TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

COMPLEXITIES OF COMPETENCE: A STUDY ON FINNISH UPPER-SECONDARY SCHOOL STUDENTS' LEXICAL DEVELOPMENT AND USE OF L2 ENGLISH

A Mixed-Methods Examination

Marja-Leena Niitemaa

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 655 | HUMANIORA | TURKU 2024

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

COMPLEXITIES OF COMPETENCE: A STUDY ON FINNISH UPPER-SECONDARY SCHOOL STUDENTS' LEXICAL DEVELOPMENT AND USE OF L2 ENGLISH

A Mixed-Methods Examination

Marja-Leena Niitemaa

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 655 | HUMANIORA | TURKU 2024

University of Turku

Faculty of Humanities
School of Languages and Translation Studies
Department of English
Doctoral Programme in Languages and Translation Studies (Utuling)

Supervised by

Professor Päivi Pietilä
University of Turku
Turku, Finland

Docent Maarit Mutta
University of Turku
Turku, Finland

Professor Pekka Lintunen
University of Turku
Turku, Finland

Reviewed by

Professor Katalin Doró
University of Szeged
Szeged, Hungary

Adjunct Professor Katja Mäntylä
University of Jyväskylä
Jyväskylä, Finland

Opponent

Adjunct Professor Katja Mäntylä
University of Jyväskylä
Jyväskylä, Finland

ISBN 978-951-29-9597-4 (PRINT)
ISBN 978-951-29-9598-1 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Painosalama, Turku, Finland 2024

UNIVERSITY OF TURKU
Faculty of Humanities
School of Languages and Translation Studies
Department of English
MARJA-LEENA NIITEMAA: Complexities of Competence: A Study on Finnish Upper-Secondary School Students' Lexical Development and Use of L2 English. A Mixed-Methods Examination
Doctoral Dissertation, 152 pp.
Doctoral Programme in Languages and Translation Studies (Utuling)
January 2024

ABSTRACT

The dissertation examines Finnish upper-secondary school second language use from four perspectives: the development of lexical knowledge in relation to extramural activities (Article I); the role of lexical recognition in searching online sources for words and information (Article II); digitally identified cohesive features in writing (Article III); and lexical sophistication as examined by traditional lexical tests and digital analyses (Article IV). The study is situated in the context of second language acquisition with implications for teaching, and assessment building on previous research by examining lexical recognition skills across the high-, mid-, and low-frequency bands, triangulating information from questionnaires and video-observations with data elicited from traditional lexical tests and digital analyses (TAACO and TAALES). The studies in Articles I and IV were conducted longitudinally, while Articles II and III focus on examinations conducted in the second year.

Article I demonstrates that using English in cognitively demanding extramural activities, such as reading and gaming, develops both overall lexical recognition skills and recognition of infrequent lexis. In the second year, these activities collectively explained 45% of the variance in the scores for infrequent lexis. Article II focused on using online sources in diverse indirect writing tasks. The results indicated that consulting online dictionaries and informational web-pages required rapid lexical recognition, multiple reading strategies and adequate digital skills to formulate queries and evaluate the search results, and that these abilities were directly associated with the participant's lexical recognition skills. According to the analyses, participants scoring less than 60–64% in the lexical recognition test (the VLT), did not benefit from consulting online dictionaries and other sources in diverse writing tasks.

In Articles III and IV, writing skills were also examined using digital analyses. Article III investigated cohesive devices in essays written in the second year. Digital analyses (TAACO) showed that using diverse referential devices, e.g., adverbs and connectors, explained 37% of the variance in the essay ratings. In Article IV, written production was examined longitudinally using traditional lexical tests and a digital tool (TAALES). Digital analyses combining three indices (function word frequency, infrequent content words and typical English two-word combinations) collectively

explained 46% in the variance of the first-essay scores and 44% of the variance in the second-essay scores. Corresponding traditional tests on associative word knowledge and recognition of two-word combinations explained 56% in the variance in the first essay scores and 61% of the second year scores. Thus, the results elicited by traditional tests were in line with the digital results. This finding implies that traditional tests for recognition and associative skills can also be used to assess lexical sophistication in language teaching.

KEYWORDS: assessment, associative knowledge, cohesion, digital analyses, essays, lexical knowledge, L2 English, mixed-methods, online sources, TAACO, TAALES

TURUN YLIOPISTO
Humanistinen tiedekunta
Kieli- ja käännöstieteiden laitos
Englannin kieli
MARJA-LEENA NIITEMAA: Complexities of competence: A study on Finnish upper-secondary school students' lexical development and use of L2 English. A mixed-methods examination
Väitöskirja, 152 s.
Tohtoriohjelma Utuling
Tammikuu 2024

TIIVISTELMÄ

Väitöskirjani tarkastelee suomalaisten lukio-opiskelijoiden englannin kielen osaamista neljästä eri näkökulmasta: miten sanoja omaksutaan vapaa-ajan harrastuksissa (Artikkeli I); kuinka hyvää sanaston tunnistamistaitoa tarvitaan digitaalisissa ympäristöissä (Artikkeli II); mitä kielellisiä keinoja käytetään johdonmukaisesti etenevän ja lukijaystävällisen tekstin tuottamiseen digitaalisen analyysin perusteella (Artikkeli III); miten monipuolista ja tekstilajiin sopivaa sanastoa kirjoitelmat sisältävät perinteisten sanastotestien ja digitaalisen analyysin valossa (Artikkeli IV). Osana vieraan kielen omaksumisen tutkimusta väitöskirja tarjoaa sovelluksia opetukseen ja arviointiin ja laajentaa tietoa englannin kielen osaamisesta testaamalla myös harvinaisten sanojen tunnistamista yhdistäen määrällisiä ja laadullisia menetelmiä sekä vertailemalla digitaalisten analyysien (TAACO ja TAALES) sekä perinteisten testien tuloksia. Artikkelit I ja IV ovat pitkittäistutkimuksia, kun taas Artikkelit II ja III tarkastelevat testisuorituksia lukion kakkosvuonna.

Artikkeli I osoittaa, että kognitiivisesti vaativat harrastukset kuten englannin käyttö lukiessa ja videopelejä pelatessa kehittävät harvinaisten sanojen merkitysten tunnistamista. Tilastollisessa analyysissä nämä harrastukset selittivät 45 % harvinaisten sanojen tunnistamisen varianssista toisena opiskeluvuonna. Artikkeli II todistaa, että verkkosanakirjojen käyttö ja tiedonhaku Internetissä vaativat käyttäjältä nopeaa merkitysten ymmärtämistä, monipuolisia luku- ja tiedonhakutaitoja sekä kykyä arvioida hakutuloksia kriittisesti. Analyysien mukaan sanaston tunnistustestissä (VLT) pitäisi onnistumisprosentin olla 60 %–64 %, jotta sanakirjojen ja lähteiden käyttö tuottaisi tulosta.

Artikkeleissa III ja IV tutkittiin lukiolaisten kirjoitustaitoja englannin kielessä myös digitaalisten analyysien avulla. Artikkeli III tarkasteli tyypillisiä koheesiokeinoja englannin kirjoitelmissa toisena opiskeluvuonna. Digitaalinen analyysi (TAACO) osoitti, että taitavasti laadittujen tekstien sidosteisuus perustui pääasiassa adverbien ja konjunktioiden monipuoliseen käyttöön eli kykyyn organisoida tekstiä. Nämä piirteet selittivät 37 % kirjoitelmien arvosanojen varianssista. Artikkeli IV tarkasteli kahtena vuonna laadittuja kirjoitelmia sekä perinteisin testein että digitaalisen työkalun avulla (TAALES). Regressioanalyysissa kolmen piirteen yhdistelmä (adverbit, apuverbit, konjunktiot; harvinaiset sisältösanat; tavalliset kollokaatiot) selitti 46 % ensimmäisten kirjoitelmien arvosanojen varianssista ja

toisena vuonna 44 %. Verrattaessa kirjoitelmien arvosanoja vastaavien perinteisten testien tuloksiin, assosiaatioiden tuottaminen ja sanayhdistelmien tunnistus selittivät 56 % kirjoitelmien arvosanojen varianssista ensimmäisenä vuonna ja 61 % toisena vuonna. Digitaalisten ja perinteisten analyysien tulokset olivat siis samansuuntaiset. Tulokset osoittavat, että perinteiset sanaston tunnistus- ja yhdistelytehtävät sopivat myös vaativan sanaston käytön arviointiin kielenopetuksessa.

ASIASANAT: arviointi, assosiaatio, digitaalinen analyysi, englanti vieraana kielenä, internet, koheesio, monimenetelmätutkimus, sanaston oppiminen, TAACO, TAALES

Acknowledgements

During the long period of writing this dissertation, I have received valuable help, support and encouragement from so many people. Firstly, I would like to express my gratitude to the external examiners of my PhD study, Associate Professor Katalín Doró from University of Szeged, Hungary, and Adjunct Professor Katja Mäntylä from University of Jyväskylä, Finland. I am deeply indebted to both of them for their detailed and encouraging reviews as well as the suggestions to clarify my goals and argumentation during the finalization process.

I have also been fortunate to work with excellent supervisors. My first supervisor was Professor Päivi Pietilä, who was the first person to suggest starting PhD studies. She helped me to start the dissertation and also encouraged me to give talks about my research in two ESSE and AFinLa conferences. In the department, Professor Pietilä organized an SLA colloquium, in which we, PhD students, had a chance to talk about our projects, get feedback on article drafts and widen our perspectives on second language acquisition. This was particularly important for the few of us who worked as teachers outside the English department. Professor Pietilä tragically passed away when I was finalizing my second article. I shall never forget our last meeting at the department. She had invited me to her office to discuss the new version of my manuscript. After a lengthy conversation, she photocopied her comments for me. As it was quite late in the evening, I encouraged her to go home to get some rest. She stayed to clean her desk. I left home. I never saw her again.

Although I had lost a mentor, I was fortunate to get two new experts to supervise my research, Dr. Pekka Lintunen and Dr. Maarit Mutta. Since then they have pushed me through the writing process by sending comments on my texts by email and arranging meetings at the department. All this they have done in a very determined but gentle way. Although busy with teaching, assessing, organizing, and carrying out their own research, they have always had time for my questions. I am also grateful to Dr. Lydia Kokkola from University of Oulu for commenting my second article at the time she was teaching at University of Turku. Moreover, I want to thank Dr. Heini-Marja Pakula for organizing teacher trainees to assess the essays analysed in my studies. I am also indebted to Jouko Katajisto for advice and consultation sessions regarding statistical analyses particularly at the start of my research. In

addition, I want to thank Meri Heinonen and Tarja Ahonen for their help. When it comes to the rules and procedures in the administrative system, they know all that is worth knowing in organizing practical matters.

While conducting my PhD study, I have also received valuable assistance from numerous people outside of Academia. I am most grateful to the school principal for the permission to organize data collection at the already overcrowded school premises and overbooked computer rooms. Without the students who took part in all the tests and permitted me to video-record their working process conducting this research would have been impossible. Lastly, I want to thank DKG, International Society for Key Women Educators, for financial support to buy a decent computer when I started to work on my PhD.

My family members, friends and former colleagues have patiently supported my efforts and encouraged me to carry on throughout the PhD process. I also want to thank Mikko-Ville Niitemaa for finalizing the figure in the study.

Turku, December 2023
Marja-Leena Niitemaa

Table of Contents

Tables

Figures

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Niitemaa, M-L. (2020). Informal acquisition of L2 English. Exploring the relationship between online out-of-school exposure and words at different frequency levels. *Nordic Journal of Digital Literacy*, *15*, 89–105. https://doi.org/10.18261/issn.1891-943x-2020-02-02

II Niitemaa, M-L., & Pietilä, P. (2018). Vocabulary skills and online dictionaries: A study on EFL learners' receptive vocabulary knowledge and success in searching electronic sources for information. *Journal of Language Teaching and Research*, *9*, 453–462. DOI: http://dx.doi.org/10.17507/jltr.0903.02

III Niitemaa, M-L. (2022). Cohesion in Finnish EFL essays: Digital analyses and observations on the use of online sources. *EduLingua*, *8*, 1–16. https://DOI:10.14232/edulingua.2022.1.1

IV Niitemaa, M-L. Lexical sophistication in Finnish upper-secondary L2 school learners' essays: A study on receptive and associative lexical development. Manuscript.

1 Aims of the Study

The general aim of this dissertation is to examine the development of Finnish upper-secondary school students' lexical competence in L2 English in diverse contexts. The first article concentrates on the expansion of lexical knowledge via free-time use of English in extramural activities. The second article focuses on the importance of receptive vocabulary size in searching online sources for words and information. The last two articles investigate receptive and associative lexical knowledge as a prerequisite for the emergence of higher-order productive skills; the third article focuses on lexical recognition skills and the ability to produce cohesive text and the fourth article analyses receptive vocabulary in relation to lexical sophistication.

1.1 Background of the study

Finnish upper-secondary school students are expected to have reached the CEFR level B2.1 in L2 English before taking the high-stakes national school-leaving examination. The assessment criteria follow the guidelines of The Common European Framework of Reference for Languages (2018a). Regarding L2 English, CEFR emphasizes differentiating between levels of formality, adapting language use according to the circumstances and avoiding errors that hinder communication by causing misunderstandings. Such abilities develop late at the upper-intermediate (B2) level (e.g., Chen & Baker, 2016), and to reach the advanced level (C1), writers need to possess a broad lexical repertoire including common idiomatic expressions and lexical combinations. This goal, however, requires far more encounters than language classes can provide, but is attainable with a combination of formal teaching and extramural activities (e.g., Ellis, 2017). As a lot of word learning takes place in informal contexts today, an important pedagogical task is to convince both students and teachers that in-class and out-of-class practices support one another.

The present study is based on the hypothesis that the expansion of receptive word knowledge has a cumulative interactive effect on all lexical competences, as illustrated by the Figure on page 14. When encountering English, we pick up new words; appropriating more words facilitates language use in reading and writing; when reading diverse texts, we are exposed to cohesion, advanced words and diverse registers; when writing, we deepen and strengthen associative lexical knowledge.

Regarding methodology, the study combines traditional lexical tests, e.g., the Vocabulary Levels Test (Schmitt et al., 2001) and Lex30 (Fitzpatrick & Meara, 2009) with digital analyses to detect cohesive devices (TAACO 1.5, Crossley et al., 2016a) and identify lexically sophisticated features (TAALES 2.2, Kyle et al., 2017). In addition, video-recordings are used to observe how participants with diverse lexical skills can use online dictionaries and informational sources in simulations of real-world tasks and essay writing.

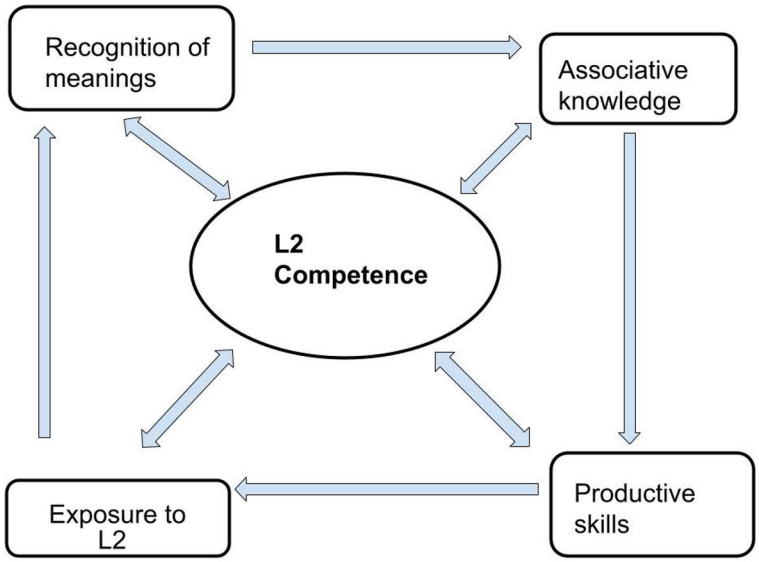

Figure. Interacting factors affecting L2 competence.

1.2 Language learning context in Finland

In Finland, comprehensive schools provide nine years of general basic education with six years at primary level and three years in middle school (Ministry of Education and Culture; https://minedu.fi/en/comprehensive-school). The curricula are based on the premise that language learning promotes thinking and communication skills as well as cultural literacy (Finnish National Agency for Education; https://www.oph.fi/english). Since the curriculum reform in 2020, children start learning a foreign language (L2) in the first grade of the primary school. English is the most widely studied L2 in the country. In larger communities, schools can also offer French, German, Russian or Spanish as the first foreign language (L2) depending on their resources. In the fourth grade, pupils can choose another foreign language as an optional subject. According to legislation, Finland has two national languages, Finnish and Swedish. In the middle school from sixth

grade onwards, Swedish is compulsory for those who speak Finnish as their first language (L1), and vice versa, Swedish speaking children study Finnish. In the eighth grade, the pupils have another chance to start studying French, German, Russian or Spanish as a shorter course.

After comprehensive school, pupils may apply for upper-secondary school, where they continue studying the second and foreign languages, and at least in larger schools they can also choose an additional foreign language. After finishing upper-secondary school, students take a matriculation examination, i.e., a national school-leaving test in five subjects including L2. These high-stakes examinations have been completed on computers since 2018. The tests are taken on a Linux operating system so that candidates can access only the applications and materials that are installed on the system (https://www.ylioppilastutkinto.fi/en/matriculation-examination/digitalmatriculation-examination). The results of the national examination are considered when students apply to universities, colleges, or universities of applied sciences.

In the national school-leaving test, the examination of English as a foreign language measures listening and reading comprehension, grammar, vocabulary, and writing proficiency. In the writing test, the examinees write a composition of 700–1300 characters on one of four given topics. In addition to traditional text types like description, discussion or argumentation, the topics may include digital genres such as various media postings. The topics are sometimes linked to extra material such as videos, images or song lyrics, which the test-taker may or may not employ. Despite the digital format, written production is still assessed by human raters, first by language teachers at schools and then by experts of the Matriculation Examination Board. The assessment criteria are currently being updated. The essays under scrutiny in the present study were assessed on the content and structure of the text, lexical richness and accuracy, and the candidate's ability to communicate the message clearly. More details on the scale are provided in subsection 3.4.2.

1.3 Research aims and outline of the study

Researchers commonly suggest that receptive lexical knowledge associates with L2 learners' ability to use English in multiple ways (e.g., Alderson, 2005; Schmitt et al., 2001). Therefore, recognition of word meanings is regarded as an important factor in the four sub-studies (c.f., Table 1). The skill is measured by the Vocabulary Levels Test (Schmitt et al., 2001; Nation, 1983), henceforward referred to as the VLT. To provide more nuanced information on recognition skills, the analysis is extended to low-frequency lexis, whereas most previous research among the same age group focuses on high- and mid-frequency lexis.

