Cohesion in Finnish EFL essays: Digital analyses and observations on the use of online sources

Marja-Leena Niitemaa[1]

School of Languages and Translation Studies, Department of English, University of Turku, Finland

The study investigated cohesion in Finnish upper-secondary school EFL learners' essays (N=46). Cohesive devices were digitally identified using TAACO 2.0.4, and robust correlations were run to examine how the devices related to human-rated holistic essay quality. The analyses found that the two most important predictors of writing quality were the use of modifying adverbs and adverbials as referential devices across paragraphs, and a wide array of connectives to organise the text. Further, the writing sessions were video-recorded to examine the role of consulting digital sources in cohesion-building. The recorded data suggested that consulting online dictionaries and informational pages assisted cohesion-building if the writer possessed adequate vocabulary knowledge and computer skills and knew how to exploit the sources efficiently. Pedagogically, the findings indicated that learners need more instruction and practice not only on writing cohesive texts but also on how to search for information and lexis effectively.

Keywords: cohesion, EFL essays, online dictionaries, video-recorded data, writing process

1. Introduction

In written language, cohesive features are particularly important for readers with insufficient vocabulary or background knowledge (cf., McNamara, Kintsch, Songer, & Kintsch, 1996), while in speech, cohesive markers can decrease misunderstandings between first (L1) and second language (L2) speakers (Crossley, Salsbury, & McNamara, 2010). Although crucial in all communication (e.g., Lintunen, Mutta, & Peltonen, 2020; Council of Europe, 2018b), learning to employ cohesive devices is one of the major challenges for learners of English as a foreign language (EFL).

In Finland, the L2 curricula are based on the guidelines of the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2018a). Overall, CEFR emphasizes the ability to produce clear, detailed text on a variety of topics. Accordingly, Finnish upper-secondary school students should reach the CEFR level B.2 in EFL written production by the national school-leaving examination. Regarding cohesion, B2 descriptors highlight the ability to employ cohesive devices to produce clear, coherent text following the established conventions of the genre. This includes, e.g., structuring texts in paragraphs, using

---

[1] Author's e-mail: maleni@utu.fi; https://orcid.org/0000-0002-2822-2735

contextually appropriate lexis, avoiding errors disrupting readability, and linking clauses and paragraphs with a wide range of connectives (Council of Europe, 2018a). All these skills take a long time to develop for EFL learners. Moreover, language teachers often consider cohesion one of the most strenuous issues to teach, e.g., Finnish learners often forget to write EFL essays in paragraphs, although they do so in L1.

Better understanding of EFL writing processes may offer new insights into approaching cohesion in the classroom. The present study sets out to examine, which cohesive features characterize Finnish upper-secondary school EFL learners' essays and how the CEFR descriptions for cohesion are met at level B2, i.e., the level that Finnish students are expected to reach before taking the national school-leaving examination in English. Another aim is to examine the role of consulting online dictionaries and other digital sources in cohesion-building. For these purposes, cohesive devices are digitally identified and compared against the human-rated holistic writing scores and the CEFR requirements at level B2, and the authentic writing sessions are video-recorded to examine consulting the Internet during writing. Further, the rate of successful consultations of digital sources is compared against the writer's vocabulary knowledge.

2. Literature review

Cohesion commonly denotes linguistic signposts that help readers notice relationships between the ideas and information presented in texts. In the framework of Halliday and Hasan (1976), cohesive features are categorized to referential devices (e.g., demonstrative reference by *this/these, similarly/otherwise*); substitution (replacing words instead of repeating them); ellipsis (omitting words); connective links within and between clauses and paragraphs; and lexical cohesion, i.e., using different but semantically related words and collocations. Recent studies have employed digital tools to identify cohesive features in tertiary-level EFL students' texts and to examine how such features relate to holistic writing quality and text organization, and moreover, investigated the writing process in real time to develop teaching and feedback practices, which could help writers produce cohesive texts.

