency), and teachers' practices.

2. Methods and Analysis

2.1. Research Questions and Predictions

We pose five research questions, which are presented below along with our predictions.

2.1.1. RQ1: What Characterizes the Use of Extramural English in the Two Sites, Norway and Flanders?

Building on prior studies[11, 14], we hypothesize that the primary forms of EE exposure will include watching television and films (regardless of streaming), gaming, listening to music, and engaging with English-mediated internet platforms, particularly YouTube. Additionally, we anticipate that learners' EE preferences in Norway and Flanders will shift from a preference for English-language television with L1 subtitles towards non-subtitled, and to a lesser extent, English-subtitled television. We also expect a decrease in gaming activity from grade 6 to grade 10, concurrent with an increase in the reading of English-language books during this period[29]. We do not anticipate differences between Norway and Flanders in this regard.

2.1.2. RQ2: Is There a Difference in *Vocabulary Knowledge*, *Reading*, and *Speaking*, between the Norwegian and Flemish L2 English Learners? Do the Language Proficiency Measures Develop Differently in the Early Start Context, Norway, Compared to the Late Start Context, Flanders?

We hypothesize that in grade 6, early starters will outperform late starters in vocabulary, reading, and speaking skills[4], but see Jaekel, et al.[32]. Concerning speaking performance in grade 6, we anticipate differences in holistic scores—specifically for the test construct of *oral production and interaction*[33, 34] —that favor early starters. Additionally, we hypothesize that by grade 10, these differences will have dissipated as late starters are expected to catch up[4–6].

2.1.3. RQ3: Which Variables Predict the Language Proficiency Measures *Vocabulary Knowledge*, *Reading*, and *Speaking*?

Our hypothesis is that instruction as well as EE will affect language proficiency positively[11, 29]. We think EE (rather than instruction) will be a determinant predictor of *vocabulary knowledge*, but not of *reading* and *speaking*.

2.1.4. RQ4: How Do (Reported) English Language Practices Relate to Identity Work and Beliefs about Language Learning among Children and Adolescents in Norway?

Previous research[35–37] has shown the importance of considering identity in L2 learning. There are also clear links between identity and motivation for L2 learning, as explained by L2 motivational self-system theory[38–40] and empirical work that has tested the theory[39]. In the context of Norway, Rindal[41–43] has spearheaded research on identity work amongst adolescent L2 English learners in conjunction with questions about attitudes towards various accents of spoken English, and RQ4 is closely connected with this specific area of research. We hypothesize that both children (in grades 1 and 6) and adolescents (in grade 10) will consider learning English as important, but that their language practices relating to identity work and beliefs about

language learning will be slightly different depending on their age and degree of engagement in EE activities. Further, we anticipate that English will play an important role in the everyday lives of children and adolescents, and that the position of English in a social hierarchy will differ in relation to other languages in the linguistic repertoires of individuals. In addition, we anticipate that EE is closely related to the participants' interests and hobbies, and therefore that English plays a considerable role in identity work and social positioning, among adolescents especially. Finally, we anticipate that language practices will be influenced by media genres (see, e.g., reports about media use[44]), in particular among participants in grades 6 and 10, and that these practices will differ from expectations of language use in school.

2.1.5. RQ5: How Does Learners' Interactional Competence Align with Experiences of EE and English Instruction?

Drawing on recent advancements in the operationalization of L2 interactional competence grounded in the principles of social interaction[45], we will explore specific sub-competencies, such as aspects of topic management, the demonstration of mutual understanding, and the resources employed for constructing actions, such as turn-taking or repair[46], which are indicative of L2 interactional competence. We will contrast EE profiles, holistic speaking test scores, and instructional background data with qualitative analyses of test interactions. Our focus will be on grade 10 students who have different experiences with EE. Since holistic test scores do not specifically indicate what learners can achieve interactionally with partners, RQ5 will provide qualitative insights into the potential impact of EE on the range of competencies that learners exhibit. We hypothesize that extensive exposure to EE, particularly through activities involving dialogue and interaction (such as television, YouTube, and interactive L2 gaming), will be reflected in learners' relative interactional competence during speaking tests. Given that even learners with low proficiency can exhibit skills in responding to a co-interactant's contributions, we do not presume that an early start alone will result in higher interactional competence. Instead, we anticipate that a longer duration of instruction may be advantageous for certain features of interactional competence, such as more advanced topic management.

2.2. Empirical Studies

To answer the research questions, five empirical cross-sectional studies will be conducted. Study 1 will concentrate on learners' EE exposure (RQ1). Study 2 will elucidate the impact of starting age on learners' performance across various language proficiency measures (RQ2). Study 3 will explore the relationship between starting age, EE, and the different measures of language proficiency (RQ3). Study 4 will provide insights into participants' identity construction and their perspectives and beliefs regarding the roles of EE and formal English instruction in proficiency development (RQ4). Lastly, Study 5 will evaluate speaking and interactional competence in the context of learners' experiences with EE and formal instruction (RQ5).

2.3. Sampling: Inclusion Criteria and Strategy

The total sample size across both sites (Norway and Flanders) will comprise 900 students. We will sample 50 student participants from grade 1 (age 6–7 years) per site (i.e., $n = 100$), 200 student participants from grade 6 (age 11–12 years) per site (i.e., $n = 400$), and 200 student participants from grade 10 (age 15–16 years) per site (i.e., $n = 400$). Because we will use multilevel modelling (hierarchical regression analysis) with an unknown number of independent variables (see Section 2.4), we have not been able to estimate the statistical power using specialized software like *MLPowSim*[47]. Using G*Power[48] for a less accurate estimate of the statistical power (assuming a *t*-test between two groups [i.e., in this case: sites]), the sample size should be enough to detect a difference of $d = 0.57$ for grade 1, and $d = 0.28$ for grade 6, and grade 10, respectively. It should be noted that we hypothesize that grade 1 students will not differ across sites because none of them has had formal schooling (see below).

For students in grades 1 and 6, eligible for inclusion in the sample will be all students at partner schools of the project whose parents/caregivers provide written consent, and whose teacher deems the participation to be feasible. Conversely, students whose parents/caregivers do not present a written consent, or whose teacher deems participation to interfere with a students' learning, will not be eligible for inclusion. As an example of the latter, consider a teacher who might assess it to be disadvantageous for a student who is largely

absent from instruction (for various reasons, for instance, struggling with motivation) to participate in the project. For students in grade 1, the plan is to collect data in the fall semester. This means that the participants from Norway will have had some formal instruction, but very little. The data collected in grade 1 are considered as baseline data. It should be noted that participants born very late in the year may only be 5 years old at the time of data collection. For students in grade 6, the plan is to collect data in the fall semester, while data collection for students in grade 10 will take place in the spring semester. For grade 10, this means that all students will be 15 or 16 years when data are collected; this, in turn, means that they will be old enough to make their own decision regarding participation, as 15 years of age is the "cutoff" for making an own informed consent in Norway. All students aged 15 or older will, thus, present their written consent themselves and those who agree to participate will be eligible for inclusion.