The first study is a longitudinal examination of the impact of extramural digital activities on the VLT scores. The second study analyses the connection of the VLT and effective use of online sources in tasks simulating real-world conditions. The examination was conducted during the first year at upper-secondary school. The third article on cohesive devices focuses on participants' written production during the second year. Cohesive features were identified digitally and the ability to produce cohesive text was analysed against the VLT results. The fourth study examines the development of lexical sophistication in writing over two years. The general aim of the present study is to examine Finnish upper-secondary school students' lexical competence in English as a second language (L2) from four perspectives: how lexical knowledge develops through digital out-of-school activities (Article I); the role of receptive lexical knowledge in searching online sources for words and information (Articles II–IV); the cohesive devices upper-secondary school students use and how such features relate to writing quality (Article III); lexical sophistication in writing quality; comparing commonalities between digital analyses and findings from traditional tests (Article IV).

Regarding terminology, L2 English refers to English as a foreign language in the Finnish context, while EFL is used when it appears in the source texts. In the following, theoretical perspectives on lexical competence are discussed in Chapter 2, the methods are introduced in Chapter 3, and the results in Chapter 4.

Article I

The first study focused on the role of out-of-school activities in advancing lexical knowledge (N = 46). Earlier research has shown that encounters with English include multiple elements that promote vocabulary learning, e.g., comprehensible input (Krashen, 1985), language production (Swain, 1985), noticing (Schmidt, 1990), task-induced involvement (Laufer & Hulstijn, 2001), exposure to and frequent repetition of vocabulary (Eckerth & Tavakoli, 2012), motivation (Dörnyei & Chan, 2013), and self-selected topics and meaningful content (Lee & Pulido, 2017). The participants' extramural activities were surveyed longitudinally using questionnaires (see subsection 3.2). Moreover, participants' self-reported experiences of vocabulary acquisition were surveyed. The lexical recognition skills were measured twice by the VLT and the scores across high-, mid-, and low-frequency bands were examined in relation to the type, frequency and English content of the online activities.

Article II

The second study investigated how the participants (N = 22) employed online reference sources to search lexis and information when the choice of the sources was

not controlled. The purpose was to examine the connection between lexical recognition skills and the ability to find words and facts. Simulating real-world tasks, the tests required inferencing skills to comprehend the texts as a whole (see subsection 3.4.1 for the tasks and rating). Consulting dictionaries and informational sites was part of a cognitively complex reading process, as the tasks involved recognizing word meanings, information processing and problem-solving, i.e., competences that are essential to reading comprehension in general (e.g., Tono, 2011). The video-recorded data on the working process revealed which reference sources were consulted, how the students searched for words and what they found. The findings were then analysed in relation to the VLT results.

Article III

The third article concentrated on cohesive features in upper-secondary school L2 learners' essays (N = 46) written on a familiar topic (see subsection 3.4.2 for the essays). The aim was to find out which cohesive devices the writers employed, how these devices related to holistic human-rated essay scores and the vocabulary recognition skills (the VLT), and whether the essays had achieved the CEFR expectations for cohesion at level B2.1 (Appendix B). Cohesive features were first identified using a natural language processing tool, TAACO 2.0.4 (Crossley, et al., 2019), and the findings were then correlated with the human-rated essay scores. The digital tool is introduced in subsection 3.6.1. The writing process of 31 participants was video-recorded. The observational data were used to examine whether or how accessing online reference sources helped the participants to increase cohesion during writing.

Article IV

The study examined lexical development in upper-secondary school L2 English essays (N = 46) over two years triangulating traditional lexical vocabulary tests with digital analyses of lexical sophistication (TAALES 2.2; Kyle, et al., 2017). The study investigated how human-rated essay scores related to the results of the traditional descriptive tests analysed the essay scores in relation to digital data and compared the results between traditional tests and digital methods. Recognition of familiar words and lexical combinations increased over time, whereas associative skills and recognizing infrequent words improved mor slowly. TAALES analyses ascertained that higher-scoring essays included diverse sophisticated lexis, typical English word combinations and a wide range of function words. These features collectively explained 46% of the variance of the first-year essay scores and 44% in the second-year. However, infrequent lexis was used appropriately only in higher-scoring texts and TAALES cannot judge whether lexis is used properly. Overall, TAALES indices correlated positively with

features that are also appreciated by human raters such as advanced vocabulary, a wide range of function words and typical English two-word combinations.

Table 1. The main foci, research questions and contributions of the sub-studies.

The main foci	Article I	Article II	Article III	Article IV	Contributions
Out-of-school online activities	RQ1. To what extent are Finnish upper-secondary school students engaged in out-of-school online activities?	RQ2. What are the participants' self-reported experiences of online out-of-school vocabulary acquisition?			Longitudinal examination across the high-, mid-, and low-frequency bands among upper-secondary learners.
The connection between online activities and receptive vocabulary size	RQ3. What is the connection between out-of-school online activities and recognition of English words at high-, mid- and low-frequency levels?				Longitudinal surveys on activities. Self-reports on learning new words.
Searching online sources for words and information	RQ2. What is the relationship between word recognition skills and successful use of online sources?	RQ1. What online dictionaries and informational sites do the participants use? In what way and how successfully do they use them?	RQ2. Under what conditions can using digital sources enhance cohesion?		Simulating real-world conditions.
Text production in relation to cohesion and lexical sophistication			RQ1. Which cohesive features characterize Finnish upper-secondary school essays? How do the essays fulfill the CEFR expectations at level B2?	RQ1. How do human-rated essay scores relate to the results of receptive and associative lexical tests?	Receptive lexical skills are examined longitudinally in relation to higher-order language skills.
Comparison of methods				RQ 2. How do human-rated essay scores relate to digitally identified features of lexical sophistication?	Comparing results elicited by traditional tests and digital analyses. Longitudinal examination.

2 Theoretical Perspectives on Lexical Competence

Researchers commonly acknowledge that all L2 competences increase along with lexical development (e.g., Henriksen, 1999, 2006; Meara, 2009; Read, 2000). Accordingly, the learner's receptive vocabulary size is regarded as one of the strongest positive correlates of text comprehension (e.g., Alderson, 2005; Nation, 2006; Schmitt et al., 2011) and writing quality (e.g., Allen et al., 2014). In the DIALANG context[1], Alderson (2005) suggests that learners' receptive vocabulary size is directly connected not only with writing ability in general but also with textual organization and cohesion in essays. Similarly, Crossley et al. (2016b) argue that lexical development likely indicates knowing how to use cohesive devices. Moreover, recent research on word recognition indicates that receptive lexical knowledge relates to multiple features of lexical sophistication (Berger et al., 2019; Hashimoto & Egbert, 2019). However, L2 development is also affected by multiple extra-linguistic elements, such as the learner's age, aptitude, cultural context, learning styles, motivation, opportunities to study, peer-pressure, socio-economic background, and opportunities and willingness to use the language (e.g., Marek & Wu, 2014).

Regarding theoretical representations of lexical competence, some models are compact, such as Henriksen's three-dimensional model (1999, 311–314) consisting of partial–precise knowledge, depth of knowledge and receptive–productive knowledge, or Meara's model with two components, vocabulary size and organization (Meara, 1996, 35–53). Aitchison's (1994) model was originally meant to describe the L1 learning process in childhood: an object or thing being connected to a word (labelling); the ability to connect a word to other words (packaging); and how a wide range of connected words develops into a system (network building). Large-scale models have been developed, e.g., by Nation (2001), Ringbom (1983, 1991) and Richards (1976). These models indicate that lexical knowledge is an

[1] DIALANG lexical proficiency tests (https://dialangweb.lancaster.ac.uk/ are based on the proficiency scales of the Common European Framework of Reference (2001). The tests provide freely available diagnostics in reading, writing, listening, vocabulary, and grammar in several European languages enabling independent testing at home.

extremely complex phenomenon, which entails not only meaning recognition but also knowing the constraints on the use of the word and the different meanings the word covers. Nation's model is based on three principal components, the form, meaning and use of a word, each of which is divided into subcategories corresponding to spoken–written and receptive–productive skills.

Table 2. Richards' and Ringbom's models of knowing a word.

	Richards	**Ringbom**
1	L1 vocabulary constantly expands through various activities whereas there is little development of syntax	
2	Recognizing that words are not equally frequent and that words are associated with other words	The word is accessible within a specific context only –> The word is accessible regardless of context
3	Understanding geographical, sociolinguistic or temporal variation in meanings	
4	Syntactic behaviour of a word	Knows no syntactic constraints –> Knows some constraints –> Knows all syntactic constraints
5	A word has a basic form and derivations can be made from of it	Knows no collocational constraints –> Knows some constraints –> Knows all collocational constraints
6	Recognition of other characteristics of a word, e.g., variation and register	Knows one form of a word –> Knows the possible derivations of a word
7	Awareness of the basic semantic features of a word (e.g., animate, non-animate)	
8	Word meanings are defined through their relationships with other words, e.g., synonym / antonym, superordinate / coordinate, proper / improper	Knows no associative constraints–> Knows some constraints –> Knows all associative constraints
9	Knowledge of the most frequent ways in which a word realizes a particular concept. Awareness of polysemy	Knows one meaning only –> Knows approximate meanings –> Knows all possible meanings

In Richards' and Ringbom's models, lexical knowledge is regarded as a process in which lexical skills develop individually at a different pace. As shown in Table 2, these models emphasize the same developmental traits including the linguistic, psycholinguistic, and sociolinguistic aspects of word knowledge. Moreover, the models underline the role of associative knowledge, which aligns with the current understanding of lexis as a changing network system (Henriksen, 2006; Larsen-Freeman, 2006; Meara, 2009; Sigman & Cecchi, 2002; Wolter, 2006; Wray, 2008). Overall, compared to more concise representations, models with multiple traits may

be more helpful when trying to understand lexical use in individual texts at different developmental stages.

In Table 2, the first assumption is targeted at L1 speakers. However, the idea of life-long learning also applies to L2 learners (c.f., Sundqvist, 2022) aligning with the principles of the Common European Framework of Reference (2001). Otherwise the two models share the same conceptions of language learning as a dynamic individual process.

Resonating with Aitchison's packaging, the second assumption refers to the developmental phase when the learner starts to understand that words are associated with certain other words. This assumption can also be read as an introduction to the further ones, which either elaborate on frequency, or associative knowledge in a network system. The third assumption concerns awareness of diverse constraints of meanings, such as variation in temporal, geographical, and sociolinguistic aspects: a different word can be used to name the same object, e.g., a *tap* is a British synonym for the American *faucet*; a *mirror* sounds more modern than a *looking glass*; and *Hi*! and *Good morning*! represent different registers. The fourth assumption emphasizes understanding how words behave in clauses and sentences, e.g., understanding that some verbs are only used transitively or that certain nouns do not take the s-plural. The fifth assumption entails knowing that a word has a basic form and derivations can be made from it, and that the derivations form a word family in which the members may not be equally frequent. For example, *school* comes from the first thousand frequency band, *scholar* and *scholarly* from the third thousand band, *scholarship* from the fourth thousand band and *scholastic* from the ninth thousand band. The sixth assumption suggests that lexical knowledge is an associative network. Meanings are defined through relationships with other words, such as contrast (*wet – dry*), similarity (*flower – blossom*), subordination (*cat – animal*) or superordination (*spinach – vegetable*). The seventh assumption is about knowing the semantic features of a word. This entails understanding distinctions between *animate – inanimate*, *human – non-human*, *synonym – antonym*, *superordinate – coordinate* or *proper – improper* on a sociolinguistic scale. The eighth assumption concerns being aware of polysemy, i.e., a word can have multiple different meanings. For example, the noun *subject* has several diverse meanings such as topic, theme, substance and matter.

Richards and Ringbom regard lexical competence as a complex developmental process from receptive word knowledge to the ability to use depth-related lexical features. In other words, appropriating lexis is more about system-building than memorizing new words. Another emphasis is on the importance of understanding polysemy, a phenomenon causing problems for multiple L2 learners who believe that the meaning they know is the only one (e.g., Chan, 2014). In line with these frameworks, later research assumes that extensive receptive vocabulary knowledge is connected to deeper lexical knowledge. Research finding show that learners who

recognize more words also tend to know more about these words (e.g., Perfetti, 2007), and that features of vocabulary size and depth are interrelated (Meara, 2006; Milton, 2009). Regarding other contributions to practice and theory, Richards suggested that vocabulary expansion should be the major target of L2 curricula (1976). He also reminded that extralinguistic factors, such as the strong emotional associations evoked by certain words may affect learning. Ringbom was interested in the word learning process as building a lexical network (1983), and the effects of a non-Indo-European L1, Finnish, on the learning process of a Germanic language like L2 English (1991).

2.1 Lexical competence as recognition of word meanings

Recognition of word meanings is a central feature of lexical competence. According to Meara (2009, 29–39), the larger the receptive vocabulary is in relation to the productive lexis the more useful it becomes for productive purposes. Meara posits that there will always be a gap between receptive and productive vocabulary sizes and the gap becomes wider as receptive vocabulary grows, suggesting that L2 students likely employ from 50 to 75 percent of the lexis they recognise. However, the relationship between receptive and productive vocabulary is not linear. The former may grow in spurts while the latter develops individually between the spurts. Regarding the ratio between receptive and productive vocabulary. In the same vein, Lowie et al. (2011, 113) found that a certain amount of receptive vocabulary would be a prerequisite for productive skills as receptive knowledge supports lexical recall, recall supports controlled production, and recall supports controlled production, which in turn supports free production.

In the present dissertation, recognition of word meanings is one of the central variables. Lexical recognition is examined longitudinally across high-, mid- and low-frequency bands in Articles I and IV. In each sub-study, lexical recognition is considered a principal factor explaining individual variation, for example, in searching electronic dictionaries for words, cohesion building or using lexically sophisticated vocabulary. Articles III and IV examine associative lexical knowledge in relation to cohesion and lexical sophistication relating to the seventh, eighth and ninth assumptions.

2.2 Lexical competence as associative knowledge

In Table 2, the second, third, seventh and eighth assumptions focus on associative knowledge, which relates to the depth of lexical competence needed in productive tasks. The sixth assumption relates to word families, i.e., knowing how to make

derivations from the basic word form. The ninth assumption emphasizes awareness of polysemy. It was observed that words with multiple meanings can easily complicate dictionary use, as exemplified in Article II. Articles III and IV concentrate on associative lexical knowledge in relation to cohesion and lexical sophistication relating to the seventh, eighth and ninth assumptions.

Richards and Ringbom were among the first researchers who treated associative knowledge as a network system. Instead of logical relationships, words come to mind according to different associative links, and retrieving words from the learner's mental lexicon is easier when such links are strong. Scholars in this field suggest that the size and the organisation of a lexical network affect the learner's overall language proficiency (Henriksen, 2006; Larsen-Freeman, 2006; Meara, 2009; Wray, 2008). When the network is large and densely organized, the learner can appropriate new words more easily and retrieve lexis from memory for language production.

Lexical networks are found to follow the pattern of a small-world type of network, which are individual, self-organising and nonlinearly developing, and based on associative connections (Sigman & Cecchi, 2002). In such systems there are only a few nodes, i.e., words with a large number of connections. This means that the central nodes are high-frequency words, which attract new connections to cluster around the existing lexis. Further, the network reorganizes itself after major interventions. Sigman and Cecchi (2002) reported a massive reorganization after adding polysemous words to the network. Wolter (2006) suggests that knowledge of multi-word units indicates development of the lexical network. According to Wray (2008), information in the learner's lexicon is stored iteratively in multiple forms, such as morphemes, separate words, and multi-word sequences. In a densely linked lexicon, all information is interconnected, which makes word meanings easily accessible. In this dissertation, two studies concentrate on associative lexical knowledge. Article III investigates cohesive features in writing and examines, e.g., the writer's ability to use semantically related words instead of repeating the same lexical items. Article IV analyses the learner texts for diverse types of advanced lexis, i.e., lexical sophistication.

2.3 Lexical competence and reading comprehension

The current view on literacy emphasizes the importance of understanding written information in multiple real-life situations. Literacy skills entail critical reading of offline and online texts either independently or with reference sources (Council of Europe, 2018a, 2018b). At the CEFR scale, B2 learners are expected to know how to change reading styles according to the purpose of the task, e.g., scan quickly through articles, websites, and longer texts to locate relevant details, or read

magazines, novels, and lyrics for pleasure. When encountering unknown words, they can infer the meanings from the context. The Finnish curricula also encourage students to employ digital technology to search for information and study online (Finnish National Agency of Education, 2017).

Readers encounter multiple texts in the digital mode. Thus, it is assumed that online reading comprehension depends largely on the same factors as reading offline, e.g., lexical knowledge, the purpose of reading, how accurate comprehension is needed, and various reader characteristics, such as motivation to carry out the task, working memory, the ability to make inferences and metacognitive skills (e.g., Alderson, 2005; Coiro, 2011a; Schmitt et al., 2011). However, there are differences between the two reading conditions. Printed text is static and linear, whereas online information is changing, interactive, multi-layered, and multimodal (Leu et al., 2013). Therefore, reading on the Internet may require even more complex versions of traditional literacies and additional skills such as ability to generate effective queries and evaluate the reliability of the information (e.g., Coiro & Dobler, 2007). Furthermore, when online, readers independently choose the sites they find useful, employ diverse sources and read different amounts of text to reach the same goal.

Online inquiry requires several distinctive reading modes: reading to understand the problem posed in the task, reading to locate information, reading to evaluate the findings, and reading to combine information from various sources. All these components are vital. Without comprehending the task learners cannot decide what to search for; without locating task-relevant information they cannot answer the questions or fulfil the tasks (Leu et al., 2011, 6–8; Leu et al., 2013, 1165); without critically evaluating the options in the dictionary entry**,** learners may choose words that do not fit the context, and without the ability to combine information from different sources, their text comprehension remains shallow (Pelttari & Mutta, 2014). These requirements align with Richards' (1976) model of knowing a word: being aware of limitations and the variability of meanings, syntactic function, and polysemy. Whether offline or online, effortless access to word meanings makes reading fluent and facilitates comprehending the content (Alderson, 2005). In the present dissertation, Article II examines upper-secondary school learners' ability to read for information online, in particular, searching for words in electronic dictionaries (see subsection 4.2), and Articles III and IV provide examples of dictionary use during the writing process.