2.1 Cohesion in relation to essay quality and text organization

This subsection introduces four studies using digital analyses to examine cohesion in EFL texts. The first two focus on the longitudinal development in using cohesive devices in relation to human rating, while the last two examine the role of elaboration in teaching cohesion. To facilitate comparison between these and the present results, the review is limited to studies using different versions of TAACO, and the titles for the cohesive devices are italicized. Examples of cohesive features in the present data are provided below in section 4.2.

Crossley, Kyle and McNamara (2016b) examined three sets of 30-minute descriptive essays written by university students attending English for Academic Purposes courses within a semester. The analyses conducted by TAACO 1.0 (Crossley, Kyle, & McNamara, 2016a) showed that 44% of the variance of holistic essay ratings was collectively explained by *Function words* across paragraphs and sentences, and *Pronouns* across paragraphs and the whole text. Regarding text organization, the best predictors were *Function words* at the paragraph and text levels and *Pronouns* and *Coordinating conjuncts* between sentences

explaining 36% of the variance of the scores. In these analyses *Function words across paragraphs* appeared to be the best single predictor. To explain this, the researchers suggested that human raters may show bias for organizational devices in L2 writing, since content may not be as rich as in L1 essays. As for longitudinal development, the analyses reported significant growth over a semester for about a half of the cohesive features examined such as *Nouns* and *Synonyms across paragraphs*. However, the increased occurrences did not necessarily correlate with human ratings.

Kim and Crossley (2018) analyzed the joint effect of lexical, syntactic, and cohesive variables on 30-minute argumentative essays and 20-minute source-based texts using structural equation modelling (SEM). The essays were written by tertiary-level students with diverse L1 backgrounds. Four indices (TAACO 1.0) were employed to examine cohesion: *Overlap of lexis across sentences and paragraphs*, and the incidence of *Positive* and *Negative logical connectives*. These devices are thought to increase readability via referential links. Overlap, i.e., repetition of the same lexical items, helps readers notice connections between the ideas presented, while connectives provide explicit links to them.

The results also showed that sentence-level, *Overlap of lexis* correlated positively with the scores of source-based texts but not with the argumentative essays, while *Overlap of lexis across paragraphs* correlated positively with the scores of both types of writing and thus, was selected to measure cohesion in the SEM analysis. This index together with lexical sophistication and syntactic complexity explained 82% of the variance of the scores of both argumentative and source-based essays. However, lexical sophistication and syntactic complexity explained a greater share of the variance compared to referential cohesion.

Two studies, Crossley and McNamara (2016), and Crossley, Kyle and Dascalu (2019), examined whether elaboration can be used to draw EFL writers' attention to cohesion during writing. Both the studies used the same essays written by American university students. The texts were written on computers but using notes or the Internet was not allowed. When ready, the participants were to spend 15 minutes to elaborate the main ideas by adding two more paragraphs. The same procedure was repeated with a new prompt so that each writer produced two original essays and two elaborated versions. Next, an expert was asked to manipulate the texts and increase referential cohesion via lexical overlap across the text segments. Finally, the original essays, elaborated texts and manipulated versions were digitally analysed for cohesion.

In the earlier study from 2016, all the versions were analyzed for *Lexical overlap across sentences and paragraphs* using TAACO 1.0. The analyses indicated that the elaborated essays did not score significantly higher than the original ones, whereas the essays with expert-added cohesion scored higher than the original and elaborated versions. The index measuring *Lexical overlap between paragraphs* was a strong predictor of essay scores. Analysing the same texts, the study from 2019 employed measures of *Semantic similarity* provided by TAACO 2.0. Contrary to the earlier results, the elaborated versions scored higher points than the original ones, but the expert-added essays still scored the highest. *Semantic similarity (word2vec)* across paragraphs appeared to be an important predictor of the essay ratings.

In sum, the TAACO analyses indicate that higher-rated writing is associated with two cohesive features, referential cohesion, i.e., the use of pronouns and lexical repetition, and organizational tools, such as connectives and function words. These devices are thought to enhance comprehensibility of the text. The researchers suggest, however, that different textual genres may require different cohesive means.

2.2 Cohesion-building in the writing process

Employing different methods, the following three studies examine EFL learners' text production processes in order to improve writing assessment, teaching practices and feedback. The essays analysed in these studies were written on computers without access to the Internet.