The recruitment of participants will be done through a non-probability sampling method, since probability sampling will not be feasible due to time, cost, and logistical constraints. Specifically, we will approach schools in the researchers' networks, and accordingly achieve a convenience sample. We will use publicly available data sources to retrieve information about the schools' size, the socio-economic status of neighborhoods of participating schools, and other relevant data (e.g., proportion of certified teachers). The finite list of school level variables will be contingent upon what type of data that will be available across both sites. At the time of this reporting, that is still unclear.

The school-level data will be used to describe the final sample in relation to the national population of schools, and to describe similarities and differences between schools from the two sites (i.e., Norway and Flanders). The school-level data will also serve as independent variables in the ensuing analysis, which will allow us to control for difference between schools.

2.4. Measures

All RQs will be answered through designated work packages (WPs). The account below explains in more detail the outcomes to be measured and data to be collected within each WP, the data management plan for each WP, and the data analytical strategies for each WP.

2.4.1. RQ1 (Work Package 1)

To answer the first research question ("What characterizes the use of Extramural English in the two sites, Norway and Flanders?"), we will survey parents/caregivers (grades 1 and 6) and students (grades 1, 6, and 10) regarding student participants' experience of extramural English. As part of WP 1, we will also conduct a systematic literature review targeting educational policy on the starting age of formal English language teaching world-wide.

(1) Outcomes to be Measured

We will tap into learners' EE and demographic background information through a parental questionnaire in grades 1 and 6 (available in the **Supplementary File S4**). The participating children will answer an EE questionnaire in school in grade 1 (orally, each participant 1-on-1 with a researcher) and in grade 6 (individually, on paper) (available in the **Supplementary File S1** and **File S2**, respectively). In grade 10, the EE questionnaire will be digital and include the EE Scale [49]. The questionnaire for grade 10 (**Supplementary File S3**) will also tap into teaching practices to control for any student-reported differences in teaching between the two sites, as well as participants' views on and beliefs about L2 English learning.

Further, we will ask participants in grades 1 and 6 to complete an identical, analog (paper and pencil) one-week EE Language Diary based on Hannibal Jensen's [2] language diary for young learners in Denmark (for a sample page, see **Supplementary File S5**). The Language Diary will be filled out in the home. Especially for the youngest participants in grade 1, the Language Diary is filled out with the help of the parents/caregivers. In grade 10, the EE Language Diary (see **Supplementary File S6**) is loosely based on Sylvén and Sundqvist's [24] language diary for Swedish adolescent learners, administered digitally every day for the duration of a week, asking for a daily report (or log) of which EE activities the participants have been engaged in, and for how long time per activity.

Furthermore, we will interview a selection of students from all three grade levels about their EE habits, beliefs, and more (for practical reasons, in Norway only). In grade 1, we will use some of the interview materials provided in the validated European Language Portfolio for preschool class and grade 1 in Sweden [50]. In grades 6 and 10, we will create interview guides building on, for example, Sundqvist [21] (see

Supplementary File S7). **Table 1** summarizes the outcome measures that will be used as dependent variables.

Table 1. Outcome Measures from Work Package 1.

Site	Grade	Level		EE	
Norway	1st Grade	Student	Time spent on EE activities survey (filled out with the help of caregivers) (Language Diary)	EE activities frequency survey (Learner EE Questionnaire)	Interviews at school + Informal interviews using artefacts in combination with observations in the home
		Parent/Caregiver	Parental Questionnaire		
	6th Grade	Student	Time spent on EE activities survey (Language Diary)	EE activities frequency survey (Learner EE Questionnaire)	Interviews*
		Parent/Caregiver	Parental Questionnaire		
	10th Grade	Student	Time spent on EE activities survey (Language Diary)	EE activities frequency survey (Learner EE questionnaire)	Interviews*
Flanders	1st Grade	Student	Time spent on EE activities survey (filled out with the help of caregivers) (Language Diary)	EE activities frequency survey (Learner EE Questionnaire)	
		Parent/Caregiver	Parental Questionnaire		
	6th Grade	Student	Time spent on EE activities survey (Language Diary)	EE activities frequency survey (Learner EE Questionnaire)	
		Parent/Caregiver	Parental Questionnaire		
	10th Grade	Student	Time spent on EE activities survey (Language Diary)	EE activities frequency survey (Learner EE Questionnaire)	

* Subsample (n = 15; Norway only).

(2) Data Management Plan

The data management plan (DMP) for the whole project was developed closely after the funding for this project was obtained, in compliance with data protection regulation following the General Data Protection Regulation (GDPR) in the EU (https://gdpr-info.eu/) and DMP guidelines (https://sikt.no/en/data-management-plan) provided and approved by the Norwegian Agency for Shared Services in Education and Research, Sikt (https://sikt.no/en/home). The data classification guide provided by the host institution (https://www.uio.no/english/services/it/security/lsis/storage-guide.html) points to suitable solutions for storing research data, and this guide was helpful in developing the DMP for this project (i.e., for all the data in the five work packages).

Primary data will be stored at a centralized storage facility which is designed for secure and efficient storage of primary data as part of a shared infrastructure (Norw. *primærdata samhandlings infrastrukturer (PSI)*; also *lagringshotell*). This shared space is at the host institution and has frequent backups. Different levels of accessibility will be put on the different types of data files being stored. For instance, only the principal investigator and project members employed by the host institution will have access to the main data folder, which will contain the original data files containing participants' (and at times also parents'/caregivers') names and contact information.

Detailed guidelines will be used to make sure all data are treated in a secure way, from data collection to storage. Physical (paper-based) data will be digitalized at the university, and digital data will be collected using an online data collection tool called Nettskjema (https://www.uio.no/english/services/it/adm-services/nettskjema), which is maintained by the University Information Technology Center at the University of Oslo. Nettskjema has a high level of security and is approved by the Norwegian Agency for Shared Serviced in Education and Research. To access data, two-factor authentication will be used. Data collected on paper will be stored in locked cabinets until they have been digitalized, after which they will be destroyed. Security measures such as two-step authentication and encrypted files will be employed.