2.4 Lexical competence and written production

According to the CEFR descriptions on writing proficiency at B2 level, L2 writers are expected to produce clear, official, and semi-official texts in multiple contexts, e.g., write detailed descriptions on subjects related to their individual interests, and

compose a review of a film or a book expressing a personal view on the creative works. They should also know how to synthesise and evaluate information and arguments, interpret information from diagrams and visually organised data in writing, explain the viewpoints articulated in a complex text as well as summarise an L1 text in L2. Regarding written interaction, B2 students should learn to avoid errors causing misunderstandings and employ cohesive devices to help the reader understand connections between sentences and paragraphs. However, some less appropriate expressions and an occasional lack of cohesion are allowed, as L2 texts are not expected to be as expressive and idiomatic as texts composed by L1 students (CEFR, 2018, see p. 75 for written production; pp. 134–136 for vocabulary, and p. 142 for cohesion).

Differences between receptive and productive vocabulary sizes are difficult to examine due to the paucity of valid tests. Using translation tests, Webb (2008) found that the difference between receptive and productive vocabulary sizes is small for common words at the first thousand frequency band but noticed that the gap widens progressively at the second and third thousand frequency bands. Other findings have indicated that receptive vocabulary and individual differences increase simultaneously, and that lexical knowledge develops from word recognition to recall, from recall to controlled production and then from a high value of controlled production to free production (Lowie et al., 2011, 116). Research has detected various factors explaining variation in writing proficiency. In addition to lexical knowledge, writing quality correlates positively with stronger reading skills (Alderson, 2005; Coiro & Dobler, 2007) and writing-specific knowledge (Saddler & Graham, 2007). Difficulties with typing and problems with using online reference sources also cause problems in writing (Niitemaa & Pietilä, 2018).

Regarding Richards' framework, a proficient writer should be able to recognize whether a word is constrained by register or sociolinguistic restrictions, know how to create derivations of words, avoid repetition by choosing semantically related options, and have rapid access to meanings of polysemous words as well as multiple extra-linguistic factors such as greater flexibility (e.g., Allen et al., 2014). Researchers working in the writing-to-learn strand (Manchón, 2012) have shown that writing tasks can promote such deeper lexical knowledge (Mäntylä et al., 2020). Requirements for vocabulary also depend on the text type, as the use of lexis depends on the communicative demands of the genre, subject matter, and the audience the text targets (e.g., Ryshina-Pankova, 2015, 58). Descriptive and narrative texts often share concrete individual experiences, while argumentative and expository texts are expected to be more reflective and distanced from the writers' feelings. As shown in Crossley (2020, 432), expository texts may contain less lexical repetition than narrative writing and syntactic coordination is more common in narratives while syntactic subordination is typical of expository texts.

2.5 Cohesion and lexical sophistication

Cohesion refers to lexical devices that assist readers in noticing ideas and arguments between sentences and paragraphs, and across the whole text. In the framework of Halliday and Hasan (1976), these cohesive signposts are grouped into five categories: referential devices, e.g., demonstrative reference via *this*/*these*, *similarly*/*otherwise*; substitution, i.e., replacing words instead of repeating them; ellipsis, i.e., omitting words; connective links within and between clauses and paragraphs; and lexical cohesion, which refers to using different but semantically related words and collocations in the text. Research suggests, for example, that L1 and EFL students use different cohesive means: native writers employ referential means and substitution, while EFL students use coordinating conjunctions and demonstrative pronouns (e.g., Bowen & Thomas, 2020). Overall, the analyses indicate that higher-scoring texts are mostly characterized by two cohesive features, referential cohesion, i.e., the use of pronouns and lexical repetition, and organizational tools, such as connectives and function words (Crossley & McNamara, 2016; Crossley et al., 2019). However, different textual genres are likely to require different cohesive means.

The use of lexically sophisticated words is another strong predictor of EFL writing quality (e.g., Crossley, 2020). Traditionally, lexical sophistication in essays has been operationalized as the use of diverse advanced words, such as low-frequency lexis and academic vocabulary (e.g., Coxhead, 2000). Currently, digital analysis tools can provide more fine-grained information on diverse types of infrequent lexis, e.g., words that elicit longer lexical decision and word naming reaction times, and vocabulary that is learnt at a later age in L1. In addition to advanced lexis, digital tools can also identify typical English word combinations (Kyle & Crossley, 2016) and combinations containing strongly associated word-pairs (e.g., Garner et al., 2020; Kim & Crossley, 2018). In sum, recent findings demonstrate that the features relating to writing quality are also associated with lexical proficiency in terms of cohesion and lexical sophistication.

Productive use of cohesive devices (e.g., Kim & Crossley, 2018) and lexically sophisticated vocabulary (e.g., Berger et al., 2019; Hashimoto & Egbert, 2019) associate with writing quality and lexical proficiency. Research findings indicate that higher-scoring EFL learners are able to employ various referential devices and connectives to create cohesive texts and choose appropriate lexis from high-frequency vocabulary to rare words. The present examinations employ digital tools to identify cohesive devices (Article III; subsection 4.3) as well as lexically sophisticated vocabulary (Article IV, subsection 4.4) to examine how these features relate to human-rated essay scores.

2.6 Development of lexical competence in extramural contexts

In Europe, upper-secondary school teenagers are active users of the Internet, which provides daily encounters with English (Statista, 2016, 2022). In the Nordic countries, most teenagers have access to the Internet both at home and at school (Statista, 2022). In the extramural context, adolescents can choose their online activities and digital communities freely. The most popular activities are similar across Europe: listening to music with English song lyrics, viewing films and multi-episodic series with or without textual aid, browsing the Internet for fun, reading for information, playing computer games, and social networking (Statista, 2016, 2022).

Theoretically, appropriating English via self-chosen online activities refers to outcomes via incidental learning, i.e., picking up lexis without a conscious intent to learn or learning of one thing when the intention is to learn something else (e.g., Laufer & Hulstijn, 2001). According to Gass (1999), L2 vocabulary can be acquired as a by-product of other cognitive exercises involving comprehension. Online activities provide multiple elements that are thought to assist vocabulary acquisition, e.g., comprehensible input (Krashen, 1985), language production (Swain, 1985), noticing (Schmidt, 1990), involvement in the task (Laufer & Hulstijn, 2001), exposure to and frequent repetition of lexis (Eckerth & Tavakoli, 2012), motivation (Dörnyei & Chan, 2013), as well as self-selected topics and materials with meaningful content (Lee & Pulido, 2017).

Research confirms that frequent participation in extramural activities has a positive effect on various L2 skills such as communicative competence (Jabbari & Eslami, 2019), lexical knowledge (e.g., Sundqvist, 2019), reading comprehension (e.g., Brevik, 2019; Peters, 2018), and written production (Kim et al., 2018). Moreover, self-chosen activities allow adolescents to interact with peers sharing the same interests and adopt the role of a language user without having to stress over formal assessment (Cabot, 2018; Brevik, 2019; Hannibal-Jensen, 2017; Jalkanen & Vaarala, 2013). The more self-chosen language contacts the individual has, the more automatic L2 performance likely becomes. Research also indicates that digital contacts assist language acquisition across all age groups from young learners (Hannibal–Jensen, 2017; Lindgren & Muñoz, 2013; Sylvén & Sundqvist, 2012) and teenagers (Brevik, 2019; Peters, 2018; Sundqvist, 2019), to adults (Elgort & Warren, 2014; Rankin et al., 2006; Zheng, 2015) and the elderly (Sundqvist, 2022). Sundqvist (2019) found that playing lexically demanding computer games associated positively with the results of receptive and productive vocabulary tests. Brevik (2016, 2019) discovered that gamers tended to become better readers in L2 English than in L1 Norwegian. The relationship between diverse extramural digital activities and lexical knowledge across high-, mid-, and low-frequency vocabulary among Finnish teenagers is examined in Article I (see subsection 4.1).

3 Data and Methods

The examinations were conducted among 32–46 upper-secondary school students. A convenience sample was the only possibility due to scarcity of space in the computer rooms, as the writing tasks were conducted on similar desk-top computers to ascertain that all the participants had equal working conditions, and some test sessions were video-recorded. To establish the validity and reliability of the findings among a fairly small group of participants, the study triangulates qualitative and quantitative data gathered during the first and second school years.

Table 3. Procedure of collecting data: Tests, surveys and writing tasks.

	First Year		Second Year	
Test type	**First term**	**Second term**	**First term**	**Second term**
Questionnaires and self-reports	Out-of-school online activities _1	Dictionary use	Out-of-school online activities _2	Digital literacy and dictionary use. Online learning experiences
Receptive vocabulary	VLT_1			VLT_2
Associative vocabulary	Lex30_1		Lex30_2	
Recognition of word combinations	Collex _1	Collmatch_ 1	Collex_ 2	Collmatch_ 2
Indirect writing video-recorded	Gap-filling task 1	Gap-filling task 2	Proofreading task	
Video-recorded writing process, human rating, digital analyses	Essay_1	Essay_2		Essay_3

As shown in Table 3, the data comprise answers to questionnaires, results of receptive and associative lexical tests, scores in diverse indirect writing tasks and human-rated essays, as well as video-recordings to observe the working processes in the writing tasks. The essays were assessed by human raters and examined for cohesion and lexical sophistication by natural language processing tools. The first and fourth articles investigate lexical development longitudinally, whereas the second and third studies are based on the second-year tests.

3.1 Participants

The research began when the 16–17-year-old participants had started their first year at the upper-secondary level and ended when they were finishing their second academic year. By the end of the second year, all participants had completed six compulsory English courses (c. 180 hours) and were thus entitled to take the high stakes examination. It was agreed, firstly, that the test performance would not affect the participants' English grades or be shown to their English teachers; secondly, the tests would be conducted during school hours; and thirdly, the participants' anonymity would be ensured during the analyses. Participating in the examinations was arranged as an optional school course called 'Test your English'. Active participation was worth one credit toward the total needed before graduation. Conducting the examinations followed the ethical principles of the school and general ethical procedures for research.

3.2 Questionnaires

The participants answered four questionnaires prepared by the researcher and piloted with a different group a year before. The first questionnaire was an online survey of the participants' extramural activities. They were asked what activities they had, how frequently they engaged in the activity and how much English the activities provided (c.f., Appendix A). The same questionnaire was administered at the end of the first term in the new school, and then at the beginning of the spring term in the second year. Before answering, the participants were instructed to think about how they spent their free time during a typical school week. They were shown a list of extramural activities and asked to indicate how often they engaged in the activity and estimate the proportion of English in the activities. The list included online reading (Sites), playing computer games (Games), watching films, video clips, or multi-episodic series (Films), listening to music with English song lyrics (Music) and communicating on social network sites (Social Networks). Regarding Games, the participants were also asked to name their favourite games and indicate at what age they had started playing them. The frequency of exposure was measured on a three-point Likert scale, the options being Daily, Weekly or Rarely. The proportion of English content was assessed by choosing one of the three options: All English, Half English or Less English. Based on the responses, three variables were emerged: the type of activity, the frequency of the activity, and the extent of English content in the activity.

The second questionnaire gathered information on using online dictionaries. The questions concerned the types of dictionaries they were in the habit of consulting (e.g., online, bilingual, monolingual) and how often they consulted them. Moreover, they were asked if they had been instructed on how to search words in online

dictionaries and if they were given opportunities to practise dictionary use at school. The response options to the latter question were *I have received instruction and training*, *I have received some instruction but no training*, and *I have received neither instruction nor training*. Moreover, extra space was provided in case the participants wanted to name the dictionaries they were in the habit of using.

The third questionnaire surveyed participants' personal experiences of learning words via online activities. They were asked to evaluate lexical appropriation by choosing one of the following options: *A lot of words*, *Certainly some words*, *Perhaps some words* or *I don't know*. They were also asked to give examples of words they remembered having learned or to name semantic contexts pertaining to these new words. The fourth survey concerned the participants' digital skills and success in learning English. Computer skills were rated on a three-level scale: *I do not need assistance*, *I sometimes need assistance*, or *I often need assistance*. When comparing how successful they were in English studies versus other school subjects, the student chose one of these four options: *As strong as in other subjects*, *Somewhat weaker*, *Somewhat stronger*, *Much stronger*. Responses to this question were thought to reflect the learners' attitudes and motivation towards English studies (Dörnyei & Chan, 2013).

3.3 Vocabulary knowledge

Receptive vocabulary knowledge was investigated at high-, mid- and low-frequency levels using The Vocabulary Levels Test (Schmitt et al., 2001; Nation, 1983). A receptive test was chosen for the followings reasons: first, tests of productive word knowledge tend to show how well the test words are mastered but cannot describe the learner's entire vocabulary, or show which words the learner does not yet know (e.g., Meara, 2009; Milton, 2009); second, it is easier to compare the present results with previous findings, as the VLT is widely used in earlier studies and commonly considered a reliable and valid measure of lexical knowledge (e.g., Meara, 2009; Read, 2000); third, research has shown that receptive vocabulary knowledge is closely connected to L2 learner's ability to use English in multiple ways including productive language use (e.g., Alderson, 2005; Crossley et al., 2016b; Read, 2000, 2007; Schmitt et al., 2001). The VLT sections are scalable, i.e., recognition of infrequent words implies knowledge of more familiar words. Moreover, the VLT offers two different versions for the purposes of longitudinal testing.

3.3.1 Recognition tests

The Vocabulary Levels Test (VLT) measures vocabulary recognition of high-frequency, mid-frequency, and low-frequency words, i.e., lexis from the second

thousand (2K), third thousand (3K), fifth thousand (5K) and the tenth thousand (10K) frequency bands. According to Schmitt & Schmitt (2014), high-frequency vocabulary comprises words from 2K to 3K; mid-frequency vocabulary includes lexis between 4K and 8K, and low-frequency vocabulary covers words from 10K onwards. At each level, the test provides ten groups of words with six words and three definitions, and the task is to match the words to the definitions.

The VLT was administered twice with a one-year interval so that the test administration coincided with the writing sessions. The test was answered with pen and paper in a regular classroom. Two different test versions were used. The VLT also provides a separate test section for academic vocabulary (AWL). As AWL contains lexis from multiple frequency bands, this part was employed mainly in Article IV, which focuses on advanced words. Based on the VLT scores, an approximation of the learner's receptive vocabulary size in word families can be calculated using a method introduced in Laufer & Ravenhorst-Kalovski (2010). The instructions for calculation are provided in Appendix C.

The participants were also assessed on their recognition of collocations using two tests, COLLEX and COLLMATCH (Gyllstad, 2007). These tests were conducted longitudinally. COLLEX (version 5) includes 50 test items, each consisting of three parallel phrases marked a, b, c, in an answer box. The task is to indicate which of the three options is an English collocation by underlining the corresponding answer.

Example 1.

a. receive a cold	b. fetch a cold	c. catch a cold

COLLMATCH (version 3) consists of 100 lexical combinations. The test-takers are to decide whether the lexical combination is used in the English language or not. If they think that it is an existing English word combination, they tick the 'yes' box, if not, they tick the 'no' box.

Example 2. catch importance

yes	no

3.3.2 Productive lexical test

In addition to the VLT, the participants' lexical knowledge was examined longitudinally using a test to measure associative lexical knowledge. Fitzpatrick and Meara's (2009) Lex30 resembles a lexical production task based on word associations. The stimuli are common high-frequency words, which are likely to elicit as responses words from the mid- or low-frequency bands. The participants are

given 30 stimulus words followed by four empty lines. The task is to fill in the four lines with the words that first come to the test-taker's mind.

Example 3. attack _____ _____ _____ _____

In this test, proper nouns, numerals and the most familiar words from K1, score zero points, whereas all words from mid-, and low-frequency bands score one point each. The results of the collocation tests and Lex30 are triangulated with the digital findings in Article IV.

3.4 Writing tasks

The writing tasks comprised indirect writing such as gap-filling and proofreading, while productive writing refers to traditional compositions on a given topic. During the process, the students had free access to online dictionaries and informational web-pages. On the one hand, accessing online tools is thought to assist the writer in finding lexis and facts. On the other hand, using such tools combines the writing process with a complex reading task: gap-filling requires good reading skills to infer the overall meaning of texts with missing words; consulting online dictionaries and informational sources during any writing task entails locating task-relevant information, combining information from various sources and critical evaluation of the findings.

3.4.1 Indirect writing tasks

The participants performed two gap-filling tasks and one proof-reading and editing task on a computer on separate occasions. Gap-filling was chosen as it is a familiar task type for participants, and used in earlier dictionary research (e.g., Atkins & Varantola, 1998; Dziemianko, 2010). In contrast, the proofreading task was an experiment to simulate a real-world condition of language use. As no standardized tests were available, the researcher designed the tests so that the target items represented various degrees of objective complexity (Singer et al., 2012), involving several queries, choosing from multiple alternatives and meanings, or searching for a rare word or a common word with an unfamiliar meaning. In contrast, a non-complex target item would be easy to find if the entry offered only one translational counterpart or more than one suitable option. However, a simple item may be subjectively complex due to the participant's inadequate word recognition skills (Singer et al., 2012), or internal influences such as self-efficacy or low motivation to conduct the task (Marek & Wu, 2014).

The gap-filling tasks represented two different text types. The subjects were instructed to fill in the gaps according to the L1 (Finnish) prompts so that the additions fitted both the context and register. The first gap-filling with eight target items was a formal letter to a potential customer from a company providing investment services. One noun (annuity) and one pragmatic formula (a formal letter ending) were regarded as a search for factual information, as looking for the answers was expected to require consulting not only dictionaries but also other informational sources. The second task with 13 target items was a blog post in which a school-boy reflected on the advantages and disadvantages of obtaining a college degree. All the target items in the gap-filling tasks included finding or checking the meanings of individual words or phrases. The experimental proof-reading task was based on poorly written instructions for the use of a small electrical device, a laser pointer. The original version was found on the Internet. The text included eight inappropriate word choices, which the participant was expected to find. The test-takers were asked to proofread, edit, and rewrite the text so that it would be easy for readers to understand.

The indirect tasks were rated in the following way. Successful use of online sources was defined as an ability to navigate online sites using menus and links (medium-related skills), and to locate and select the required items and evaluate whether the findings fitted the context (content related skills). The responses were divided into those involving looking-ups information and those without consulting any reference sources. The answers were then categorized as *Fully successful*, *Partly successful*, or *Unsuccessful*. In a *Fully successful* answer, the meaning and use of the word were accurate and appropriate to the context and register. If the meaning was slightly inaccurate, or the verb was used in the wrong tense, the answer was regarded as *Partly successful*. For example, the collocation *make a mistake* fitted the context only in the past tense. This may be considered overly strict, but on the other hand, the participants had ample time to check their answers. If the phrase did not fit the context or the word meaning was incorrect, the answer was rated *Unsuccessful*.

3.4.2 Productive writing and assessment

Written production includes three compositions written at school. The participants were asked to write 150–250 words on familiar topics according to prompts given using 3–5 bullet points. To hinder priming effects, the prompts were given in L1. The texts were written on similar computers in a computer room and the writing process was video-recorded. The writing time was 60 minutes, of which ten minutes was used for instructions. The participants were encouraged to consult online sources for lexis and information and to revise the texts using editing functions. They were also informed that the texts would be checked for plagiarism.