Bowen and Thomas (2020) investigated in real time how L1 and L2 students developed clauses, added information, and used lexis to refer backward and forward across the paragraphs. The participants were three L1 and three L2 students (Chinese) at a British university, the latter scoring intermediate points on the International English Language Testing System. Using keystroke logging (Inputlog), the researchers analysed writing essays which were to include description of the data presented in charts, responding to the prompt, and revising the text by adding information. The data was then used to examine how the text evolved from one section to another. The findings suggested that the writers could produce informationally dense and interconnected texts when revising but that L1 and L2 students employed different cohesive means: L1 writers preferred substitution and modification of complex noun groups, while L2 students relied on coordinating conjunctions and demonstrative pronouns.

Abdel Latif (2021) analysed the basic components of text production and composing strategies among a group of 30 university students with Arabic as L1 and intermediate proficiency in English. The participants were asked to write an argumentative essay, as this text type requires more textual organization compared to other genres. The researcher employed think-aloud data to investigate how the writing process evolved. The participants were first trained on how to verbalize and record their concurrent actions during writing. The transcribed think-aloud data showed that cohesion emerged via adding, changing, or deleting words and phrases, in other words, using referential devices, substitution, ellipsis, and connectors. The findings also suggested that such revisions were associated with the writers' linguistic resources, as proficient writers were able to monitor and revise their texts, whereas less skilled writers had difficulties in finding alternative choices for words and expressions and needed to verify L2 meanings using L1.

Lyashevskaya, Panteleeva and Vinogradova (2021) examined features of lexical, morphological, syntactic, and discursive complexity in EFL texts written by Russian university students. The aim was to develop a digital feedback application to provide recommendations for revising the text and alerting for L1 interference. For this purpose, the researchers analysed over 3000 essays including descriptions of graphical material and expressing opinions on social and cultural problems. The texts were then divided in the best and non-best essays assessed by human raters. The findings indicated that discursive complexity emerged from three cohesive features: the use of discourse-organizing nouns, i.e., semantically unspecific abstract nouns like fact, issue, and argument referring to information given in different text sections, multi-word connectors, e.g., *on the other hand*, or *in my opinion*, and single-word connectives to communicate addition, cause and effect, clarification, or contrast such as moreover or consequently. Overall, the best essays contained more discourse-organizing nouns and diverse linking tools compared to the non-best essays.

To sum up, the findings suggest that writing quality is strongly associated with cohesion. Research on various aspects of the writing process and findings on text complexity indicate that higher-scoring texts progress logically from one topic to another and employ multiple types of

cohesive devices, while lower-scoring texts demonstrate difficulties in connecting ideas across paragraphs, although the learners manage writing at the sentence level.

3. The present study

The present study aims to contribute to cohesion research in two respects. Firstly, we examine essays written by upper-secondary school learners, while previous research mostly focuses on writing at tertiary-level. Secondly, the participants are allowed to use the Internet during writing, as there is an overall scarcity of examinations simulating authentic writing sessions with free access to online dictionaries and webpages.

Subsection 4.1 reports on how digitally identified cohesive features are connected to holistic writing quality in Finnish secondary-school EFL essays, and how the CEFR expectations for cohesion at level B2 are met (RQ 1). Subsection 4.2. focuses on the role of consulting the Internet in cohesion-building. As employing cohesive devices may be connected to EFL learners' lexical development (e.g., Crossley et al., 2016b) and linguistic resources (e.g., Abdel Latif, 2021)**,** successful use of digital sources is compared against the writer's receptive vocabulary knowledge (RQ 2). The research questions are formulated as follows:

RQ 1. Which cohesive features characterize Finnish upper-secondary school EFL essays? How do the essays fulfill the CEFR expectations for cohesion at level B2?

RQ 2. Under what conditions can using digital sources enhance cohesion?

3.1 Participants

The participants were 46 students, aged 16−17 with Finnish as L1, at a typical Finnish municipally maintained upper-secondary school. They volunteered to be tested for lexical knowledge and writing tasks during the first two academic years. Previously they had studied L2 English circa 600 lessons (45 min.) in the general basic education and circa 120–150 lessons (75 min.) during the first and second year at the upper-secondary level. To persuade the students to take all the tests, they were offered a credit of one course in English. It was also agreed that the test performance would not affect their English grades and that the tests would be conducted during school hours. The present examination is based on the data from the second year.