A codebook will be used to label and save all data files systematically. All digitalized data will then be pseudonymized; that is, personal data, such as participants' names, will be replaced by codes. A physical keycode (including the participants' names and their own unique code) will be stored separately in a locked space to which only the principal investigator has access, with a hard copy stored as backup in a locked space in the home of another project member employed by the host institution.

Pseudonymized data files will be shared with all project team members using an online storage service that the host institution recommends for so-called restricted (or "yellow") data, that is, data that need a certain level of protection, for example, test scores and research data that remain to be published. Video data will be stored in the storage hotel only; researchers not employed by the host institution who will need access, will be assigned temporary guest accounts with the host institution.

(3) Data Analysis

Because of the clustered nature of the data set (participants will be recruited at school level), all data will be analyzed using a multilevel modelling approach[51]. As dependent variables, the ones listed in **Table 1** will be employed (i.e., the levels of engagement in various EE activities, measured through questionnaires and language diaries, providing data about frequency as well as duration/time).

(4) Timeline

Data collection will begin in the fall semester of 2022 and last to the spring semester of 2024. We will clean and begin analyzing some data when they have been pseudonymized, but the lion's share of data analysis will take place after all data have been pseudonymized (planned for early fall semester, 2024). We will start disseminating the results in 2025.

2.4.2. RQ2 (Work Package 2)

We will answer the second research question ("Is there a difference in the language proficiency measures *vocabulary knowledge*, *reading*, and *speaking*, between the Norwegian and Flemish L2 English learners? Do the language proficiency measures develop differently in the early start context, Norway, compared to the late start context, Flanders?") by measuring the vocabulary knowledge, reading comprehension, and speaking proficiency of children in grades 1, 6 and 10 in Norway and in Flanders. The outcomes to be measured

are detailed below and summarized in **Table 2**.

(1) Outcomes to be Measured

We will use the *Picture Vocabulary Size Test (PVST)* (Version 1.2.0, Test Set 1)[52] to measure participants' receptive vocabulary knowledge at the level of meaning recognition. The PVST has proven to work well with children in primary school[14]. It is a multiple-choice test and we will use the original version for grades 6 and 10, in which a total of 96 words are tested. For participants in grade 1, we will use an age-appropriate truncated version (item [k = 31]) based on age-of-acquisition ratings of the target words[53]. The PVST is specifically designed to measure the receptive vocabulary size of L1 as well as L2 learners. This test estimates knowledge pertaining to word families, sampling target items from the 6,000 most frequently occurring word families in English, derived from a corpus comprising five million tokens of spoken and written language. This corpus includes an array of sources such as movies, television programs, children's books, educational materials, and colloquial spoken American and British English texts.

In the PVST test format, each target word is embedded within a non-defining sentence that indicates the part of speech. The target item (and its sentence) is accompanied by four images. Participants will answer by selecting the correct image, that is, the image that corresponds to the meaning of the target item. While showing the slide with the target item embedded in a sentence plus the four images, there is also accompanying audio/sound, so both the spoken and written forms of the target word will be presented to the participants when they answer the PVST. We decided to add an "I don't know" (IDK) option. Even though the use of an IDK option has both advantages, such as suppressing excessive random guessing, and drawbacks, such as discouraging drawing on partial word knowledge, our participants will be clearly instructed to avoid guessing and choose the IDK option if the target word meaning is unknown to them. In a study by Zhang[54], the use of an IDK feature decreased the number of random guesses, and the author recommended using the IDK in cases where this is a desirable goal. In grade 10, we will also use the *Vocabulary Size Test (VST)*[55] in order to better discriminate between more advanced participants, as this test also targets more low-frequency words. The VST exists in monolingual and bilingual versions (see Paul Nation's webpage: https://www.wgtn.ac.nz/lals/resources/paul-nat

ions-resources/vocabulary-tests). The monolingual version furthermore comes in two subversions: one targeting words (word families) from the 1–14K frequency bands, which is the one that will be employed by us, and another targeting words from the 1–20K frequency bands. Just like the PVST, the monolingual VST is a receptive multiple-choice test. Each test item provides the targeted English word in a short, non-defining sentence followed by four meaning alternatives, and for the same reasons mentioned above when describing the PVST, an IDK option will be added. Thus, the modality tested is meaning recognition. In the 1–14K version, the test has 14 sections with 10 target words in each, amounting to a total of 140 items. The words in each section have been sampled from a corresponding frequency band of 1,000 words. The scores in each section are added up and extrapolated to indicate an individual's total vocabulary size. For example, a total score of 90 out of the total 140 means that the test-taker is estimated to have a vocabulary size of 9,000 word families. The VST has undergone a number of validation studies, some of which yielded positive evidence of various facets of validity[56, 57], whereas others highlighted some issues and limitations[58, 59]. On balance, due to the lack of a competitive alternative to the VST, the test is arguably the only viable test that can capture more advanced learners' vocabulary sizes.

In addition, we will develop and trial a new picture-based vocabulary test intended for young learners around the age of 6–7 (1st grade). The purpose of the new test is to tap into young learners' meaning recognition and meaning recall vocabulary knowledge, and potentially have a section where their budding fluency can be tapped into. There have been several calls for such a test to be developed, pointing to issues with some existing tests (see, e.g., Aus der Wieschen[60], and Goriot, et al.[61]). The aim is to create a test that can be used for various populations, meaning that it should contain target words corresponding to concepts that can be argued to be more or less universal and known to the test-takers. Other considerations are of a more practical nature; it is well-attested that young learners have a limited attention span[62]. Consequently, a new test must be fairly short (∼30 items). Other criteria are that preschool language learners in most cases lack a precise mastery of their first language, and that the selection of targeted words be age-appropriate and relating to children's cognitive development[63].

We will use the reading test from the *European Survey of Language Competences*, ESLC[12], to measure reading comprehension. This test has several sections, each corresponding to a CEFR proficiency level: A1, A2, B1, B2[31]. Learners will be given all test sections. However, given that learners in grade 6 cannot be expected to be able to do a B1 or B2 reading task, they will be asked to tick off a box indicating that they are willing to move on to the next test section. It is possible that some grade-6 learners will be proficient enough to complete one (or two) reading task(s) at the B-level; if so, the [reading] experience is likely to be highly motivating for them. This approach allows us to compare reading between two grade levels (i.e., grades 6 and 10) and between two settings (i.e., Norway and Flanders) without demotivating grade-6 learners.