The first essays were composed at the beginning of the autumn term of the first year in the new upper-secondary school. The students were asked to describe their feelings and experiences in the new school and discuss differences between studying at two different school levels. The second essays were written in the spring term of the first year on the topic What's in a name? Before writing, the students were asked to find out, e.g., why their parents had chosen the particular names, and what the origin of their family name was. During the writing session, they were given additional prompts, e.g., to search the Internet to find out if the names were original Finnish names or Finnish versions of international names. As these essays were rated by the researcher, they were not analysed in the present study. The third essays were composed at the end of the spring term in the second year, when the participants were already preparing for the national school-leaving language tests. The participants were asked to discuss, for example, their academic success, describe the second-year traditions at school, and express their feelings in advance of the upcoming school-leaving examinations and/or their plans for future studies. The first and third essays are analysed in this study.

The essays were rated on a scale from 0 to 99 points, accounting for the content and structure of the text, lexical richness and accuracy, and ability to communicate clearly. The scale at the upper intermediate level (80–82–85–88) corresponds to the CEFR level B2. The same criteria were also used in the national school leaving examination. Currently, the scoring has been slightly changed. The raters were 28 teacher trainees finishing their studies at the Faculty of Education at a Finnish university. After scoring the essays on their own, they were asked to discuss the assessments in small groups. They were, however, encouraged to give their own scores independently. The interrater reliability (Cronbach's alpha) was strong, ranging from 96% to 99%. Before running the automated analyses, the texts were cleaned for spelling errors that changed the word meaning, e.g., taught instead of thought. The raters used the cleaned version, as the purpose was to draw attention to the structure, cohesion and lexical richness instead of accuracy. Regarding further analyses, only the first and third essays were analysed digitally for cohesion and lexical sophistication, as both topics were related to school and education. Unfortunately, the high-stakes essays were not available for research.

3.5 Video-recorded data

To examine writing as a process, working on the indirect writing tasks and composing the third essay were video-recorded using freely downloadable video-recording software, CamStudio. The participants were first shown how to switch on the video function and then asked to start working as usual. The recording software is unobtrusive for the writer. The recorded data were used to monitor, e.g., which online

sources the test takers employ, what they need to search for, how many times they search, and whether they decide to use the findings in the writing task. Finally, we evaluated whether the search results were appropriate for the context. However, due to the scarcity of time and space in the computer rooms, we were able to video-record only 22 out of 46 participants for the indirect writing tasks, and 31 students for the third essays. The rest of the participants conducted their task without a recording.

3.6 Natural language processing tools

TAACO and TAALES are freely available digital tools, which work without an Internet connection on Linux, Mac, and Windows operating systems and analyse batches of essays as plain text files (txt.). The tools can be found at https://www.linguisticanalysistools.org/.

3.6.1 TAACO 2.0.4

To analyse the essays for cohesive features, TAACO 2.0.4 (Crossley et al., 2019) was employed in the third study. The tool is designed to detect cohesion across sentences (local cohesion), paragraphs (global cohesion) and the entire text (text cohesion). The tool provides 194 indices for analysis altogether. Based on earlier research results (e.g., Crossley et al., 2016b; Kim & Crossley, 2018), the essays were analysed for the cohesive features explained in Table 4.

Table 4. Measures of cohesive features.

Index	The index measures
Lexical overlap across sentences Lexical overlap across paragraphs	repetition of all words and various parts of speech (Adjectives, Adverbs, Content words, Function words, Nouns, Pronouns and Verbs)
Semantic overlap across sentences Semantic overlap across paragraphs	semantic relations, e.g., repetition of noun and verb synonyms
Connectives	linking clauses and sentences
LSA semantic similarity	the average similarity between adjacent paragraphs measured by Latent Semantic Analysis
TTR	referential cohesion measured by type-token ratio
Determiners	the incidence of articles (referential cohesion)
Demonstratives	the incidence of demonstrative pronouns (referential cohesion)
Pronoun to noun ratio	number of third person pronouns divided by number of nouns
Pronoun density	number of third person pronouns divided by number of words

Most of these indices measured lexical overlap across sentences and paragraphs. Lexical overlap across paragraphs in particular has been demonstrated to have a positive correlation between writing quality. In this context, overlap means repeating the same lexical items to link text segments together, which is thought to help readers notice connections between ideas and information in the text (e.g., Crossley & McNamara, 2012). Moreover, a large group of different connectives, such as additive, causal, logical as well as the conjunctions *and* and *but*, were analysed, as connectors provide explicit links to the ideas and topics presented. Determiners and demonstratives were analysed to detect text-level cohesion.

The calculation of lexical overlap is explained briefly in the following. Lexical overlap across paragraphs measures the average number of words that are repeated in adjacent paragraphs. After importing the essays into the analysis programme, the texts are lemmatized and separated into paragraphs so that each paragraph includes only unique lemmas. In each paragraph, which is followed by another paragraph, all the words are checked to see if the same items also occur in the following paragraph. Each word that is repeated in the following paragraph at least once, increases the overlap count by one. The total score of the index is the number of words repeated in adjacent paragraphs divided by the number of words considered in the paragraphs. For calculation, TAACO also provides other procedures, which differ in terms of the number of paragraphs considered at a time and the denominator by which the sum of overlapping words is divided (Crossley et al., 2016a). For example, to receive a normed score, the total number of overlapping words is divided by the number of paragraphs considered. All the calculation procedures are explained in the TAACO 2.0.4 manual available at www.kristopherkyle.com.

3.6.2 TAALES 2.2

TAALES 2.2 (Kyle et al., 2017) provides 484 indices of lexical sophistication based on large corpora such as The British National Corpus (BNC) and The Corpus of Contemporary American English (COCA). The tool processes plain text files counting the number of incidences of the chosen indices. All TAALES measures are normed by text length. Durrant, Moxley and McCallum (2019) found that indices based on different corpora often overlap, and thus the number of indices can be diminished. Based on earlier findings, the variables (Table 5) that are shown to associate with essay quality in previous research (e.g., Berger et al., 2019; Kim et al., 2018; Kyle & Crossley, 2016) were chosen for analysis of lexical sophistication in Article IV.

Table 5. Measures of lexical sophistication.

Index	The index measures
SUBTLEXus Freq. FW log	how frequently a Function Word occurs across documents within SUBTLEXus corpus (Brysbaert & New, 2009)
SUBTLEXus Range FW	the number of texts within the corpus in which the item occurs
COCA Acad. Bigram Freq.	how frequently a Bigram occurs across documents in the *Corpus of Contemporary American English* (COCA; Davies, 2009).
COCA Acad. Bigram Proportion 100k	the proportion of Bigrams among the 100 000 most frequent bigrams within documents in the subcorpus of COCA
COCA Acad. Bigram AC	association strength of two-word combinations (spoken subcorpus of COCA)
MRC Imageability AW	how perceptible the mental representation of the word is, and how likely it activates other words. MRC Psycholinguistic Database (Coltheart, 1981)
Kuperman AoA FW	the estimated age of acquisition (AoA) of a function word in L1 (Kuperman et al., 2012).
LD Mean RT CW zsc	standardized mean lexical decision reaction times for content words. (The English Lexicon Project; Balota et al., 2007).
EAT FW types	function words elicited in free association tasks (Edinburgh Associative Thesaurus, Kiss & al., 1973)
Sem D	the variability of semantic contexts in which the word occurs. Based on LSA (Landauer et al., 1998; LSA). Calculated by Hoffman, Ralph, and Rogers (2013).

One of the benefits of TAALES 2.2 is the ability to detect n-grams, i.e., common English word combinations, which are also strong predictors of language proficiency (Durrant et al., 2019; Garner et al., 2020; Kim et al., 2018).

3.7 Statistical analyses

In Article I, the qualitative information was arranged into exposure categories according to the type and frequency of the activity, and the extent of English content, and coded numerically for each participant. The participants' vocabulary recognition scores (VLT) were then correlated to the coded categories (Pearson's r). Further, simple linear regression analyses were run to examine how the exposure variables (type, frequency, English content in the activity) relate to the VLT scores of high-, mid-, and low-frequency lexis. Finally, a stepwise regression was run to examine how diverse combinations of two activities predict lexical recognition.

In Article II, Finnish L2 learners' success rates in consulting online reference sources were examined in relation to the VLT scores across the high-, mid-, and low-frequency bands. The participants' responses were divided into answers with and without consulting online sources, and categorized as fully successful if the meaning of the word was accurate and contextually appropriate; partly successful, if the meaning was inaccurate, the word was not a common collocate, or it was used in a wrong form; and unsuccessful, when the word or phrase did not fit the context or register or was given with a wrong meaning (see examples in Niitemaa & Pietilä, 2018, 456). In Articles I and II, correlations (Pearson's *r*) and regression analyses were conducted using SPSS, version 23.

In the last two studies, correlational analyses were conducted to examine how the VLT results, and the essay ratings related to the automatically identified cohesive indices (Article III) and features of lexical sophistication (Article IV). In these studies, the correlations and regression analyses were calculated using the robust bootstrapped method (Larson-Hall, 2016, 213–214). Robust tests do not assume normal distribution, but the regression analyses need to be checked for outliers, homogeneity of variances, and normality and independence of residuals This method is thought to provide more accurate inferences when the data are not well-behaved or when the sample size is small (Larson-Hall, 2016, 587–600)**.** Robust tests also provide confidence intervals (CI), which indicate that the actual correlation coefficient is within the bounds of the CI with a probability of 95%. Although CI is wider in smaller samples, it indicates that the correlation coefficient is within the bounds of the CI with a probability of 95%. The result is statistically significant if it does not pass through zero. The values for the effect (R2) are interpreted according to Plonsky and Oswald's (2014) guidelines: R2 = 0.06 is small, R2 = 0.16 is medium and R2 = 0.36 is large. The bootstrapped analyses were conducted using SPSS, version 27.

4 Results

In the four research articles, vocabulary recognition capacity is examined in relation to developing lexical knowledge via out-of-school activities, consulting English reference sources, and written production with special reference to cohesion and lexical sophistication. These variables are interconnected: Frequent language use in free-time activities expands lexical recognition skills; efficient consultation facilitates finding and appropriating lexis, which results in better writing scores. To examine the development of receptive lexical competence, the VLT was examined in both the first and in second school years. The results over time showed no significant differences between the mean scores of high-frequency words (2K–3K), while the scores of mid-frequency (5K) and low-frequency vocabulary (10K) were significantly higher in the second year compared to the first test. The total mean score (77.48) was 65% of the maximum points. However, the standard deviation (30.18) was wide, as roughly a third of the participants scored 50% or less of the total score, while another third scored 80% or more.

4.1 Article I

Article I focused on the development of English vocabulary skills in out-of-school contexts (N = 46). To attain the expected level, CEFR B2, by the national school-leaving examinations requires substantial vocabulary knowledge. It is likely that meeting these lexical standards will require not only formal instruction but also frequent informal use of English.

The first research question asked to what extent Finnish upper-secondary students engaged in extramural activities in English. The participants' activities were surveyed over two years focusing on three variables: the type of the activity, the frequency of the activity, and the extent of English content (see subsection 3.2). The answers were largely consistent over two years: approximately half of the students encountered *All-English Daily Content* through *Music* and watching *Films*; and slightly less than half through reading online *Sites* and playing computer *Games*. Regarding the small changes in frequency, several participants reported that they had less time for any activities in the second school year. As for *Social Networking*, only

a small number of students encountered English, as most participants preferred using Finnish in this activity.

The second research question inquired about the participants' personal experiences of extramural word learning. This survey was answered only in the second year. Over half of the participants (65%) thought that they had learned "a lot of words", about a third (31%) chose the option "certainly some words", and 4% answered that they may have learned "some words". In the same questionnaire, the participants were asked to give examples of words they had appropriated via activities. Most of the participants (78%) answered by giving single words, such as *queue* or *promptly*, or generalizations like useful phrases or slang. Most examples were common words from the high- or mid-frequency bands. As for semantic fields, the participants reported having learned vocabulary related to current affairs and society, battle and warfare, information technology, crime and law enforcement, gaming, music, and hobbies.

The third research question examined the connection between extramural activities and recognition of English words across the high-, mid-, and low-frequency bands. For this purpose, correlations were run between the VLT scores and the *type of the activity*, *frequency* of participation and the *extent of English content* (see 3.2). Pearson's correlations reported significant positive correlations, firstly, between the VLT scores and encountering English through *Films* and *Sites* on the condition that the activity was *Daily*, and secondly, between the VLT scores and playing computer *Games* with *All-English content*, even if gaming was less frequent. In contrast, listening to *Music* with English lyrics and *Social Networking* showed either non-significant positive correlations or negative significant correlations between the VLT scores.

Compared to the other activities, playing computer *Games* was typically a long-term Daily activity with All-English content although some gamers informed that they had less time for the activity in the second school year. Moreover, a group of gamers reported playing games that require extensive English vocabulary and interaction in English with the gaming community. They had also begun regular gaming at the age of 8–10, assisted by their family members. Thus, an additional category, *Games+*, was formed to examine this group separately in the statistical analyses.

Simple regression analyses were conducted to find out how different activities predicted the VLT scores of high-, mid-, and low-frequency vocabulary**.** As shown in Table 6, language contacts through *Films*, *Sites*, *Games* and *Games+* promoted recognition of English vocabulary across all frequency levels, but the effects (R2) varied over time. In the first year, the largest positive effect on high-frequency vocabulary was found for *Games*, while *Sites*, *Films* and *Games+* showed smaller effects. *Games* was also the best predictor of recognition of mid-frequency

vocabulary, while *Sites*, *Films*, and *Games+* reported smaller effects. Effects on low-frequency vocabulary are of special interest, as infrequent words are rarely examined among younger learners. The analyses showed that *Games*, *Games+* and *Sites* exhibited practically equal effects (around 30%) on recognition scores of low-frequency vocabulary, and a smaller positive effect was found for *Films*.

Table 6. Effects of daily extramural activities on vocabulary recognition. Pearson's *r* and adjusted R2.

	High-frequency words				**Mid-frequency words**				**Low-frequency words**			
	1st year		2nd year		1st year		2nd year		1st year		2nd year	
Activity	*r*	R2	*r*	R2	*r*	R2	*r*	R2	*r*	R2	*r*	R2
Films	.518**	.252	.371**	.118	.485**	.218	.451**	.185	.476**	.209	.351*	.103
Games	.691**	.466	.514**	.248	.667**	.433	.571**	.310	.560**	.299	.390**	.133
Games+	.520**	.254	.355*	.155	.554**	.291	.468**	.201	.558**	.295	.596**	.341
Sites	.548**	.284	.536**	.271	.497**	.230	.524**	.258	.552**	.289	.421**	.158

In the second year, the largest effects on high-frequency vocabulary were found for *Sites* and *Games*, while *Films* and *Games+* reported positive smaller effect sizes. The same activities also showed the largest effects on mid-frequency lexis. The largest effect on infrequent words was observed for *Games+*, while *Sites*, *Games* and *Films* showed considerably smaller effects. Lastly, stepwise regression was run to examine the combined effect of *Games+* and *Sites* on recognition of infrequent words. The analysis yielded a significant model showing that reading online and playing lexically demanding computer games explained 45% of the variance in the score of low-frequency words in the second year.

The results indicate that using English in cognitively demanding contexts, e.g., reading and gaming, develops not only overall lexical recognition skills but also recognition of infrequent lexis. Reading is traditionally considered one of the best ways to increase vocabulary knowledge (Nation, 2006, 2016). Although reading online is often done in shorter stretches, it demands diverse reading strategies, e.g., scanning to locate information and careful reading to find facts (Tono, 2011). Gamers often excel at vocabulary tests, as performing well in more complex games requires mastering key vocabulary, which motivates gamers to appropriate the essential lexis (e.g., Brevik, 2016). Full-length films may contain infrequent vocabulary even up to the 9K, but simultaneous exposure to language, gestures, sound, and visual clues facilitates understanding content with rare words (Sweller, 2005). Multi-episodic TV series provide familiar vocabulary, but they are also rich in phraseology, and repetition of formulaic language facilitates comprehension and learning (Lin, 2014). In Finland, films and series are subtitled, which gives the

audience a chance to compare whether what they heard was the same as what was said (e.g., Peters & Webb, 2018).

4.2 Article II

The second article concerned indirect writing tasks based on gap-filling and proofreading. The former was a familiar test type, while the latter was an experimental test in which the task was to proofread a short text and correct the mistakes in it. In contrast to earlier studies, the participants were allowed to use the Internet and choose online sources freely. Consulting references was regarded as reading for information and instructions (cf., CEFR, 2018, 63–64), and consulting dictionary entries was considered a cognitively complex process involving vocabulary knowledge, information processing and problem-solving, in other words, competences that are essential to reading comprehension in general (Tono, 2011). Due to scarcity of time in the computer room, the recording process was conducted among 22 upper-secondary school students from 17 to 18 years of age during their second year at the upper-secondary level.

The first research question enquired, firstly, which online dictionaries and informational sites upper- secondary school participants usually consult when they choose the sources freely, and secondly, how successfully they use online dictionaries and translation tools. The participants' self-reports indicated that 14 out of 22 participants preferred bilingual online dictionaries (Finnish-English-Finnish); two students consulted monolingual dictionaries and three used both, whereas three participants did not consult any dictionaries. Moreover, the participants preferred Google Translate as a translation tool, Google.com as a search engine, and Wikipedia in either English or Finnish as a source for information. Eighteen participants had received neither information nor training on how to use online dictionaries, while four students had been told that such dictionaries exist.

The video-recordings (N = 22) demonstrated that even if 11 different dictionaries were used in the gap-filling tasks, most subjects had one principal source, which was either the only source or the source they always consulted first. The most frequently employed dictionary was Sanakirja.org with 109 look-ups. For proofreading, which appeared to be the most difficult task, the participants queried three main sources, Sanakirja.org, Google Translate and Google.com. Google Translate was the most popular translation tool with 120 consultations.

The recorded data revealed that, in the gap-filling test, the participants consulted online sources if they did not know a word, e.g., the adjective *entrepreneurial* in the formal letter, or when they were not quite sure of the meaning, some writers checked the word *funds* after using it. The third reason was to check the spelling when the participant had an approximate idea of what the item would sound or look like, e.g.,

sincerely instead of **sincinerally*. Regarding the success rates, nearly half of the look-ups were fully successful, one third were partly successful, and one fourth were rated unsuccessful (see subsection 3.4.1 for assessment). The proofreading task elicited fewer consultations, which is likely due to the new test type, and, in contrast to gap-filling, the target items were not indicated. The most often searched item was the verb *become*. This common verb is usually difficult for Finnish learners, as the meanings *come* and *become* are not differentiated in L1.

The second research question examined the relationship between the participant's word recognition skills and successful use of online sources. The correlational analyses between the VLT scores across the high-, mid-, and low-frequency bands and successful queries showed that higher VLT scores were associated with successful use of online dictionaries and informational sources. In contrast, the lower the VLT score, the more unsuccessful answers were produced. Overall, the analyses showed that to succeed in 50% of look-ups required scoring at least 60% on the VLT. This is higher than the proportion suggested by Nation's (2015) study on using monolingual dictionaries. Another decisive factor was the type of the target item. The results suggested that finding single words was more closely associated with the VLT score than finding word combinations. Searching for collocates often requires awareness of the fact that certain words collocate, and the process of checking collocates requires extra effort with several queries. Moreover, as evidenced by the recorded data, inadequate vocabulary skills were often combined with poor digital skills, which complicated the process of locating and evaluating information. Table 7 outlines traits of efficient and inefficient users of online sources.

Table 7. Factors affecting the use of digital sources.

	Effective user	**Ineffective user**
Receptive lexical knowledge	**Adequate recognition skills** Scores over 60% on the VLT	**Low recognition skills** Scores under 60% in the VLT
Associative lexical knowledge	**High degree of language awareness** Awareness of register Awareness of the fact that certain words collocate	**Low degree of language awareness** Not aware of register Not aware of the fact that certain words collocate
Mechanics	**Good computer literacy** Types in the query quickly and efficiently Exploits the spell checker Uses the copy & paste function	**Poor computer literacy** Types in the query with several spelling mistakes Does not exploit the spell checker Difficulties with the copy & paste function
Dictionary use	**Multifaceted use of sources** Uses two or more information sources	**Limited use of sources** Uses one primary information source
Locating information	Uses bilingual and monolingual information Searches for additional information	Uses mainly bilingual information Does not search for additional information
Evaluating information	Compares the alternatives Crosschecks the findings	Chooses the first alternative Takes the finding at face value
Reading	Reads carefully highlighting task relevant parts and / or moves the cursor along the lines	Reads carelessly moving the cursor all over the entry without focusing on task relevant parts
Disposition and extramural activities	Resilient. Conducts successive queries. In free time, uses English frequently in cognitively complex contexts, e.g., reading and gaming.	Gives up easily. Stops querying if the item is not found immediately. In free time, uses English frequently when listening to music and watching TV series.

To summarize, succeeding in the gap-filling and proofreading tasks required several competencies: rapid lexical recognition, diverse reading strategies to comprehend the text in dictionary entries and informational pages, content-, and medium-related digital skills to formulate queries and evaluate the search results. These abilities were associated with lexical knowledge. Compared to the participants scoring low on the VLT, the high-scorers looked up fewer words and conducted fewer successive queries. However, if necessary, they consulted several sites before decision-making.

4.3 Article III

Article III examined Finnish upper-secondary school learners' (N = 46) associative word knowledge in relation to cohesive features and human-rated essay scores. Moreover, the study searched for the point at which the learner's lexical recognition capacity would be large enough for the emergence of associative word knowledge. The writing process was video-recorded (N = 31) to find out whether and how consulting online sources helped to enhance cohesion.

The first research question asked, which cohesive devices characterised Finnish upper-secondary school L2 essays, and how the essays fulfilled the CEFR expectations for cohesion at level B2. To answer the first part, the essays were analysed for cohesion using a digital tool, TAACO (c.f., subsection 3.6.1). The analyses showed that roughly a quarter of the features analysed correlated significantly with the essay scores. The strongest positive correlation was found for adverbs and adverbials ($r = .525$, $p = .000$). Items like *compared to, definitely, firstly, in addition, nevertheless, regarding* and *regardless* were scarce or non-existent in lower-scoring essays, while higher-scoring essays included several types of such words. In addition, the incidence of function words ($r = .515$, $p = .000$) and verb synonyms ($r = .513$, $p = .000$) demonstrated highly significant correlations with the essay ratings.

In contrast to previous research (Crossley et al., 2016b), no connection was found between writing scores and the type-token ratios or the use of pronouns and determiners. The analyses indicated that higher-scoring essays included diverse types of connectors, and semantically related content words to specify meanings. In contrast, the correlation between essay scores and the conjunctions and/but was negative. Bootstrapped multiple regression analyses found that the incidence of adverbs across paragraphs and conjunctions and/but, collectively reported the best significant model (F (3.42) = 14 008, $p < .001$, R2 = .37) explaining 37% of the variance of the essay scores. The other combinations with two indices yielded smaller effect sizes from 30% to 34%. Higher-scoring writers employed from four to ten diverse types of connectors. The most widely used connectives in the these texts were *because*, *when*, *if*, *as* and *so*, in that order. In contrast, using only the conjunctions *and* / *but* had a negative effect on the essay scores predicting 23% of the variance. Lower-scoring writers tended to start sentences in a colloquial mode with *and*, e.g., "*And I have good memories*…" Demonstrative reference, such as using adverbs as modifiers across paragraphs (*especially, definitely, hopefully*) was a positive predictor of the essay scores. The adverbs and connectors used in the second essays are provided in Appendix E.

To answer RQ1, the analyses by TAACO showed that a rich array of adverbs as referential devices across paragraphs and a wide range of connectors were the best predictors of the essay scores. Regarding the CEFR descriptors for cohesion

(Council of Europe, 2018), Finnish L2 writers at level C1 employed referential and lexical cohesion, substitution, and a wide range of connectives, B2 writers used fewer types of referential devices and connectives but were able to avoid errors disrupting readability, while B1 writers overused the conjunctions and/but, and often forgot to structure the text in paragraphs, which diminished the clarity of the text.

The second research aim was to use video-recordings to observe how cohesion emerged during the writing process, and under what conditions using online sources enhanced cohesion. The participants' VLT results were compared to the essay scores and the lexis in the essays. These comparisons revealed that the participants scoring 80% or more in the VLT were familiar with most of the topic-related words they needed, but if they consulted the sources, they were able to choose appropriate words and expressions. In contrast, the participants scoring 64% or less on the VLT did not benefit from consulting the reference sources, so the participants with intermediate lexical recognition skills seemed to benefit the most from consulting online reference sources.

The observations via video-recordings demonstrated that consulting online dictionaries was closely associated with the participant's VLT scores (see subsection 4.2 on indirect writing), as word recognition skills were essential for choosing appropriate lexical items to improve referential cohesion across paragraphs, and the entire text. However, consulting dictionaries was not simple. Most of the participants employed a non-expert constructed dictionary[2] which provided uncontextualized lists of words in the entries. For example, one of the most frequently searched words was the noun (school) *subject*. The Finnish counterpart (aine) is a polysemic word with many meanings, e.g., *matter*, *material*, *substance* and *essay*. Among the participants who consulted this dictionary, five writers chose the noun **material* and one chose **substance*. The latter was also the primary suggestion made by Google Translate. Moreover, the writers using Google Translate often relied on the suggested literal translations. For example, one student tried to explain his future plans via a phrase, which Google Translate formulated as "*I read myself a building engineer*." When reading informational pages such as Wikis or the home pages of educational institutions, word recognition skills also assisted some participants with noticing items that they were not actually looking for. For example, *advanced course*, *University of Applied Sciences* and *citizenship* were found by chance and used in the essays.

There were only a few cases in which changing the search term helped the writers to find the appropriate target. The observations also revealed that word combinations were rarely searched for and the choice of the collocates was strongly affected by L1

[2] Sanakirja.org is a free multilingual source operated by a leading Finnish newspaper.

interference. As an example, in Finnish, you "*write an examination*" and "*read a school subject*". When consulting Wikipedia for *matriculation examination*, one student found the collocate *take*, but also used the incorrect combination in another sentence. Regarding the collocating verb for school subject, Google Translate suggested read, e.g., *read mathematics**.

In sum, diverse issues complicated the use of online reference sources: inadequate lexical meaning recognition, not knowing how to use dictionaries, a lack of digital skills, relying on one source without crosschecking and evaluating the information, and forgetting to consult definitions and examples, which would have provided collocates and information on the register. In comparison with the CEFR descriptions for cohesion (Council of Europe, 2018a), approximately 60% of the essays had reached at least level B2, which is the stage that Finnish EFL students are expected to reach in English by the national school-leaving examination.

4.4 Article IV

The fourth article is a longitudinal study of lexical sophistication, e.g., the incidence of infrequent words that elicit longer response times in lexical decision tests, and common English two-word combinations (n-grams). Lexical sophistication was digitally identified by TAALES 2.2 (Kyle, Crossley, & Berger, 2017) and triangulated with results from traditional lexical tests measuring receptive and associative lexical knowledge. The essay topics are explained in subsection 3.2.

The first RQ asked how human-rated essay scores related to the results of the traditional lexical tests over time. As shown in Table 8, all the second-year mean scores are higher compared to the first results. A comparison of the means[3] showed that the difference was significantly higher for the VLT (Z =. –4.055, p =.000), recognition of collocations (Z = –3.754; p = .000), low-frequency words (Z =.–2.993, p =.003) and the essay scores (Z =.–4.694, p =.000). The growth was non-significant only for Lex30 tests due to the high p-value (Z = –1.115; p = .265) indicating that infrequent lexis was difficult even in the second year, as the maximum score remained under 50% both times.

[3] The distributions were first examined for normality (the Shapiro–Wilks test suggested for groups with 50 or fewer participants) and compared using a non-parametric test (Wilcoxon Signed Ranks test), as the first and second VLT test scores were non-normally distributed ($p \leq .050$).

Table 8. Results of traditional receptive and productive tests.

Recognition	Time	Mean / % of the maximum	Median	Standard Deviation	Minimum	Maximum
The VLT (max. 150)	Year 1	91.4 / 60.9	93	33.9	35	146
	Year 2	101.1 / 67.4	117.5	36.8	22	150
Low-frequency lexis K10 (max. 30)	Year 1	10.2 / 34	8	6.2	2	0
	Year 2	12.4 / 41.3	9	8.1	26	30
Collex, Collmatch (max. 150)	Year 1	92.2 / 61.5	89.5	16.1	65	65
	Year 2	97.2 / 64.8	93	18.8	132	138
Production	**Time**	**Mean / % of the maximum**	**Median**	**Standard Deviation**	**Minimum**	**Maximum**
Lex30 (max. 120)	Year 1	46.5 / 38.8	49.5	25.6	0	85
	Year 2	50.3 / 41.9	53	21.7	0	94
Essays (max. 98)	Year 1	72.7 / 74.2	75	10.3	48	95
	Year 2	77.7 / 79.3	80	9.2	58	95
Estimated vocabulary size in word types	Year 1	4911	4992	1826	1950	7700
	Year 2	5332	6217	1983	1333	8000

Next, the proportion of word types across frequency bands and academic word lists was analysed using Vocabprofile Compleat (Cobb, Tom, n.d.). The percentage of familiar words (K1–K2) was over 98% in the first year and under 90% in the second measurement. The percentages were similar across proficiency levels in the first year, whereas in the second year, the differences were small but perceptible (Table 9).

Table 9. Lexical frequencies in the essays by proficiency groups.

CEFR level	Number of essays / %		Percentage of K1–K2 types		Percentage of K3 types		Percentage of K4–K14 types		Percentage of academic word types	
Year	1st year	2nd year	1st year	2nd year	1st year	2nd year	1st year	2nd year	1st year	2nd year
C1	2 / 4%	7 / 15%	98.7	86.8	0.6	4.5	0.6	8.7	3.2	9
B2.2	17 / 37%	21 / 47%	98.2	88.9	1.1	6.3	0.4	4.4	5.9	8.2
B1.2	17 / 37%	9 / 19%	98.6	89.4	0.8	5.6	0.1	3.8	4.9	5.4
B1.1	5 / 11%	9 / 19%	98.8	90.4	0.8	5.3	0.4	3.6	3.6	3.7
A2	5 / 11%	-	98.5	-	0.6	-	0.7	-	3.8	-

In sum, the data suggest that a growth in the VLT allowed productive use of less familiar words, as the percentage of mid-frequency (K3), low-frequency (K4–K14) and academic words in writing increased over time.

The second RQ asked how human-rated essay scores relate to digitally identified lexical sophistication over time. The following examples were taken from the second-year essays. The substantial proportion of advanced word types explained the positive correlation between the writing scores and lexical decision mean reaction times for content words (*LD Mean RT CW zsc*). The higher-scoring (C1, B2) second-year essays exhibited lexis from high-frequency to low-frequency bands: *affect*, *complicated*, *option* (K2); *adjust*, *anxiety*, *relevant* (K3); *straightforward*, *vocation* (K4); *assimilate*, *mandatory* (K5); *compulsory* (K6) or even rarer words like *biotechnology* (K7); *elective* (K8), or *astrophysics* (K10). Infrequent words were used idiomatically in appropriate contexts, although the percentages of infrequent words and academic lexis at C1 were higher compared to B2. Lower-scoring essays included literal translations from L1 to L2, e.g., *numbers* instead of *grades* or *fulfil studies* instead of *completing* them. L1 words were also used without explaining their meanings, e.g., *lukio*, denoting the upper-secondary school, or *wanhat*, which means the second-year students' traditional celebration day. These errors were often combined with syntactical mistakes, such as using an infrequent noun as a verb, "*I decided to baccalaureate Swedish*." Another participant meant to express "it does not prevent me from" but wrote "*it do not hamper to me*." In this case, a K6 verb, *hamper*, was the only rare word in the otherwise simple text. Lower-scoring texts were mostly written in informal style. Occasionally intermediate writers also mixed registers, using the expressions *stuff like that* or *a little kid*.

The second predictor analysed the incidence of function words (*SUBTLEXus Freq. FWlog*). Function words include, e.g., prepositions, pronouns and articles. The following examples focus on auxiliaries and modals. In advanced texts, these verbs were used skilfully even in complex structures: "*I would have liked to be able to*…; *I do wonder from time to time if I should*…". Intermediate writers seemed to avoid linguistic risks by employing common auxiliaries in simple structures: "*Second-year studies have not been much different*…; *I have had more homework*…". In the weak texts, elementary errors occurred even in common constructions: "*because they were been great*…; *I have already wrote*…". More examples across proficiency levels can be found in Appendix D.

The third predictor analysed the essays for the proportion of common bigrams (*COCA Acad. Bigram Prop.100K*). The findings indicated that higher-scoring essays exhibited typical English two-word combinations. In the following excerpt from an advanced text, each two-word combination is possible: "*I have had some courses in which I was the only one of my age group, but it has not really affected my grades or my motivation, even though I would prefer to sit next to someone I know*". In an

intermediate excerpt, the words are combined properly, but the lexis and structures are simpler: "*My second-year studies have been really easy compared to first year studies and I have not stressed so much*". Despite the academic words, the following sentences from a weak text include literal translations from L1 and four non-target-like bigrams: "*In student writings I will write math and biology. I participate to writing next year on autumn*".

Bootstrapped multiple linear regression analysis (see subsection 3.7) was run to examine how the three features predict L2 essays scores overtime. The analyses showed that these features collectively explained 46% of the variance in the first-essay scores ($F(3,42) = 13\,593$, $p < .001$) and 44% of the variance in the second-essay scores ($F(3,42) = 12\,762$, $p < .001$) after adjusting for the sample size. Another regression analysis was run to compare the combined effect of traditional tests on the essay scores. As no corresponding test was found for function words, the analysis was conducted using only two features, associative lexical knowledge (Lex30), and the sum of collocation tests, COLLEX and COLLMATCH. These features collectively explained 56% ($R2 = .559$) of the variance of the first-year essay scores and 61% ($R2 = .614$) of the second-year essay scores. All the effect sizes were large (Plonsky & Oswald, 2014). The analyses showed that the digital findings align with the results of traditional lexical tests. The finding implies that traditional tests for recognition and associative skills can also be used to assess lexical sophistication in the classroom.

4.4.1 Comparing the findings elicited by diverse methods

TAACO and TAALES were not designed for assessment in education, although they provide valuable information on productive writing. Therefore, we compared findings elicited by traditional lexical tests and digital methods. As Lex30 is a measure of productive knowledge of advanced words, it can be regarded as a rough counterpart for LD Mean RT CW zsc, i.e., the TAALES index for content words taking longer reaction times in lexical recognition tests.

Table 10. Comparison of correlations between the essay scores and diverse test measures over time.

First year		Second year	
Traditional test / 1st essay score	**Digital index / 1st essay score**	**Traditional test / 2nd essay score**	**Digital index / 2nd essay score**
Lex30 ***r* = .628, *p* = .000**	**LD Mean RT Zsc. CW** *r* = .555, *p* < 001	**Lex30** *r* = .594, *p* = .000	**LD Mean RT Zsc. CW** *r* = .528, *p* < 001
Collex/Collmatch ***r* = .712, *p* = .000**	**COCA Acad. bi prop 100K** *r* = .304, *p* = 040	**Collex/Collmatch** *r* = .764, *p* = .000	**COCA Acad. bi prop 100K** *r* = .396, *p* = 006

As shown in Table 10, the correlations between the essay scores and Lex30, and between the essay scores and the corresponding digital index, are large and highly significant over time. The finding supports the assumption that these tests measure the same construct, i.e., production of advanced lexis. Regarding word combinations, the correlations between the essay scores and the collocation tests Collex / Collmatch are positive, large and highly significant, while the digital tests show moderate positive correlations between the essay scores. The result was expected as Collex / Collmatch measure meaning recognition, while the digital index indicates productive use of two-word combinations. However, the test scores show a rising trend in the second year. Bootstrapped regression analyses run with these tests showed that Lex30 and Collex/Collmatch collectively explained 56% of the variance in the first essay scores, and 61% in the second essay scores. The use of function words (*SUBTLEXus Freq. FW log*) was not included in this experiment, as no traditional test is available for these words, although based on the coverage they provide, function words form an important part of lexis for language learners (Kremmel, 2016). High scores in the VLT were closely related to essays scores. In the first year, the VLT score explained 64% of the variance in the essay scores. In the following year the percentage was 56%. Thus, the results elicited by traditional tests are in line with the digital results. The finding implies that traditional tests for receptive and associative skills can be used to assess lexical sophistication in language teaching at school.

In sum, the findings implied, firstly, that digital indices correlated positively with features that are also appreciated by human raters, such as the use of advanced vocabulary, a wide range of function words and typical English two-word combinations. Secondly, function words played a key role both in cohesion and lexical sophistication. Appendix D illustrates how the use of auxiliaries and modal verbs varies across frequency levels, and Appendix E shows how wide a range of adverbs and connectors were found in the essays. Thirdly, the analyses proved that traditional tests for receptive and associative skills can be used to assess lexical sophistication in language teaching at school.