For Speaking, we will use the freely available and standardized *National English Speaking Test* (NEST) from Sweden, used in grades 6 and 9 ("NEST6" and "NEST9," respectively)[64]. The NEST is aimed at measuring the test construct "oral production and interaction" (for details, see chapter 1[65]). A passing grade for NEST6 corresponds to CEFR A2.1 (basic user) and a passing grade for NEST9 corresponds to CEFR B1.1 (independent user). Participants in both grades 6 and 10 will take the NEST in groups of 3 and performances (video recorded) will be assessed using a holistic grade, from A to F. The assessment procedure will be in accordance with the instructions created by the test constructors for teachers in Sweden[64]. Both NEST6 and NEST9 are typically assessed by the students' own L2 English teacher, who thus acts as an examiner of their own students' NESTs[66–68]. We intend to recruit in-service English teachers from Sweden, who have experience of assessing NEST6 and/or NEST 9, as raters of the NEST data collected in this project. This means that the recruited raters will be very familiar with the assessment procedures and assessment criteria.

(2) Test Development

As indicated above, WP 2 will also include developing and trialing two tests for young learners aged 6–7 (1st grade): a new picture-based vocabulary test and a new picture-based speaking test. The two tests will most likely employ some words/pictures that are identical, to allow for comparison across the two tests and across L2 English abilities (receptive vocabulary knowledge measured through meaning recall and

productive vocabulary knowledge measured through speaking). All pictures will represent age-appropriate words that are known by same-age native speakers of English[53].

Table 2. Data collection instruments and number of participants per Test/Grade/Site.

Outcome Measure	Grade 1	Grade 6	Grade 10
Picture vocabulary size test (PVST)*	X	X	X
New picture-based vocabulary test	X		
Vocabulary size test (VST)			X
Reading test (ESLC)		X	X
Speaking test (NEST)		X#	X#
New picture-based speaking test	X		
Total (N) participants per site	50	200	200
Total (N) participants per grade	100	400	400
Total (N) participants, two sites			900

Note: * = Shorter version for Grade 1; # = fewer for Speaking ($n = 42$).

(4) Data Management Plan

Again, we will employ a multilevel modelling approach due to the nested nature of the data, but also modern language testing theory techniques, such as item response theory (1PL Rasch)[69]. Each outcome measure will be dependent variables in the analyses. We will control for clustering, but not include additional explanatory variables, as these will be included in the analysis done in WP 3.

(5) Timeline

Test development is time consuming and planned for three years, 2022–2024. Data collection will take place in several rounds for each stage of test development (i.e., for tapping into young learners' meaning recognition and meaning recall vocabulary knowledge, with an inclusion of a section targeting emerging oral skills) and is planned for 2023 and 2024. We expect the data analysis to be ongoing during the different stages of test development and trialing, and to be finally completed in the beginning of 2025. Results will then be disseminated.

2.4.3. RQ3 (Work Package 3)

In the third WP 3 we will utilize already collected data (Studies 1 and 2).

(1) Data Analysis

To answer the third research question ("Which variables predict the language proficiency measures *vocabulary knowledge*, *reading*, and *speaking*?"), we will use the data collected in WP 1 and WP 2 to model language proficiency. We will run two types of analyses. First, we will use multilevel modeling[51, 70] to predict language proficiency with

(3) Data Management Plan

The DMP connected with WP 2 follows the same principles as the DMP described in Section 2.4.1. (2) above.

data collected in WP 1 as predictors and data collected in WP 2 as dependent variables. Second, we will use Structural Equation Modelling[71] to model the relationship between language proficiency factors and factors made up from WP 1 data.

(2) Timeline

The analyses will be conducted in the fall of 2024, and dissemination will take place in 2025.

2.4.4. RQ4 (Work Package 4)

We will use interview and ethnographic data collected in Study 4 to investigate the overarching aim of the proposed project from the stakeholder perspective of learners. As argued by Pinter[72], children are indeed able to contribute to research that pertain to their own L2 learning, but they are rarely the focus of second language acquisition studies. Learner views will be compared across age groups but, for practical reasons, not across settings (e.g., the involved researchers in this WP do not speak Dutch, and recruiting a Dutch-speaking researcher to conduct interviews on site in Flanders on the behalf of our project was deemed suboptimal from a cost-benefit research perspective).

While we will have some survey data on participants' views on formal and informal learning of English and their own attitudes towards English, we will rely most heavily on qualitative data from interviews and observations in answering RQ4.

(1) Data Analysis

To answer the fourth research question ("How do (reported) English language practices relate to identity work

and beliefs about language learning among children and adolescents in Norway?"), we will draw on data collected in Study 1 and Study 4. As mentioned in the previous section, this investigation will mainly be qualitative. We plan to use a thematic analysis approach[73]. We will code interview and observational data in an iterative process where close readings of the material will generate codes, which will lead to a re-reading of the material, potentially discovering new codes. This iterative process will stop when we have codes sufficient enough to cover the material.

For our youngest 1st-grade participants, in addition to individual interviews at the school, informal interviews with artefacts in combination with observations in the home will be used[74]. The purpose is to tap into the participants' engagement in various extramural English activities and their identity work, as well as their views on and beliefs about L2 learning. In the home interview, the researcher and the participant will wear lapel microphones to record sound from oral interactions between the researcher and the participant, plus sound from interactions with the different family members. Informed consents will be collected from all family members at the time of the interview. To analyze identity work, views, and beliefs amongst 6th- and 10th-grade participants, individual interviews will be used (Norway only).

(2) Timeline

Data collection is planned for 2022 and 2023, and data analysis for 2023 and 2024. Dissemination will take place in 2025.

2.4.5. RQ5 (Work Package 5)

We will use data collected in Studies 1 and 2 in grades 6 and 10 to examine learners' L2 English speaking (test construct: oral production and interaction) using assessment data and conversation analysis, and contrast these findings with learners' exposure to EE and English instruction. In addition, we will use conversation analysis to specifically analyze interactional competence (grade 10 only, both settings). **Table 3** shows what data will be collected as part of WP 5.

Table 3. Data collection for examining L2 English speaking and interactional competence.

Data/Instrument	Grade 6	Grade 10
National English Speaking Test (NEST)	X	X
In-service teachers' assessments (3 per participant)	X	X
IC experts' assessments*		X

Note: * Two experts will assess the interactional competence of a subset of the NESTs.