5 Discussion and Conclusions

In his article on the role of vocabulary teaching, Richards (1976, 84) suggests that "a major feature of a second language program should be a component of massive vocabulary expansion". In this study, insufficient lexical knowledge predicted low essay scores and inability to use English online sources effectively. The variation between the VLT results was unexpectedly wide, and some second-year participants scored even lower compared to the first year, although the reason likely lies in motivational problems. The estimated vocabulary sizes in Table 8 illustrate individual differences more concretely. Presuming that a word family includes three word types , the lowest-scoring participant would recognize about 4000 words, which allows understanding easy authentic texts, asking simple questions and understanding the answers to them. Receptive knowledge of around 3000 is considered the threshold needed to progress beyond elementary levels in reading English (Milton 2010). The highest scorer would recognize circa 24000 words. If learners can employ 50–75% of the lexis they recognise (Meara 2009), this student could use 12000–18000 words productively. According to Meara (2006, 2009), vocabulary size and depth are interrelated; the more words we recognise, the better we can use them.

Moreover, the results of traditional lexical tests and digital analyses suggested that skilful use of lexis is connected with writing proficiency. The learners' developmental path seems to follow Richards' (1976) and Ringbom's (1983, 1991) frameworks (c.f., Table 2), which emphasize the importance of associative vocabulary knowledge, and thus, align with the current conception of individual lexis as a changing network consisting of, e.g., derivations from basic word forms, polysemous words and synonyms from different registers.

The general aim of the four studies was to examine Finnish upper-secondary school L2 students' receptive and associative lexical knowledge from diverse perspectives: Article I monitored the impact of extramural use of English on lexical development longitudinally; Article II examined the role of lexical proficiency in consulting online sources; Article III investigated cohesive features in the second-year essays; and Article IV analysed L2 essays for lexical sophistication longitudinally. In subsections 5.1–5.3, the main findings are briefly summarized,

relating them to previous results and theoretical perspectives on lexical knowledge. Subsections 5.4 and 5.5 conclude the dissertation by acknowledging some limitations, suggesting further issues for research, and discussing pedagogical implications for teaching L2 English.

5.1 Development of lexical knowledge in extramural contexts

It is thought that the most frequent 2500 – 3000 words are learned at school, whereas the recognition of less common words requires both formal and informal learning (e.g., Nation, 2015; Schmitt, 2008). Thus, the first study surveyed Finnish upper-secondary school L2 learners' extramural contacts with English through diverse activities. The purpose was to examine to what extent the informal use of English develops participants' lexical recognition skills (the VLT) when the type, frequency, and extent of English content in the activity are accounted for. The study contributed to previous research by including infrequent vocabulary (K10) in the analyses.

The longitudinal surveys showed that the common daily activities were listening to music with English lyrics, watching films and multi-episodic TV series, reading online sites, playing computer games, and social networking. The results indicated that three variables, the type and frequency of the activity, and the extent of English content, were important for appropriating lexis. Watching films or TV series and reading websites showed positive correlations with the VLT scores only if the activity provided all-English content daily. As for playing computer games, all three variables were essential, and the positive correlation between the VLT scores and gaming was stronger than the correlations between any other activity. In the second year, playing lexically demanding games (Games+) had the largest effect on the scores of low-frequency lexis. The combined effect of reading and playing more demanding games explained 45% of the variance of the recognition score of infrequent lexis. In contrast, the correlations between the VLT scores and listening to music or social networking, were non-significant even if the contact was daily and the content in English. However, the first year analyses detected one unexpected connection between recognition of low-frequency lexis and daily contact with English song lyrics ($r = .323$, $p = .028$). Further investigation revealed that one group of participants appeared to be fans of a genre specializing in peculiar lyrics. They also made music and wrote words for their own songs.

In research, gaming and reading are traditionally considered effective means of appropriating infrequent lexis (e.g., Nation, 2006, 2016; Brevik, 2016; Jabbari & Eslami, 2019). In the present study, Games+ predicted 36% of the variance of the low-frequency scores, while the percentage for less demanding games was 15%, which is on a par with Films and Sites. The finding indicates that the participants in

the Games+ category were able to increase their knowledge of infrequent vocabulary due to daily language use in a cognitively demanding context. The participants with lower VLT scores were more interested in Music and Films. The VLT scores improved over time but showed considerable variability in the second year, as the standard deviation was 30.1. It is noteworthy that recognition of infrequent lexis (K10) improved as the mean score was 34% of the maximum in the first year and increased to 41% in the following year.

Exposure to a larger number of words is feasible through reading. Browsing the Internet provides repeated exposure to diverse vocabulary so that appropriating lexis is possible even if the reader is interested in a limited number of topics (Nation, 2014) or visits the same websites repeatedly (Sockett & Toffoli, 2012). As for gaming, players are motivated to memorize the essential lexis in order to perform well and to communicate with the gaming community (Brevik, 2016, 2019). Although the corpus of game vocabulary is not available for analysis in the same way as there are transcripts for films, researchers assume that multiplayer games (MPGs), Massively Multiplayer Online games (MMOs) and Massively Multiplayer Online Role-Playing Games (MMORPGs) provide a large amount of infrequent lexis. Studies on gaming report that active gamers are good at recognizing rare words (Brevik, 2016; Coxhead & Bytheway, 2014), and also excel in productive vocabulary tests (Sundqvist, 2019).

From a theoretical perspective, appropriating lexis in extramural contexts is incidental learning, in other words, learning without a conscious intent to learn, or learning one thing when the intention was to learn something else (e.g., Laufer & Hulstijn, 2001). It is likely that a considerable proportion of L2 vocabulary is acquired as a by-product of other cognitive activities involving comprehension (e.g., Gass, 1999) and deliberate effort to improve performance (Ericsson, 2006). Activities also provide exposure to and frequent repetition of vocabulary (e.g., Eckerth & Tavakoli, 2012), comprehensible input (Krashen, 1985), noticing (Schmidt, 1990), materials with meaningful content (Lee & Pulido, 2017), task-induced involvement (Laufer & Hulstijn, 2001) and motivation (Dörnyei & Chan, 2013). Moreover, researchers have proved that informal encounters with English promote language acquisition in all age groups from young learners (Hannibal-Jensen, 2017; Lindgren & Muñoz, 2013; Sylvén & Sundqvist, 2012) to teenagers (Brevik, 2019; Peters, 2018), adults (Elgort & Warren, 2014; Rankin et al., 2006; Zheng, 2015), and the elderly (Sundqvist, 2022). Extramural activities provide opportunities for language use in the learner's own niche(s) and a chance to communicate with peers sharing the same interests. In such contexts, the learner can function as an active language user (Cabot, 2018; Jalkanen & Vaarala, 2013) and participate in something that is meaningful to them (e.g., Dufva et al., 2014) without having to stress over formal assessment. As illustrated in the Figure on page 14,

appropriation and competence are interactive: extramural exposure promotes L2 competence, and, in turn, lexical competence makes all encounters with English more accessible and enjoyable.

5.2 Receptive lexical knowledge and consulting online sources

The second study examined the relationship between lexical recognition skills and the ability to search online sources for words and information. The tests included diverse indirect writing tasks: gap-filling tasks with L1 (Finnish) prompts representing informal and formal registers, and a simulation of a real-world task, in which the participants were to proofread, edit and rewrite a poorly written text (see subsection 3.4.1 for details). Unlike most previous research, Article II examined dictionary use among teenaged L2 learners, allowed the participants to consult self-chosen reference sources, observed the consultation process employing video-recording, and examined the results in relation to lexical recognition across high-, mid-, and low-frequency bands.

According to the findings, the connection between the VLT scores and the success rates in consulting online sources differed depending on the type of item that was searched (see 3.4.1 for the rating system). If the target item was a single word, nearly half of the consultations were *Fully successful* showing a positive moderate correlation (Pearson's *r*) with the VLT score. In contrast, a negative moderate correlation was found between the VLT and *Partly successful* responses, while *Unsuccessful* answers showed a larger negative correlation with the VLT score. When searching for lexical combinations, the findings correlated negatively with the VLT, and conducting several queries for the same combination did not change the result. This indicates that the participants' receptive knowledge of lexical combinations was poor and that they did not know how to search for combinations. For example, when looking for *make a mistake*, some participants searched or checked the noun mistake but did not scroll further down on the page in the dictionary to find the whole combination "make a mistake".

Regarding the fact-finding tasks, the VLT and *Fully successful* responses showed a positive moderate correlation. In contrast to single words and word combinations, the correlation was slightly larger if the same item was queried more than once. It seems that the participants found it easier to read continuous explanatory text than concise dictionary entries and uncontextualized examples. A further analysis suggested that the recognition of infrequent lexis correlated positively with fact-finding, and moreover, conducting successive queries strengthened the correlation.

Despite varied materials and methods, the above findings align with previous results in that L2 learners' dictionary skills are inadequate (e.g., Boonmoh, 2012;

Chan, 2014; Nesi & Meara, 1994). Not knowing how to use dictionaries properly means that some learners stop searching or choose one of the first meanings in the entry instead of examining all the senses (e.g., Tono, 2011). From a theoretical perspective, consulting dictionaries and other digital sources requires good meaning recognition skills and associative word knowledge to combine words properly. "Reading dictionaries" is a cognitively complex task. The entries are consulted to gain a quick access to the word, which underlines the importance of the automated recognition of meanings. Moreover, a dictionary user needs to recognise the differences between the options in the entry in order to choose the most appropriate word for the text in question. In gap-filling tasks, the participants also needed to comprehend the text as a whole to know what to look for. The proofreading task required reading and inferencing skills to single out the words that did not suit the context. The results demonstrated that dictionary use was particularly difficult for learners with inadequate lexical skills. As L2 dictionary users often misinterpret what they read in the entry, Nesi and Meara (1994) indicate the need for reconstructing dictionaries so that they provide more information in a more accessible way.

Difficulties and errors in dictionary use are a showcase of Richards' (1976) and Ringbom's (1987) assumptions about knowing the constraints of meanings, having an awareness of polysemy, and understanding the syntactic behaviour of words. Polysemy, in particular, caused problems in both L1 and L2. Lexically less proficient participants believed that the meaning they knew was the only one (e.g., Chan, 2014), as demonstrated in the following examples. One of the target words was the noun *investment*. The Finnish counterpart "*sijoitus*" also means **placement*. If the participants had read the examples provided further down in the entry, they would have found the proper option. Another common problem was unawareness of register. One of the tasks was adding a word to the sentence in a formal text, the expected answer being either *assets* or *funds*. Some students completed it with *money* or *moneys*. As pointed out in Nesi and Meara (1994), semantic and usage errors tend to occur in the same sentence. In the proofreading task, only a few participants noticed that the word *deaf* should have been replaced by *blind*. Some lexical errors were due to the tip-of-the-tongue phenomenon. This means that the writer has an approximate image of a word but cannot remember it properly (Aitchison, 1994). In an example from our data, one participant wrote **sincinerally* instead of *sincerely*. Due to the complexities of dictionary use, the frequency of search queries did not relate linearly to the number of correct answers (e.g., Atkins & Varantola, 1998; Liou, 2000; Pelttari & Mutta, 2014).

5.3 Written production

Articles III and IV examined productive use of lexis. Article III focused on cohesion in Finnish upper-secondary school L2 essays (N = 46) in the second year, i.e., at the time the students are expected to fulfil the CEFR expectations for cohesion at level B2. Cohesive features were identified digitally by TAACO (see subsection 3.6.1) and correlated to human-rated essay scores. Article IV investigated lexical sophistication using another digital analysis tool TAALES (see subsection 3.6.2). The aims were to study how using advanced lexis developed over time in L2 essays (N = 46), and how lexical sophistication associated with essay scores and success rates in traditional lexical tests. In the second year, the writing process of 31 participants was video-recorded to observe whether or how consulting online sources assisted in developing cohesion and lexical sophistication.

5.3.1 Writing proficiency in relation to cohesion

The TAACO analyses showed that a wide range of adverbs and connectors enhanced referential cohesion, and semantically related content words and verb synonyms across paragraphs developed lexical cohesion and substitution (see the framework of Halliday and Hasan in subsection 2.5). These features had a positive effect on the essay scores each predicting from 22–28% of the variance of the essay scores. The multiple linear regression analyses with two predictors, the incidence of adverbs and the conjunctions *and*/*but*, reported the best significant model explaining 37% of the variance in the essay scores ($F(3,42) = 14\,008$, $p < .001$, $R2 = .37$). In terms of the CEFR descriptors, C1 writers developed referential cohesion across sentences and paragraphs using a broad range of adverbs and connective links; B2 writers used more familiar adverbs and connectives but were able to avoid errors disrupting readability. B1 students wrote in an informal style overusing the conjunctions and/but. Moreover, they often forgot to structure the text in paragraphs, which diminished the clarity of the text.

The findings align with previous results in that Finnish L2 learners concentrated on textual organizers in cohesion-building (c.f., Crossley, et al., 2016b; Kyle & Dascalu, 2019). Crossley & McNamara (2016) suggest that human raters may even show bias for organizational devices in L2 essays, as the content cannot be expected to be as versatile as in L1 essays. Doró (2014) and Mäntylä et al. (2020) point out that the typology of the learner's L1 may explain the differences in text production even better than the student's proficiency level. Similarly, Pietilä (2015) suggests that non-native students may not be aware of the writing conventions of English and employ their L1 norms to L2 writing. The choice of cohesive devices also depends on the textual genre. For example, argumentative essays may require more textual organization compared to narrative texts (Abdel Latif, 2021). In contrast to previous

findings (e.g., Crossley et al., 2016b), the present analyses detected no connection between writing scores and the incidence of pronouns, type-token ratios, or determiners. This exemplifies the impact of typological differences between English and Finnish. Determiners are particularly problematic for Finnish learners of L2 English, as articles are not used in the Finnish language. However, the correlation between the incidence of articles and the essay scores was approaching the significance level, which may suggest gradual development towards employing the right types of articles in the right places.

5.3.2 Writing proficiency in relation to lexical sophistication

Advanced lexis was digitally identified by TAALES 2.2 (see subsection 3.6.2). Three features measuring diverse aspects of lexical sophistication were chosen for further analysis. The first predictor (*LD Mean RT CW zsc*) calculated content words that elicit longer reaction times in lexical recognition tests. This index showed a positive correlation with the essay scores. In particular, the second-year essays at C1 and B2 levels included lexis from the high- to low-frequency bands (c.f., Table 9). In the higher-scoring texts, advanced words were used idiomatically in appropriate contexts, whereas lower-scoring essays included inappropriate lexis combined with syntactical mistakes, L1 words without explaining the meaning, and literal translations from L1 to L2. Usage examples at different proficiency levels were provided in subsection 4.4.

The second predictor measured frequency of function words (*SUBTLEXus Freq. FW log*). The use of the most common function words correlated negatively with the essay scores. Although all the essays included adverbs, connectors, auxiliaries and modal verbs, the usage differed greatly. Advanced texts contained more diverse types of infrequent function words used in complex structures, e.g., "*I would have liked to be able to*". Intermediate texts included common auxiliaries in simple structures, e.g., "*I have had more homework*". In the low-scoring texts, elementary errors occurred even when using the most frequent auxiliary verbs "*because they were been great*". The same examples also exemplify the third predictor (*COCA academic bi prop 100k*), a measure of the proportion of typical English two-word combinations. In the first two excerpts illustrating the use of function words, each bigram is possible in English, while the third one contains a non-typical combination, "*were been"*.

A bootstrapped regression analysis was conducted with the three features introduced above. The results showed that these indices collectively explained 46% of the variance in the first essay scores ($F(3,42) = 13\,593$, $p < .001$) and 44% of the variance in the second essay scores ($F(3,42) = 12\,762$, $p < .001$). These findings align with previous research in that higher-scoring texts contain diverse advanced

lexis, typical English two-word combinations, and less frequent function words (e.g., Kyle & Crossley, 2016; Kim, et al., 2018; Durrant, et al., 2019). In contrast to Garner et al. (2020), a significant correlation was not found between typical three-word combinations and writing quality. In sum, the analyses indicate that along with increasing associative lexical knowledge, Finnish upper-secondary school L2 learners start producing essays with more sophisticated lexis, less common function words, and typical English two-word combinations.

5.4 Limitations and future perspectives

Regarding the present examinations, some limitations must be acknowledged. Firstly, the results are based on a convenience sample, which was the only option due to the participants' timetables and scarcity of time and space at the school premises. Thus, the inferences based on the results concern the sample in question. However, the results seem to align with previous studies among larger populations and different age groups. Secondly, in addition to the questionnaires, it would have been worthwhile to conduct personal interviews with each participant to find out more about their individual experiences of the benefits or problems of accessing online tools during the tasks. Such discussions might also have revealed more about extralinguistic factors affecting their performance and working strategies. Thirdly, again for practical reasons, the raters were teacher trainees finishing their studies at the Faculty of Education and working as preservice English teachers in different schools. After first assessing the essays on their own, the twenty-eight trainees were divided into eight groups and asked to discuss the assessments together. In case of disagreements, they were encouraged to give their final scores independently. However, the group discussions may have affected the strong interrater reliability (Cronbach's alpha).

In future, large-scale longitudinal studies should be conducted to shed light on vocabulary development in multiple usage contexts and allow a wider range of statistical methods and test batteries with combinations of traditional lexical tests and digital findings. Regarding research designs, differentiating between the subgroups of readers, viewers and gameplayers calls for longitudinal research with multiple measurement points. For example, the number of girl-gamers seems to be growing, but we do not know if girls and boys prefer the same game genres. Moreover, knowing the role of digital activities in word learning among participants with different L1 backgrounds would be informative.

5.5 Pedagogical implications

In his article on the role of vocabulary teaching, Richards (1976, 84) suggests that "a major feature of a second language program should be a component of massive vocabulary expansion". It is thought that the most frequent 2500–3000 words are learned at school, whereas the recognition of less common words requires both formal and informal learning (e.g., Nation, 2015; Schmitt, 2008). According to Meara (2006, 2009), vocabulary size and depth are interrelated; the more words we recognise, the better we can use them. In this study, insufficient recognition skills predicted low essay scores and an inability to use online sources effectively.

Current research has proved that new words can be appropriated in informal contexts via various online activities. However, some students may still underachieve in the classroom as free-time encounters do not automatically translate to better results in formal tests. To encourage these learners, the first implication relates to convincing them that all encounters with new words are valuable, and that sooner or later, informal language use will manifest itself in test results. This aligns with the CEFR (2001) principles encouraging teachers to appreciate and document the whole range of learners' language skills, whether attained formally or informally. Perhaps students should be asked, once in a while, to share their online learning experiences with the classmates. A survey employed in Article I indicated that the majority of students had learned new words in digital activities and most of them remembered examples of such words.

The second implication concerns raising learners' awareness of the importance of lexical skills. The VLT is practical for diagnostic purposes. A shorter version without AWL and K10 better suits for ninth graders at a comprehensive school, while the entire version could be employed at the upper-secondary level in the first and second years using different test versions. First, the students start working by themselves and mark their own papers according to the teacher's instructions, with the maximum number of points being 120. The VLT includes lexis from K2, K3, K5, K10 and AWL and each test set includes ten questions with six options. Thus, the student encounters a total of 300 words in this test. Next, the students continue working in small groups to find out whether they also recognize the extra words. The meanings that no one recognizes could be checked in online dictionaries. In the second year, the students should try an English–English dictionary for this purpose. L2 students could also be given a chance to test their associative knowledge with Lex30, a free online tool to reveal the frequencies of the words elicited by the stimulus word. The test is also available in a pen-and-paper version. When using this version, the students check the word frequencies, for example, by inserting the answers into the Compleat Lexical Tutor by Tom Cobb (see subsection 3.3.2 for the scoring). This procedure can be used to raise the students' awareness of word frequencies as suggested by Richards (1976).

Thirdly, dictionary literacy should be included in the language curricula. The present findings revealed that many upper-secondary school students did not know how to consult online dictionaries. Teachers should not only demonstrate how to search for lexis effectively but also provide opportunities to practise searching, locating, selecting, and evaluating information, not only in dictionaries but also in other informational sources. Learners might also benefit from working together, comparing their individual search paths, sharing the results, and assessing the findings and translations generated by automated tools. These abilities will be needed later in life.

The fourth implication concerns examining how in-class and out-of-class practices complement each other and how the different learning environments can best assist students to develop their own lexical repertoire. Practices at school should promote critical thinking, improve digital literacy, motivate word learning and provide meaningful contexts for using lexis. Research is needed at the interface between offline classroom work and online activities to discover how the best learning practices could be combined. Answers are needed to the following questions: How can lexis be a part of formal teaching without merely digitalizing or gamifying textbook exercises? How can gamified exercises be made meaningful and interesting for individual students? How can L2 learners' lexical competence be developed in a real-world context? What if a language game was at the same time connected to other important goals? *Freerice* (https://unric.org/en/freerice/) is a vocabulary recognition game in which every correct answer donates rice to the United Nations' World Food Program. This game helps students to test their vocabulary knowledge informally, even if done at school, and at the same time, raises consciousness about global affairs. Such an activity might also increase motivation among students who otherwise underachieve on formal tests. The learners might also be interested in organizing *Freerice* competitions between language classes or other schools in the community. Schools could also ask students to design English vocabulary games of their own. Encouraging vocabulary learning in both in-school and out-of-school contexts could help adolescents realize that different learning environments complement each other.

Lastly, future research with a larger populations and different age groups could provide more nuanced information on the connections between the use of lexis, vocabulary size and human ratings. As writing is primarily communication between human beings (Enright & Quinlan, 2010; Weigle, 2010), researchers might also want to examine whether learners write differently when they know that the reader is a robot instead of a human teacher.

List of References

Abdel Latif, M. (2021). Remodelling writers' composing processes: Implications for writing assessment. *Assessing Writing*, *50,* 1–16. https://doi.org/10.1016/j.asw.2021.100547

Aitchison, J. 1994. *Words in the mind. An introduction to the mental lexicon*. Blackwell.

Alderson, J.C. (2005). *Diagnostic Foreign Language Proficiency*. London: Continuum.

Allen, L. K., Crossley, S. A., & McNamara, D. S. (2015). Predicting misalignment between teachers' and students' essay scores using natural language processing tools. In *International Conference on Artificial Intelligence in Education* (pp. 529–532). Springer International Publishing.

Atkins, B.T.S., & Varantola, K. (1998). Language learners using dictionaries: the final report on the EURALEX/AILA research project on dictionary use (pp. 21─81). In Atkins (Ed.), *Using dictionaries: Studies of dictionary use by language learners and translators*. Tubingen: Niemeyer.

Berger, C. M., Crossley, S. A., & Kyle, K. (2019). Using Native-Speaker Psycholinguistic Norms to Predict Lexical Proficiency and Development in Second-Language Production. *Applied Linguistics*, *40*, 22–42. https://doi.org/10.1093/applin/amx005

Berger, C., Crossley, S., & Skalicky, S. (2019). Using lexical features to investigate second language lexical decision performance. *Studies in Second Language Acquisition*, *41*, 911–935. https://doi.org/10.1017/S0272263119000019

Bocconi, S., Chioccariello, A., & Earp, J. (2018). *The Nordic approach to introducing computational thinking and programming in compulsory education*. Report prepared for the Nordic@BETT2018 Steering Group. https://doi.org/10.17471/54007

Boonmoh, A. (2012). E-dictionary use under the spotlight: students' use of pocket electronic dictionaries for writing. *Lexikos*, 22, 43─68.

Bowen, N. E. J. A., & Thomas, N. (2020). Manipulating texture and cohesion in academic writing. A keystroke logging study. *Journal of Second Language Writing*, *50*, 1–15. https://doi.org/10.1016/j.jslw.2020.100773

Brevik, L. M. (2016). The Gaming Outliers: Does out-of-school gaming improve boy's reading skills in English as a second language? (pp. 39–61). In E. Elstad, (Ed.), *Educational Technology and Polycontextual Bridging*. Rotterdam, the Netherlands: Sense Publishers.

Brevik, L. M. (2019). Gamers, surfers, social media users: Unpacking the role of interest in English. *Journal of Computer Assisted Learning*, 35,1–12. https://doi.org/10.1111/jcal.12362

Cabot, M. (2018). Personal English Learning Ecologies and Meaningful Input with Digital and NonDigital Artefacts. *Nordic Journal of Digital Literacy*, 13, 94–112. https://doi.org/10.18261/issn.1891-943x-2018-02-03

CamStudio (n.d.). https://camstudio.en.softonic.com/?ex=DSK-347.0

Chan, A. (2014). How can ESL students make the best use of learners' dictionaries? *English Today*, 30, 33–37. Web.

Chen, Y., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37, 849–880. https://doi:10.1093/applin/amu065

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, 11, 38–63. http://dx.doi.org/10125/44117

Cobb, T. (2016). Numbers or numerology? A response to Nation (2014) and McQuillan (2016). *Reading in a Foreign Language*, 28, 299–304. https://www.proquest.com/scholarlyjournals/numbers-numerology-response-nation-2014-mcquillan/docview/1850833361/se-2

Coiro, J. (2011a). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research*, 43, 352–392. https://doi:10.1177/1086296X11421979

Coiro, J. (2011b). Talking about reading: Modeling the hidden complexities of online reading comprehension. *Theory Into Practice*, 50, 107–115. https://doi:10.1080/00405841.2011.558435

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42, 214–257. https://DOI:10.1598/RRQ.42.2.2

Council of Europe (2018a).Common European Framework of Reference for languages: Learning, Teaching, Assessment. Companion volume with new descriptors. Language Policy Programme. Education Policy Division Education Department. https://rm.coe.int/cefr-companion-volume-withnew-descriptors-2018/1680787989

Council of Europe (2018b). Descriptors of competences for democratic culture. RFCD Vol. 2.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–38. http://www.jstor.org/stable/3587951

Coxhead, A., & Bytheway, J. (2014). Learning vocabulary using two massive online resources: you will not blink. In D. Nunan & J. Richards (Eds.), *Language Learning beyond the Classroom* (pp. 65–74). New York, London: Routledge.

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11, 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S., & McNamara, D. (2012). Predicting second language writing proficiency: The role of cohesion and linguistic sophistication. *Journal of Research in Reading*, 2, 115─135. https://doiorg.ezproxy.utu.fi/10.1111/j.1467-9817.2010.01449.x

Crossley, S., & McNamara, D. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, 7, 351–370. https://doi.org/10.17239/jowr-2016.07.03.02

Crossley, S.A., Salsbury, T., & McNamara, D.S. (2010). The development of semantic relations in second language speakers: A case for Latent Semantic Analysis. Vigo International Journal of *Applied Linguistics*, 7, 55–74.

Crossley, S.A., Kyle, K., & McNamara, D. (2016a). The tool for the automatic analysis of text cohesion (TAACO): automatic assessment of local, global, and text cohesion. *Behaviour Research Methods*, 4, 1227–1237. https://doi-org.ezproxy.utu.fi/10.3758/s13428-015-0651-7

Crossley, S.A., Kyle, K., & McNamara, D. S. (2016b). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S.A., Kyle, K., & Dascalu, M. (2019). The tool for the automatic analysis of cohesion 2.0: integrating semantic similarity and text overlap. *Behaviour Research Methods*, 51, 14–27. https://doi.org/10.3758/s13428-018-1142-4

de Rosnay, J. (2011). Symbionomic evolution: From complexity and systems theory to chaos theory and coevolution. World Futures: *The Journal of General Evolution*, 67, 304–315. https://doi:10.1080/02604027.2011.585896

Dziemianko, A. (2010). Paper or electronic? The role of dictionary form in language reception, production and the retention of meaning and collocations. *International Journal of Lexicography*, 23, 257–273. https://DOI:10.1093/ijl/ecp040

Doró, K. (2014). The lexical richness of undergraduate student essays: the influence of writing experience and essay writing strategies (7–19). In P. Lintunen, M. S. Peltola, & M.-L. Varila (Eds.), *AFinLA-e Soveltavan kielitieteen tutkimuksia*, 6.

Doró, K. (2015). Changes in the lexical measures of undergraduate EFL students' argumentative essays (57–76). In P. Pietilä, K. Doró, & P. Pípalová (Eds.), *Lexical Issues in L2 Writing*. Cambridge Scholars Publishing.

Doró, K., & Pietilä, P. (2015). Researching vocabulary in L2 writing: Methodological issues and pedagogical implications (pp. 11–25). In P. Pietilä, K. Doró, & P. Pípalová (Eds.), *Lexical Issues in L2 Writing*. Cambridge Scholars Publishing.

Dufva, H., Aro, M., & Suni, M. (2014). Language learning as appropriation: How linguistic resources are cycled and regenerated (pp. 20–31). In P. Lintunen, M.S. Peltola & M.-L.Varila (Eds.) AFinLa Yearbook, 6. Jyväskylä, AFinLa.

Durrant, P., Moxley, J., & McCallum, L. (2019). Vocabulary sophistication in First-Year Composition assignments. *International Journal of Corpus Linguistics*, 24, 33–66. https://doiorg.ezproxy.utu.fi/10.1075/ijcl.17052.dur

Dziemianko, A. (2010). Paper or electronic? The role of dictionary form in language reception, production, and the retention of meaning and collocations. *International Journal of Lexicography*, 23, 257–273. https://doi-org.ezproxy.utu.fi/10.1093/ijl/ecp040

Dziemianko, A. (2012). On the use (fullness) of paper and electronic dictionaries. In M. Paquot & S. Granger (Eds.), *Electronic lexicography* (pp. 319─337). Oxford: Oxford University Press

Dörnyei, Z., & Chan, L. (2013). Motivation and vision: an analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, 63, 437–462. https://doi.org/10.1111/lang.12005

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental vocabulary acquisition through reading. *Language Teaching Research*, 16, 227–252. https://doi.org/10.1177/1362168811431377

Elgort, I., & P. Warren (2014). L2 vocabulary learning from reading: explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, 64, 365–414. https://doi.org/10.1111/lang.12052

Ellis, N. (2017). Cognition, corpora, and computing: Triangulating research in usage-based language learning. *Language Learning*, 67(S1), 40–65. https://doi.org/10.1111/lang.12215

Eurostat. Being young in Europe (2015). https://ec.europa.eu/eurostat/statistics-xplained/index.php/Being_young_in_Europe_today_-_digital_world

Ericsson, K. A. (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press.

Finnish National Agency for Education (EDUFI). https://www.oph.fi/english

Fitzpatrick, T., & Meara, P. (2009). Exploring the validity of a test of productive vocabulary. *VialVigo International Journal of Applied Linguistics*, 55–74.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35, 305–327. https://doi:10.1093/applin/amt015

Garner, J., Crossley, S., & Kyle, K. (2020). Beginning and intermediate L2 writer's use of N-grams: an association measures study. *International Review of Applied Linguistics in Language Teaching*, 58, 51–74. https://doi.org/10.1515/iral-2017-0089

Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21, 319–333. https://doi.org/10.1017/S0272263199002090

Gill-Flores, J., Torres-Gordillo, J-J., & Perera-Rodríguez, V-H. (2012). The role of online reader experience in explaining students' performance in digital reading. *Computers & Education*, 59, 653–660. https://doi.org/10.1016/j.compedu.2012.03.014

Godwin-Jones, R. (2019). Riding the digital wilds: Learner autonomy and informal language learning. *Language Learning & Technology*, 23, 8–25. https://doi.org/10125/44667

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *ITL International Journal of Applied Linguistics*, 166, 94–126. https://doi.org/10.1075/itl.166.1.03fer

Granger, S. & M. Paquot (eds.) (2012). *Electronic lexicography*. Oxford: Oxford University Press.

Gyllstad, H. (2007). *Testing English Collocations. Developing Receptive Tests for Use with Advanced Swedish Learners*. Lund University.

Halliday, M., & Hasan, R. (1976). *Cohesion in English*. Longman.

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

Hashimoto, B., & Egbert, J. (2019). More than frequency? Exploring predictors of word difficulty for second language learners. *Language Learning*, 69, 839–872. https://doi.org/10.1111/lang.12353

Henriksen, B. (1999). Three dimensions of vocabulary development. *SSLA*, 21, 303–317.

Henriksen, B. (2006). Exploring the quality of lexical knowledge in the language learner's L1 and L2 (pp. 13─36). In P. Pietilä, P. Lintunen, & H-M. Järvinen (Eds.) *Kielenoppija tänään – Language Learners of Today. AFinLA Yearbook*, *64*. Jyväskylä: AFinLA.

Henriksen, B., & Danelund, L. (2015). Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English (pp. 29–55). In P. Pietilä, K. Doró, & P. Pípalová (Eds.), *Lexical Issues in L2 Writing*. Cambridge Scholars Publishing.

Hicks, T., & Perrin, D. (2014). Beyond single modes and media: Writing as an ongoing multimodal text production (pp. 181–231). In E. Jakobs, & D. Perrin (Eds.), *Handbook of Writing and Text Production*. De Gruyter. https://doi.org/10.1515/9783110220674

Härmälä, M., Huhtanen, M., & Puukko, M. (2014). *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013* (pp. 11–12) [Learning outcomes of English A-course at the end of the comprehensive school]. Helsinki: Kansallinen koulutuksen arviointikeskus. English summary. https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/reference-frameworkofcompetences-for-democratic-culture

Jabbari, N., & Eslami, Z. R. (2019). Second language learning in the context of massively multiplayer online games: A scoping review. *ReCALL*, 31, 92–113. https://doi.org/10.1017/S0958344018000058

Jalkanen, J., & Vaarala, H. (2013). Digital texts for learning Finnish: shared resources and emerging practices. *Language Learning and Technology*, 17, 107–124. http://llt.msu.edu/issues/february2013/jalkanenvaarala.pdf

Jeon, E., & Yamashita, J. (2014). L2 reading comprehension and its correlates: a meta-analysis. *Language Learning*, 64, 160─212. https://doi:10.1111/lang.12034.

Jung, Y., Crossley, S., & McNamara, D. (2015). Linguistic Features in MELAB Writing Task Performances. *CaMLA Working Paper No. 2015─05*. CambridgeMichigan.org. https://www.academia.edu/23431508/_CaMLA_Working_Papers_Linguistic_Features_in_MELAB_Writing_Task_Performances

Kaarakainen, M., Kaarakainen, S., Tanhua-Piiroinen, E.,Viteli, J., Syvänen, A., & A. Kivinen. (2017). Digiajan peruskoulu 2017. Tilannearvio ja toimenpidesuositukset. *Valtioneuvoston selvitys ja tutkimustoiminnan julkaisusarja 72/2017*. [Comprehensive school digitalisation: Status review and recommendations for action for 2017, English abstract]. http://www.utu.fi/fi/Ajankohtaista/Uutiset/Documents/Digiajan%20peruskoulu%202017%20%E2%80%93%20Tilannearvio%2 0ja%20toimenpidesuositukset.pdf.

Kalaja, P., Alanen, R., Palviainen, Å., & H. Dufva (2011). From milk cartons to English roommates: context and agency in L2 learning beyond the classroom (pp. 47–58). In P. Benson & H. Reinders (Eds.), *Beyond the language classroom*. Basingstoke: Palgrave Macmillan.

Kerekes, E. (2014). Using songs and lyrics in out-of-class learning (pp. 33–42). In D. Nunan & J. Richards (Eds.), *Language Learning beyond the Classroom*. New York, London: Routledge.

Kim, M., & Crossley, S.A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and 5 independent writing. *Assessing Writing*, 37, 39–56. https://doi.org/10.1016/j.asw.2018.03.002

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical Sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal* (Boulder, Colo.), 102, 120–141. https://doi.org/10.1111/modl.12447

Krashen, S.D. (1985). *The Input Hypothesis: Issues and Implications*. London: Longman.

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: challenging conventions. *TESOL Quarterly*, 50, 976–987. https://doi:10.1002/tesq.329

Kumpulainen, K., & Mikkola, A. (2016). Toward hybrid learning (pp. 15–37). In E. Elstad, (Ed.), *Educational Technology and Polycontextual Bridging*, Rotterdam, the Netherlands: Sense Publishers.

Kuppens, A. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media, and Technology*, 35, 65–85. https://doi.org/10.1080/17439880903561876

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12–24. https://doi.10.1016/j.jslw.2016.10.003

Kyle, K., Crossley, S. A., & Berger, C. (2017). The tool for the analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*, 50, 1030–1046. https://doi.org/10.3758/s13428-017-0924-4

Lan, Y., Chen, N., Ping, L., & Grant, S. (2015). Embodied cognition and language learning in virtual environments. *Education Technology Research and Development*, 63, 639–644. https://doi.org/10.1007/s11423-015-9401-x

Larsen-Freeman, D. (2019) On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal,* 103 *(Supplement 2019*). https://doi:10.1111/modl.125360026-7902/19/61–79

Larsen-Freeman D (2006) The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27, 590–619.

Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex theory perspective. *Modern Language Journal*, 92, 200–213.

Larson-Hall, J. (2016). *A Guide to Doing Statistics in Second Language Research Using SPSS and R: Vol*. Second Edition. Routledge.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & Paribakht , T. (2008). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48, 365–391.