(1) Data Analysis

We will use Conversation Analysis (CA) to analyze the speaking data[45, 75]. CA is a key approach to the study of interaction in L2 speaking. NESTs, designed to be administered to students either in pairs or small groups, is a speaking test that aims to elicit interaction between the test-takers[65]. Research has shown that dyadic and small-group tests better resemble natural conversation than, for example, oral proficiency interviews, which as the name reveals has an interview format[76].

Since the late 1990s, CA has been recognized as an established methodology for investigating language learning and teaching[77]. It is a method that aims to describe and comprehend the sequential and temporal organization of everyday social activities[78] and focuses on the resources, methods, and practices that participants in interaction utilize and achieve, which they also render publicly visible, thus allowing these elements to be observed, analyzed, and responded to by others. A fundamental tenet of ethnomethodological CA, encompassing both theoretical and methodological dimensions, pertains to the inherent orderliness of social interaction; interaction is presumed to be orderly at all junctures[79, 80]. It is the analyst's responsibility to uncover the orderly practices through which, in this case test-takers, fulfill their tasks, roles, and relationships.

We will analyze quantitively in-service teachers' assessments of the speaking test data in order to identify participants' whose NEST ratings are not in agreement, since such differences in assessment have been shown to indicate that the specific test performances in question, for various reasons, are perceived as problematic by raters[34, 81]. Such test performances with deviant assessments will undergo CA work to gain further insights into the characteristics of test interaction and potential influence on assessment. The

relationship between the level of speaking proficiency and engagement in EE will also be examined (see also WP 2).

(2) Timeline

Speaking test data will be collected in 2022, 2023, and the spring of 2024. Over the summer of 2024, assessment data will be collected using Nettskjema from in-service L2 English teachers in Sweden with experience of assessing NEST6 and NEST9. Data analysis is planned for the fall semester of 2024 and the beginning of 2025. Dissemination will take place in 2025.

2.5. Personnel

The personnel of the STAGE project consist of co-workers in Norway (NO), Sweden (SE) and Flanders (FL). The principal investigator (PI) is presented first, followed by researchers in alphabetical order and Ph.D. students in alphabetical order.

2.5.1. Principal Investigator

Pia Sundqvist, professor, University of Oslo (NO).

2.5.2. Researchers and Project Members

Henrik Gyllstad, associate professor, Lund University (SE); Marie Källkvist, professor, Linnæus University and associate professor Lund University (SE); Elke Peters, professor, KU Leuven (FL); Ulrikke Rindal, associate professor, University of Oslo (NO); Erica Sandlund, professor, Karlstad University (SE); and Gustaf B. Skar, professor, Norwegian University of Science and Technology (NO).

2.5.3. PhD Students

Kathrine Staksrud, Ph.D. student, University of Oslo (NO) (an external position as Ph.D. student tied to the STAGE project, which means that the position is funded by her home institution, Østfold University College) and Nasrin Ulfat, Ph.D. student, University of Oslo (NO).

3. Discussion

The purpose of STAGE is to advance our understanding of the impact of an early start for English instruction and Extramural English on reading, speaking, vocabulary knowledge, and interactional competence. To do so we have chosen to employ a cross-sectional design in two contexts, one where children encounter onset of formal English instruc-

tion in first grade (Norway) and one where they encounter it in 7th or 8th grade (Flanders). While we do believe that the data collection and analysis presented above will aid us in our pursuit for more knowledge on the topic, we are aware that the sample size, the sampling strategy, and the cross-sectional design have some weaknesses. We will briefly discuss them below, as well as how those weaknesses will constrain interpretations of the results.

The sample size aimed for amounts to 900 students in total, distributed as 100 students in first grade (50 at each site), 400 in sixth grade (200 at each site) and 400 in tenth grade (200 at each site). This relatively modest sample size and the fact that students will be clustered in classes in schools will require differences between sites to be large for them to be significant (see Dorman[82] for an explanation of this phenomenon). The sample size might generate false negatives, which means that any interpretation of null findings must be conservative. A null finding, if not interpreted and communicated adequately, might encourage policy makers to suggest postponement of English instruction in Norway.

The sampling strategy will allow us to recruit the desired number of participants but will also inevitably mean built-in sampling error. Far from all students in Norway and Flanders will have the same probability to be part of the study. One might suspect that schools and teachers opting to be a part of the project and to inform parents/caregivers about it, will be favorable towards English instruction and the importance of extramural English – and not representative of the population of schools and teachers. To remedy this built-in sampling error, we will describe the characteristics of the participating schools, and compare those characteristics with national averages. While this will not reduce the sampling error per se, it will allow us to gauge how similar or dissimilar our sample is to the population average. In turn, this will help us to better communicate nuanced interpretations of the results.

The cross-sectional design allows us to collect data in a much more limited time frame than a longitudinal design would have. We will use the differences between students of different grade levels in school and students in different sites as a proxy for development. This approach is not flawless, however. First, it assumes that all students in the project at the respective site are or have been exposed to the same levels of EE and equitable instruction. Second, it assumes

few, if any, individual variations in acquisition through EE and instruction. Again, these will be caveats important to remind our audiences about.

The STAGE project is an ambitious international project with a complex research design, but it will be feasible to carry out. For example, to enhance its feasibility, the language tests used for measuring different aspects of L2 English proficiency have been validated and successfully used in previous research, and although the data collection is relatively large, it will take place within a reasonable period. Also related to feasibility, the time available to visit schools for data collection will be limited because of, for example, theme days and school trips. However, careful planning and close collaboration with teachers should alleviate any major problems.

While a longitudinal design (also referred to as a panel study) over ten years (from 1st grade to 10th grade) may have been preferred, the cross-sectional design adopted here will still allow for making careful aims as regards L2 development in relation to age. Moreover, it should be mentioned that longitudinal studies, despite some obvious advantages (e.g., robust investigation into cause and effect), are not unproblematic. For instance, longitudinal research designs entail significant financial and temporal investments, necessitating the sustained dedication of a research team over extended periods. In addition, such designs are subject to two critical threats to their validity: attrition (participants drop out of the panel in the different phases of data collection, severely reducing the final sample of the panel – and the reduction may certainly not be random) and panel conditioning (participants are affected by being part of the panel, so there is a practice effect, possibly combined with a "Hawthorne effect", which means that participants perform differently simply because they are aware of the fact that that they are being studied)[83].

4. Ethics and Dissemination

The STAGE project has been reviewed and approved by the Norwegian Agency for Shared Services in Education and Research, Sikt (approval number: 900288).