Laufer, B., & Ravenhorst–Kalovski, G. (2010). Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22, 15─30. http://nflrc.hawaii.edu/rfl

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention. *Language Learning & Technology*, 3, 58–76. http://llt.msu.edu/vol3num2/laufer-hill/index.html

Laufer, B., & Levitzky–Aviad, T. **(**2006). Examining the effectiveness of "bilingual dictionary plus" – a dictionary for production in a foreign language. *International Journal of Lexicography*, 19, 135─155. http://doi:10.1093/ijl/eck006

Laufer, B., & Yano, Y. (2001). Understanding unfamiliar words in a text: do L2 learners understand how much they don't understand? *Reading in a Foreign Language*, 13, 549–564.

Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research*, 21, 118–135. https://doi.org/10.1177%2F1362168816637381

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, 105–129. https://doi.org/10.1080/14790718.2012.679275

Liou, H. (2000). The electronic bilingual dictionary as a reading aid to EFL learners: research findings and implications. *Computer Assisted Language Learning*, 13, 467–476. Web.

Kim, M., & Crossley, S.A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and 5 independent writing. *Assessing Writing*, 37, 39–56. https://doi.org/10.1016/j.asw.2018.03.002

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical Sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal* (Boulder, Colo.), 102, 120–141. https://doi.org/10.1111/modl.12447

Krashen, S.D. (1985). The Input Hypothesis: Issues and Implications. London: Longman. Kumpulainen, K., & Mikkola, A. (2016). Toward hybrid learning (pp. 15–37). In E. Elstad, (Ed.), E*ducational Technology and Polycontextual Bridging*, Rotterdam, the Netherlands: Sense Publishers.

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: challenging conventions. *TESOL Quarterly, 50*, 976–987. https://doi:10.1002/tesq.329

Kuppens, A. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media, and Technology*, 35, 65–85. https://doi.org/10.1080/17439880903561876

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12–24. https://doi.10.1016/j.jslw.2016.10.003

Kyle, K., Crossley, S. A., & Berger, C. (2017). The tool for the analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*, 50, 1030–1046. https://doi.org/10.3758/s13428-017-0924-4

Lan, Y., Chen, N., Ping, L., & Grant, S. (2015). Embodied cognition and language learning in virtual environments. *Education Technology Research and Development*, 63, 639–644. https://doi.org/10.1007/s11423-015-9401-x

Larsen-Freeman, D. (2019) On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal,* 103 *(Supplement 2019*). https://doi:10.1111/modl.125360026-7902/19/61–79

Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex theory perspective. *Modern Language Journal*, 92, 200–213.

Larson-Hall, J. (2016). *A Guide to Doing Statistics in Second Language Research Using SPSS and R: Vol*. Second Edition. Routledge.

Laufer, B. (2008). The development of passive and active vocabulary: Same or different? *Applied Linguistics*, 19, 255–271.

Laufer, B., & T. Paribakht (2008). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48, 365–391.

Laufer, B., & G. Ravenhorst–Kalovski (2010). Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22, 15─30. http://nflrc.hawaii.edu/rfl

Laufer, B., & M. Hill (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention. *Language Learning & Technology*, 3, 58–76. http://llt.msu.edu/vol3num2/laufer-hill/index.html

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & T. Levitzky–Aviad. (2006). Examining the effectiveness of "bilingual dictionary plus" – a dictionary for production in a foreign language. *International Journal of Lexicography*, 19, 135─155. http://doi:10.1093/ijl/eck006

Leu, D.J., McVerry, J. G., O'Byrne, W. I., Kiili, C., Zawilinski, L., Everett-Cacopardo, H., Kennedy, C., & Forzani, E. (2011). The new literacies of online reading comprehension: expanding the literacy and learning curriculum. *Journal of Adolescent & Adult Literacy*, 55, 5–14. https://DOI:10.1598/JAAL.55.1.1

Leu, D.J., Kinzer, C.K., Coiro, J., Castek, J., & Henry, L.A. (2013). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. In D.E. Alvermann, N.J. Unrau, & R.B. Ruddell (Eds.), *Theoretical Models and Processes of Reading* (6th ed., pp. 1150–1181). Newark, DE: International Reading Association. http://doi:10.1598/0710.42

Lin, P. (2014). Investigating the validity of the internet television as a resource for acquiring L2 formulaic sequences. *System*, 42, 164–176. https://doi.org/10.1016/j.system.2013.11.010

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, 105–129. https://doi.org/10.1080/14790718.2012.679275

Liou, H. (2000). The electronic bilingual dictionary as a reading aid to EFL learners: research findings and implications. *Computer Assisted Language Learning*, 13, 467─476. https://dx.doi.org/10.1076/0958-8221(200012)13:4-5;1-E;FT467

Lowie, W., Caspi, T., van Geert, P., & Steenbeek, H. (2011). Modeling development and change (pp. 99–121). In M. Verspoor, K. de Boot, & W. Lowie (Eds.) *A dynamic approach to second language development. Methods and techniques*. John Benjamins Pub.Co.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.

Manchón, R. M. (2012). *L2 writing development: Multiple perspectives*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781934078303

Marek, M.W., & Wu, W.-C.V. (2014). Environmental factors affecting computer assisted language learning success: a Complex Dynamic Systems conceptual model. *Computer Assisted Language Learning*, 27, 560─578. https://doi.org/10.1080/09588221.2013.776969

McNamara, D., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43. http://www.jstor.org/stable/3233687

McQuillan, J. (2016). Time, texts, and teaching in vocabulary acquisition: A rebuttal to Cobb. *Reading in a Foreign Language*, 28, 307–318. https://www.proquest.com/scholarly-journals/timetexts-teaching-vocabulary-acquisition/docview/1850833419/se-2

McQuillan, J., & Krashen, S.D. (2008). Commentary: can free reading take you all the way? A response to Cobb. *Language Learning and Technology*, 12, 104–108. http://llt.msu.edu/vol12num1/mcquillan

Meara, P. (2006). Emergent properties of multilingual lexicons. *Applied Linguistics*, 27, 620–644. http://doi:10.1093/applin/aml030

Meara, P. (2009). *Connected Words. Word Associations and Second Language Vocabulary Acquisition*. John Benjamins Pub.Co.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol, Buffalo & Toronto: Multilingual Matters.

Ministry of Education and Culture; https://minedu.fi/en/comprehensive-school

Mutta, M., Pelttari, S., Salmi, L., Chevalier, A., & Johansson, M. (2014). Digital literacy in academic language learning contexts: Developing information-seeking competence (pp. 227–244). In J. Pettes Guikema & L. Williams (Eds.) *Digital literacies in foreign and second language education*. CALICO Monograph Series, Vol. 12. Texas State University: Computer Assisted Language Instruction Consortium (CALICO). https://www.researchgate.net/publication/279840004_Digital_Literacy_in_Academic_Language_Learning_Contexts_Developing_Information-Seeking_Competence

Mäntylä, K., Lahtinen, S., & Toropainen, O. (2020). Processing L2 vocabulary in writing (pp. 107–126). In P. Booth & J. Clenton (Eds.) *First language influences on multilingual lexicons* (1st ed.). Routledge.

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12─25.

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

Nation, P. (2006). How large a vocabulary is needed for reading and listening. *The Canadian Modern Language Review*, 63, 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed). Cambridge: CUP. https://doi.org/10.1017/CBO9781139858656

Nation, P. (2014). How much input do you need to learn the most frequent 9000 words? *Reading in a Foreign Language*, 26, 1–16. https://www.proquest.com/scholarly-journals/how-much-input-doyou-need-learn-most-frequent-9/docview/1705666520/se-2

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, 27, 136–145. https://www.proquest.com/scholarly-journals/principles-guidingvocabulary-learning-through/docview/1705665696/se-2

Nation, P. (2016). Response to Tom Cobb. *Reading in a Foreign Language*, 28, 305–306. https://www.proquest.com/scholarly-journals/response-tom-cobb/docview/1850833186/se-2

Nesi, H. (1998). Defining a shoehorn: learners' dictionary entries for concrete nouns (pp. 159─178). In Atkins (Ed.), *Using dictionaries: Studies of dictionary use by language learners and translators*. Tubingen: Niemeyer.

Nesi, H. (2012). Alternative e-dictionaries: Uncovering dark practices (pp. 363─378). In S. Granger & M. Paquot (Eds.), *Electronic lexicography*. Oxford: Oxford University Press.

Nesi, H., & Meara, P. (1994). Patterns of misinterpretation in the productive use of EFL dictionary definitions. *System*, 22, 1–15.

Niitemaa, M. L. (2020). Informal acquisition of L2 English vocabulary. Exploring the relationship between online out-of-school exposure and words at different frequency levels. *Nordic Journal of Digital Literacy*, 2, 86–105. https://doi.org/10.18261/issn.1891-943x-2020-02-02

Niitemaa, M. L. (2022). Cohesion in Finnish EFL essays: Digital analyses and observations on the use of online sources. *EduLingua*, 8, 1–16. https://DOI:10.14232/edulingua.2022.1.1

Niitemaa, M. L., & Pietilä, P. (2018). Vocabulary skills and online dictionaries: A study on EFL learners' receptive vocabulary knowledge and success in searching electronic sources for information. *Journal of Language Teaching and Research*, 9, 453–462. DOI: http://dx.doi.org/10.17507/jltr.0903.02

OECD (2015). *Students, Computers and Learning: Making the Connection*, PISA, OECD Publishing. https://doi.org/10.1787/19963777

Olsson, E., & Sylvén, L.K. (2015). Extramural English and academic vocabulary. A longitudinal study of CLIL and non-CLIL students in Sweden. *Apples – Journal of Applied Language Studies*, 9, 77–103. https://doi.org/10.17011/apples/urn.201512234129

Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, 2, 31–55. http://nflrc.hawaii.edu/rfl

Pelttari, S. & M. Mutta (2014). Language learners' seeking paths and strategies: one dimension of digital literacy (pp. 157─175). In M. Mutta, P. Lintunen, I. Ivaska, & P. Peltonen (Eds.), *Tulevaisuuden kielenkäyttäjä. Language users of tomorrow*. AFinLA Yearbook**.** Jyväskylä: Jyväskylän yliopistopaino.

Perfetti, C. 2007. Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357─383. http://DOI:10.1080/10888430701530730

Peters, E. (2007). Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning & Technology*, 11, 36–58. http://llt.msu.edu/vol11num2/peters/

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, 142–167. https://DOI.org/10.1075/itl.00010.pet

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40, 551–577.. https://doi.org/10.1017/s0272263117000407

Peters, E.; Noreillie, A-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to FL input on learners' vocabulary knowledge in two languages. *Language Learning*, 69, 747–782. https://doi.org/10.1111/lang.12351

Pietilä, P. (2015). Lexical diversity in L2 academic writing: A look at MA thesis conclusions (pp. 106─125). In P. Pietilä, K. Doró, & P. Pípalová (Eds.), *Lexical Issues in L2 Writing*. Cambridge Scholars Publishing.

Plonsky, L., & Oswald, F. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, 64, 878–912. https://doi.org/10.1111/lang.12079

Rankin, Y., Gold, R., & Gooch, B. (2006). *3D role-playing games as language learning tools*. A paper presented at the EuroGraphics 2006, September 4–8, 2006, Vienna, Austria. http://www.thegooch.org/

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Richards, J.C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77─89.

Ringbom, H. 1983. On the distinctions of item learning vs. system learning and receptive competence vs. productive competence in relation to the role of L1 in foreign language learning. In H. Ringbom (ed.) *Psycholinguistics and Foreign Language Learning*. Publications of the Research Institute of the Åbo Academy Foundation, 86, 163–173.

Ringbom, H. (1991). Crosslinguistic lexical influence and foreign language learning. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood Smith & M. Swain. (eds.) *Foreign/Second Language Pedagogy Research*. Multilingual Matters: Clevedon, England, pp. 172–181.

Ryshina-Pankova, M. (2015). A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor. *Journal of Second Language Writing*, 29, 51–63. http://dx.doi.org/10.1016/j.jslw.2015.06.005

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129─158. https://doi.org/10.1093/applin/11.2.129

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363. https://doi.org/10.1177%2F1362168808089921

Schmitt, N. (2010). *Researching Vocabulary. A Vocabulary Research Manual* (1st ed. 2010.). London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230293977

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55─88.

Schmitt, N. & Schmitt, D. (2014). Reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47, 484–503. https://doi:10.1017/S0261444812000018.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, 18, 55–88. http://dx.doi.org.ezproxy.utu.fi/10.1191/026553201668475857

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95, 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Segalowitz, N. (2003). Automaticity in second languages (pp. 382–408). In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition*. Malden (MA): Blackwell Publishing.

Shebilske, W., Goettl, B., & Regian, J.W. (1999). Executive control of automatic processes as complex skills develop in laboratory and applied settings (pp. 401–432). In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: interaction of theory and application* (pp. 401–432). Cambridge, MA: MIT Press.

Sigman, M., & Cecchi, G.A. (2002). Global organisation of lexicon. *PNAS*, 99, 1742–1747. https://langev.com/author/msigman.

Singer, G., Danilov, D., & U. Norbisrath (2012). Complex search: aggregation, discovery, and synthesis. *Proceedings of the Estonian Academy of Sciences*, 61, 89–106. https://doi.10.3176/proc.2012.2.02

Sinharay, S., Mo Zhang, & Deane, P. (2019) Prediction of Essay Scores From Writing Process and Product Features Using Data Mining Methods. *Applied Measurement in Education*, 32, 116–137, https://DOI:10.1080/08957347.2019.1577245

Sockett, G., & Toffoli, D. (2012). Beyond learner autonomy: a dynamic systems view of the informal learning of English in virtual online communities. *Procedia. Social and Behavioral Sciences*, 34, 212–215. https://doi.org/10.1016/j.sbspro.2012.02.042

Statista, 2015. Statistics Finland. (n.d.). Internet usage for searching for information on Wikipedia or other online encyclopaedias in Finland in Q1 2015, by age group. Statista – The Statistics Portal. https://www.statista.com/statistics/551756/internet-usage-for-searching-for-information-byagegroup/

Statista, 2016. Statistics Finland. (n.d.). Internet usage for listening to music in Finland in Q1 2016, by age group. In Statista – The Statistics Portal. Retrieved March 23, 2017, from https://www.statista.com/statistics/552155/internet-usage-for-listening-to-or-downloading-musicby-age-group/

Statistics Finland. (n.d.). Internet usage for searching for information on Wikipedia or other online encyclopaedias in Finland in Q1 2015, by age group. In Statista – The Statistics Portal. https://www.statista.com/statistics/551756/internet-usage-forsearching-for-information-by-agegroup/

Sundqvist, P. (2009). *Extramural English Matters: Out-of-school. English and its Impact on Swedish Ninth Graders' oral proficiency and vocabulary*. Karlstad: Karlstad University Studies.

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

Sundqvist, P. (2022). Learning across the lifespan. Age, language learning and technology (pp. 343–355). In N. Ziegler & M. Gonzalez-Lloret (Eds.), *The Routledge Handbook of Second Language Acquisition and Technology*. Routledge. https://DOI:10.4324/9781351117586-30

Sundqvist, P., & Olin-Scheller, C. (2013). Classroom vs. extramural English: teachers dealing with demotivation. *Language and Linguistics Compass*, 7, 329–338. https://doi.org/10.1111/lnc3.12031

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

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

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development (pp. 235–253). In S. Gass & C. Madden (Eds.) *Input in Second Language Acquisition*. Rowley MA: Newbury House

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning (pp. 19–30). In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press.

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

Toffoli, D., & Sockett, G. (2010). How non-specialist students of English practice informal learning using Web 2.0 tools. ASp: *La revue du GERAS*, 58, 125–144. https://doi.org/10.4000/asp.1851

Tono, Y. (2011). Application of eye-tracking in EFL learners' dictionary look-up process research. *International Journal of Lexicography*, 24, 124–153. https://doi:10.1093/ijl/ecq043.

Tumelius, R., & Kuure, L. (2021). Towards a shared vision of language, language learning, and a school project in emergence. *Classroom Discourse*, 12, 344–364. https://doi.org/10.1080/19463014.2020.1808495

Usoof, H., Leblay, C., & Caporossi, G. (2020). GenoGraphiX-Log version 2.0 user guide. *Les Cahiers Du GERAD*, 2020, 68, 1–63. https://www.gerad.ca/en/pa

Van Dijk, J. & A. Van Deursen. (2014). *Digital skills. Unlocking the information society*. Basingstoke: Palgrave Macmillan.

Vandergriff, I. (2016). *Second-language Discourse in the Digital World: Linguistic and Social Practices in and beyond the Networked Classroom*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal*, 92, 214–231. Web.

Verspoor, M. H., 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.), *English in Europe today: Sociocultural and Educational Perspectives* (pp. 147–166). Amsterdam; Philadelphia: John Benjamins Pub.Co.

Verspoor, M., Schmid, M., & Xu Xiaoyan (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21, 239–263. https://doi:10.1016/j.jslw.2012.03.007

Voulgari, I., Komis, V., & Sampson, D.G. (2014). *Learning outcomes and processes in massively multiplayer online games: exploring the perceptions of players*. Education Technology Research Development, 62, 245–270. Web.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *SLA*, 30, 79–95. https://doi.org/10.1017/S0272263108080042

Webb, S., & Rodgers, M.P.H. (2009a). The vocabulary demands of television programs. *Language Learning*, 59, 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Webb, S. & Rodgers, M.P.H. (2009b). The lexical coverage of movies. *Applied Linguistics*, 30, 407–427. https://doi.org/10.1093/applin/amp010

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: the role of L1 lexical/conceptual knowledge. *Applied Linguistics*, 27, 741─747. Web.

Wray, A. (2008). Formulaic Language: Pushing the Boundaries. Oxford: Oxford University Press.

Xue Guo-yi, & Nation, P. (1984). A university word list. *Language Learning and Communication*, 3, 93─242.

Zheng, D., Bischoff, M., & Gilliland, B. (2015). Vocabulary learning in massively multiplayer online games: context and action before words. *Education Technology Research Development*, 63, 771–790. https://doi.org/10.1007/s11423-015-9387-4

Online resources

Lexical tests available (Meara, P.) https://www.lognostics.co.uk/tools/index.htm
Lex30 (*Fitzpatrick & Meara*, *2009*). https://www.lognostics.co.uk/tools/index.htm
Vocabulary Levels Test. https://www.norbertschmitt.co.uk/vocabulary-resources
TAACO. https://www.linguisticanalysistools.org/taaco.html
TAALES. https://kristopherkyle.github.io/professional-webpage/docs/tools
Vocabulary game Freerice (https://unric.org/en/freerice/)
Vocabprofile Compleat Lextutor (Cobb, T., n.d.) www.lextutor.ca.
World Food Programme. (n.d.) FreeRice. https://unric.org/en/freerice/

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2024

ISBN 978-951-29-9597-4 (PRINT)
ISBN 978-951-29-9598-1 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)