When preparing the research application, four scholars were recruited as members of an advisory board. These four members were invited based on their respective expert knowledge to provide expertise in areas deemed relevant to the project. They will offer feedback and sharpen the research by scientific scrutiny. Members of the advisory board are professors Raphael Berthele, University of Fribourg, Switzerland (expertise: age and L2 learning), Batia Laufer, University of Haifa, Israel (L2 vocabulary acquisition and reading), Hayo Reinders, Anaheim University, USA and KMUTT, Thailand (L2 learning beyond the classroom), and Shannon Sauro, University of Maryland – Baltimore County (technologically-mediated language teaching/learning and L2 literacy). These four scholars will help ensure the scientific quality and integrity of the STAGE project.

Related to ethical considerations, we anticipate some potential challenges in the data collection. First, we will collect data from children and adolescents. This means that informed consent must be obtained from all participants, and with additional informed consent from parents/caregivers of children under the age of 15. Second, finding schools willing to cooperate in research is known to be challenging. For schools participating in the STAGE project, in exchange for participation, they will receive feedback reports and an invitation to a symposium (see below), where researchers will present the findings. During the actual data collection, as a token of gratitude, for all teachers interested, we will offer a "bonus" age-appropriate language activity for their participating class(es).

The data management plan and the rigor with which data will be handled and managed, where researchers and research assistants will be guided by well-developed instructions in the various phases of the project, will contribute to establishing and maintaining good research ethics (see also 2.4.1).

With regard to dissemination, in addition to articles planned for Studies 1–5, the project findings will be presented at national, Nordic, European, and international conferences. Towards the end of the project, the plan is to organize a free two-day symposium to disseminate the research findings to teachers and policymakers (one day), and to researchers (one day), with an aim to set up an international research network. We will publish at least one popular publication for teachers and policymakers in a professionally-oriented journal, and continuously publish updates from the project on a project website, including offering video summaries of all studies. Project updates will also regularly

be posted from the project members' various social media accounts and on the PI's home department news page.

Supplementary Materials

The supplementary materials can be downloaded at https://journals.bilpubgroup.com/public/FLS-10412.zip. File S1: EE Questionnaire (EEQ) for Grade 1 (in Norwegian and English). File S2: EE Questionnaire (EEQ) for Grade 6 (in Norwegian and English). File S3: EE Questionnaire (EEQ) for Grade 10 (in Norwegian and English). File S4: Parental Questionnaire (PQ) for Grade 1 and 6 (in Norwegian and English). File S5: Language Diary (LD) for Grade 1 and Grade 6 (in Norwegian and English, sample page for Monday). File S6: Language Diary (LD) for Grade 10 (in English). File S7: Interview guide for Grade 10 (in Norwegian and English).

Author Contributions

Conceptualization, P.S., E.P. and H.G.; methodology, P.S., E.P., G.B.S. and H.G.; validation, P.S.; data curation, P.S. and E.P.; writing—original draft preparation, P.S., E.P., G.B.S. and H.G.; writing—review and editing, P.S., E.P. G.B.S. and H.G.; project administration, P.S.; funding acquisition, P.S. and E.P. All authors have read and agreed to the published version of the manuscript.

Funding

Institutional Review Board Statement

The STAGE project has been reviewed and approved by the Norwegian Agency for Shared Services in Education and Research, Sikt (approval number/protocol code 900288, 22 April 2022).

Informed Consent Statement

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

Data Availability Statement

After the end of the project and when all key outputs from the STAGE project have been published, written anonymized data will be shared in an open data repository.

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] Sundqvist, P., 2009. Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary [PhD Thesis]. Karlstad University: Karlstad, Sweden.

[2] Hannibal Jensen, S., 2017, Gaming as an English language learning resource among young children in Denmark. CALICO Journal. 34(1), 1–19. DOI: https://doi.org/10.1558/cj.29519

[3] European Commission, 2011. Study on the use of subtitling: The potential of subtitling to encourage language learning and improve the mastery of foreign languages EACEA/2009/01. Available from: https://op.europa.eu/en/publication-detail/-/publication/afc5cf17-02f8-459c-b238-1890ee5cca2b (cited 10 June 2025).

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

[5] Muñoz, C., 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 Learning. Applied Linguistics. 29(4), 578–596. DOI: https://doi.org/10.1093/applin/amm056

[6] Pfenninger, S.E., Singleton, D., 2017. Beyond Age Effects in Instructional L2 Learning: Revisiting the Age Factor. Multilingual Matters: Bristol, UK.

[7] Muñoz, C., 2014. Contrasting effects of starting age and input on the oral performance of foreign language learners. Applied Linguistics. 35(4), 463–482. DOI: https://doi.org/10.1093/applin/amu024

[8] DeKeyser, R.M., 2011. Age effects in second language learning. In: Gass, S., Mackey, A. (eds.). Handbook of Second Language Acquisition. Routledge: London, UK. pp. 442–460.

[9] Unsworth, S., Persson, L., Prins, T., et al., 2015. An investigation of factors affecting early foreign language learning in the Netherlands. Applied Linguistics. 36(5), 527–548. DOI: https://doi.org/10.1093/applin/amt052

[10] Cadierno, T., Hansen, M., Lauridsen, J.T., et al., 2020. Does younger mean better? Age of onset, learning rate and short-term L2 proficiency in Danish young learners of English. Vigo International Journal of Applied Linguistics. 57–86.

[11] 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), 142–167. DOI: https://doi.org/10.1075/itl.00010.pet

[12] European Commission, 2012. First European survey of language competences. Final report. European Commission, Brussels. Available from: https://op.europa.eu/en/publication-detail/-/publication/42ea89dc-373a-4d4f-aa27-9903852cd2e4/language-en/format-PDF/source-116835286 (cited 10 June 2025).

[13] 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. DOI: https://doi.org/10.1017/S1366728918001062

[14] Puimège, E., Peters, E., 2019. Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related factors. Language Learning. 69(4), 943–977. DOI: https://doi.org/10.1111/lang.12364

[15] Muñoz, C., 2020. Boys like games and girls like movies: Age and gender differences in out-of-school contact with English. Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics. 33(1), 171–201. DOI: https://doi.org/10.1075/resla.18042.mun

[16] Hannibal Jensen, S., 2019. Language learning in the wild: A young user perspective. Language Learning & Technology. 23(1), 72–86. DOI: https://doi.org/10125/44673

[17] Sylvén, L.K., 2022. Very young Swedish children's exposure to English outside of school. In: Bardel, C., Hedman, C., Rejman, K., Zetterholm, E. (eds.). Exploring Language Education: Global and Local Perspectives. Stockholm University Press: Stockholm, Sweden. pp. 121–155.

[18] Enever, J., 2011. ELLiE: Early Language Learning in Europe. British Council: Manchester, UK.

[19] 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. DOI: https://doi.org/10.1080/14790718.2012.679275

[20] Bollansée, L., Puimège, E., Peters, E., 2021. "Watch out! Behind you is the enemy!" An exploratory study into the relationship between extramural English and productive vocabulary knowledge. In: Werner, V., Tegge, F. (eds.). Pop Culture in Language Education. Routledge: London, UK. pp. 199–214.

[21] Sundqvist, P., 2019. Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. Language Learning & Technology. 23(1), 87–113. DOI: https://doi.org/10125/44674

[22] Sundqvist, P., Wikström, P., 2015. Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. System. 51, 65–76. DOI: https://doi.org/10.1016/j.system.2015.04.001

[23] 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. DOI: https://doi.org/10.14746/ssllt.2021.11.2.2

[24] Sylvén, L.K., Sundqvist, P., 2012. Gaming as extramural English L2 learning and L2 proficiency among young learners. ReCALL. 24(3), 302–321. DOI: https://doi.org/10.1017/S095834401200016X

[25] Brevik, L.M., 2016. The gaming outliers: Does out-of-school gaming improve boys' reading skills in English as a second language? In: Elstad, E. (eds.). Educational Technology and Polycontextual Bridging. Sense Publishers: Rotterdam, The Netherlands. pp. 39–61.

[26] Meyer, J., Fleckenstein, J., Krüger, M., et al., 2024. Read at home to do well at school: informal reading predicts achievement and motivation in English as a foreign language. Frontiers in Psychology. 14, 1289600. DOI: https://doi.org/10.3389/fpsyg.2023.1289600

[27] Verspoor, M.H., de Bot, K., van Rein, E., 2011. English as a foreign language: The role of out-of-school language input. In: de Houwer, A., Wilton, A. (eds.). English in Europe Today: Sociocultural and Educational Perspectives. John Benjamins Publishers: Amsterdam, The Netherlands. pp. 147–166.

[28] Lyrigkou, C., 2019. Not to be overlooked: agency in informal language contact. Innovation in Language Learning and Teaching. 13(3), 237–252. DOI: https://doi.org/10.1080/17501229.2018.1433182

[29] Peters, E., Noreillie, A.-S., Heylen, K., et al., 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. DOI: https://doi.org/10.1111/lang.12351

[30] De Wilde, V., Brysbaert, M., Eyckmans, J., 2020. Learning English through out-of-school exposure: How do word-related variables and proficiency influence receptive vocabulary learning? Language Learning. 70(2), 349–381. DOI: https://doi.org/10.1111/lang.12380

[31] Council of Europe, 2020. Common European Framework of Reference for Languages: Learning, Teaching, Assessment - Companion Volume. Available from: https://www.coe.int/en/web/common-european-framework-reference-languages (cited 10 June 2025).

[32] Jaekel, N., Schurig, M., van Ackern, I., et al., 2022. The impact of early foreign language learning on language proficiency development from middle to high school. System. 106, 102763. DOI: https://doi.org/10.1016/j.system.2022.102763

[33] Sandlund, E., Sundqvist, P., 2019. Doing versus assessing interactional competence. In: Salaberry, M.R., Kunitz, S. (eds.). Teaching and Testing L2 Interactional Competence: Bridging Theory and Practice. Routledge: New York, NY, USA. pp. 357–396.

[34] Sundqvist, P., Sandlund, E., Skar, G.B., et al., 2020. Effects of rater training on the assessment of L2 English oral proficiency. Nordic Journal of Modern Language Methodology. 8(1), 3–29. DOI: https://doi.org/10.46364/njmlm.v8i1.605

[35] Norton, B., 2013. Identity and Language Learning: Extending the Conversation, 2nd ed. Multilingual Matters: Bristol, UK.

[36] Norton, B., De Costa, P.I., 2017. Research tasks on identity in language learning and teaching. Language Teaching. 51(1), 90–112. DOI: https://doi.org/10.1017/S0261444817000325

[37] Norton, B., Toohey, K., 2011. Identity, language learning, and social change. Language Teaching. 44(4), 412–446. DOI: https://doi.org/10.1017/S0261444811000309

[38] Dörnyei, Z., 2005. The Psychology of the Language Learner. Lawrence Erlbaum: Mahway, NJ, USA.

[39] Dörnyei, Z., 2009. The L2 motivational self system. In: Dörnyei, Z., Ushioda, E. (eds.). Motivation, Language Identity and the L2 Self. Multilingual Matters: Bristol, UK. pp. 9–42.

[40] Dörnyei, Z., Ushioda, E., 2009. Motivation, Language Identity and the L2 Self. Multilingual Matters: Bristol, UK.

[41] Rindal, U., 2025. English in Norway. In: Bolton, K. (eds.). The Wiley Blackwell Encyclopedia of World Englishes. Wiley-Blackwell: Hoboken, NY, USA. pp. 1–13.

[42] Rindal, U., 2010. Constructing identity with L2: Pronunciation and attitudes among Norwegian learners of English. Journal of Sociolinguistics. 14(2), 240–261. DOI: https://doi.org/10.1111/j.1467-9841.2010.00442.x

[43] Rindal, U., Piercy, C., 2013. Being 'neutral'? English pronunciation among Norwegian learners. World Englishes. 32(2), 211–229. DOI: https://doi.org/10.1111/weng.12020

[44] Norwegian Media Authority, 2022. Children and media 2022 – an investigation of 9–18-year-olds' media habits. Available from: https://www.medietilsynet.no/globalassets/publikasjoner/barn-og-medier-undersokelser/2022/231002_barn-og-medier_2022.pdf (cited 3 December 2023).

[45] Salaberry, M.R., Kunitz, S., 2019. Teaching and Testing L2 Interactional Competence: Bridging Theory and Practice. Routledge: New York, NY, USA.

[46] Pekarek Doehler, S., Pochon-Berger, E., 2015. The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In: Cadierno, T., Eskildsen, S. W. (eds.). Usage-Based Perspectives on Second Language Learning. Mouton de Gruyter: Berlin, Germany. pp. 233–268.

[47] Browne, W., Golalizadeh, M., 2021. MLPowSim (1.0) [Computer software]. University of Bristol, Center for Multilevel Modelling: Bristol, UK.

[48] Faul, F., 2020. G*Power (3.1.9.6) [Computer software]. Universität Kiel: Kiel, Germany.

[49] Sundqvist, P., Uztosun, M.S., 2024. Extramural English in Scandinavia and Asia: Scale development, learner engagement, and perceived speaking ability. TESOL Quarterly. 58(4), 1638–1665. DOI: https://doi.org/10.1002/tesq.3296

[50] Swedish National Agency for Education, 2004. European Language Portfolio 6-11 years. Available from: https://www.skolverket.se/download/18.43f8b3a718dd345a2542/1708662853163/Europeiska-sprakportfolion-6-11-2.pdf (cited 23 June 2025).

[51] Snijders, T.A.B., Bosker, R.J., 2012. Multilevel Analysis, 2nd ed. Sage: London, UK.

[52] Anthony, L., Nation, I.S.P., 2017. Picture Vocabulary Size Test (Version 1.2.0) [Computer software and measurement instrument]. Waseda University: Tokyo,

Japan.

[53] Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M., 2012. Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods. 44(4), 978–990. DOI: https://doi.org/10.3758/s13428-012-0210-4

[54] Zhang, X., 2013. The I Don't Know option in the Vocabulary Size Test. TESOL Quarterly. 47(4), 790–811. DOI: https://doi.org/10.1002/tesq.98

[55] Nation, I.S.P., Beglar, D., 2007. A vocabulary size test. The Language Teacher. 31(7), 9–13.

[56] Beglar, D., 2010. A Rasch-based validation of the Vocabulary Size Test. Language Testing. 27(1), 101–118. DOI: https://doi.org/10.1177/0265532209340194

[57] Gyllstad, H., 2012. Validating the Vocabulary Size Test: A classical test theory approach. Poster presentation, The Ninth Annual Conference of EALTA, Innsbruck, Austria, 30 May–3 June 2012.

[58] Gyllstad, H., Vilkaitė, L., Schmitt, N., 2015. Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. ITL-International Journal of Applied Linguistics. 166(2), 278–306. DOI: https://doi.org/10.1075/itl.166.2.04gyl

[59] Stewart, J., 2014. Do multiple-choice options inflate estimates of vocabulary size on the VST? Language Assessment Quarterly. 11(3), 271–282. DOI: https://doi.org/10.1080/15434303.2014.922977

[60] Aus der Wieschen, M.V., 2017. Classroom practices in early foreign language teaching in Denmark: On the role of quantity and quality of exposure to English inside the classroom [PhD Thesis]. University of Southern Denmark: Kolding, Denmark.

[61] Goriot, C., van Hout, R., Broersma, M., et al., 2021. Using the Peabody picture vocabulary test in L2 children and adolescents: effects of L1. International Journal of Bilingual Education and Bilingualism. 24(4), 546–568. DOI: https://doi.org/10.1080/13670050.2018.1494131

[62] Cameron, L., 2001. Teaching Languages to Young Learners. Cambridge University Press: Cambridge, UK.

[63] Güngör, B., Önder, A., 2022. Development of English Picture Vocabulary Test as an assessment tool for very young EFL learners' receptive and expressive language skills. Early Education and Development. 34(2), 572–589. DOI: https://doi.org/10.1080/10409289.2022

[64] Nafs Project, University of Gothenburg, 2025. National tests of foreign languages. Available from: https://www.gu.se/en/national-tests-of-foreign-languages (cited 15 May 2025).

[65] Sundqvist, P., Sandlund, E., 2024. Testing Talk: Ways to Assess Second Language Oral Proficiency. Bloomsbury Academic: London, UK.

[66] Borger, L., 2019. Assessing interactional skills in a paired speaking test: Raters' interpretation of the con-

struct. Apples. 13(1), 151–174. DOI: https://doi.org/10.17011/apples/urn.201903011694

[67] Borger, L., 2019. Evaluating a high-stakes EFL speaking test: Teachers' practices and views. Moderna Språk. 113(1), 25–57. DOI: https://doi.org/10.58221/mosp.v113i1.7621

[68] Sundqvist, P., Wikström, P., Sandlund, E., et al., 2018. The teacher as examiner of L2 oral tests: A challenge to standardization. Language Testing. 35(2), 217–238. DOI: https://doi.org/10.1177/0265532217690782

[69] Ockey, G.J., 2017. Item response theory. In: Fulcher, G., Davidson, F. (eds.). The Routledge Handbook of Language Testing. Routledge: Oxon, UK. pp. 336–349.

[70] Field, A., Miles, J., Field, Z., 2012. Discovering Statistics Using R. Sage: London, UK.

[71] Kelloway, E.K., 2015. Using Mplus for structural equation modeling. Sage: London, UK.

[72] Pinter, A., 2014. Child participant roles in applied linguistics research. Applied Linguistics. 35(2), 168–183. DOI: https://doi.org/10.1093/applin/amt008

[73] Braun, V., Clarke, V., 2022. Thematic Analysis: A Practical Guide. Sage: Los Angeles, CA, USA.

[74] Copland, F., Creese, A., 2015. Linguistic Ethnography: Collecting, Analysing and Presenting Data. Sage: London, UK.

[75] Sidnell, J., Stivers, T., 2013. The Handbook of Conversation Analysis. Wiley-Blackwell: Chichester, UK.

[76] Ducasse, A.M., Brown, A., 2009. Assessing paired orals: Raters' orientation to interaction. Language Testing. 26(3), 423–443. DOI: https://doi.org/10.1177/0265532209104669

[77] Kasper, G., Wagner, J., 2011. A conversation-analytic approach to second language acquisition. In: Atkinson, D. (eds.). Alternative Approaches to Second Language Acquisition. Routledge: New York, NY, USA. pp. 117–142.

[78] Schegloff, E.A., 2007. Sequence Organization in Interaction: A Primer in Conversation Analysis. Cambridge University Press: Cambridge, UK.

[79] Sacks, H., 1992. Lectures on Conversation, Volumes I and II. In: Jefferson, G., Schegloff, E.A. (eds.). Blackwell: Oxford, UK.

[80] ten Have, P., 2007. Doing Conversation Analysis: A Practical Guide. Sage: London, UK.

[81] Sandlund, E., Sundqvist, P., 2016. Equity in L2 English oral assessment: Criterion-based facts or works of fiction? Nordic Journal of English Studies. 15(2), 113–131. DOI: https://doi.org/10.35360/njes.365

[82] Dorman, J.P., 2008. The effect of clustering on statistical tests: an illustration using classroom environment data. Educational Psychology. 28(5), 583–595. DOI: https://doi.org/10.1080/01443410801954201

[83] Dörnyei, Z., 2007. Research Methods in Applied Linguistics. Oxford University Press: Oxford, UK.