3.2 Essays and scoring

The participants were asked to write an essay with 150–250 words on "My second school year". They were encouraged to regard this task as an additional opportunity to train their writing skills before the national high-stakes examination in the following year. The prompts included, e.g., discussing their academic success and expectations in front of the upcoming school-leaving examinations as well as their future plans for the tertiary level. To hinder priming effects, the prompts were given in L1 using 3−5 bullet points. The writing time was 60 minutes, as the participants were expected to use online sources to search for lexis and information and to revise their texts using editing functions. They were also informed that the essays would be checked for plagiarism.

The essays were rated on a four-point scale from 0 to 99 points (e.g., 80−82−85−88 at the upper-intermediate level) accounting for the content and structure of the text, lexical richness and accuracy, and the candidate's ability to communicate the message clearly. The same criteria are used in the national Matriculation Examination. The raters were twenty-eight teacher trainees finishing their studies at the Faculty of Education at a Finnish university as a part of curricular training. The trainees first assessed the essays on their own and then discussed the assessments in small groups. They were, however, encouraged to give the scores independently. The interrater reliability (Cronbach's alfa) was strong ranging from 96% to 99%.

Before running the automated analyses, the essays were cleaned for spelling errors that changed the word meaning, e.g., *taught* instead of *thought*. The raters also used the cleaned version, as the purpose was to draw attention to the structure and cohesiveness of the texts instead of accuracy.

3.3 Recording of the writing session

A freely downloadable video-recording software, CamStudio, was installed in the computers to record the individual writing processes. The participants were first shown how to switch on the software and then asked to start writing as usual. The recording is unobtrusive for the writer. We examined the recorded data to monitor which online sources were used, how many times each source was consulted, what was searched, how the writers used the findings, and evaluated whether the search results suit the context. However, due to space problems in the computer rooms, we were able to record only thirty-one students out of forty-six.

3.4 Vocabulary knowledge

To examine the use of online sources in relation to vocabulary knowledge, the participants were assessed for receptive vocabulary knowledge using the revised version of the Vocabulary Levels Test (the *VLT*; Schmitt, N., Schmitt, D., & Clapham, C., 2001; Nation, 1983) with 30 items in the 2nd, 3rd, 5th, and 10th thousand frequency bands (maximum 120 points). Each frequency band consists of ten item groups with six words and three definitions. The task is to match the words to the definitions. The *VLT* is commonly considered a reliable and valid measure of receptive vocabulary (e.g., Meara, 2009; Read, 2000). The test sections are scalable, i.e., the knowledge of rare words implies knowledge of more familiar words. The test administration coincided with the writing session. The rationale for using the *VLT* in a cohesion study is that receptive vocabulary size is closely connected to EFL learner's ability to use English in multiple ways (Alderson, 2005; Schmitt et al., 2001) and thus, may also interact with cohesion (Crossley et al., 2016b).

3.5 Automated analysis

To analyze the essays for cohesive features, the present study employed TAACO 2.0.4 (Crossley, Kyle, & Dascalu, 2019). It is a freely accessible tool designed to detect cohesion across sentences (local cohesion), paragraphs (global cohesion) and the entire text (text cohesion). TAACO also provides diagnostic output files to show how the words in the

individual essays are tagged. For detailed information on calculation of the indices, please see https://www.linguisticanalysistools.org/taaco.html.
The present study analysed 123 indices:

> • 60 indices at the sentence level: 48 lexical overlap indices, 2 semantic overlap indices, the incidence of 10 types of connectives
> • 50 indices at the paragraph level: 48 lexical overlap indices and 2 semantic overlap indices
> • 13 indices at the text level: measuring type-token ratios, determiners, demonstratives, pronoun to noun ratio and pronoun density

Lexical overlap calculates repetition of words across sentences and paragraphs, while *Semantic overlap* measures repetition of semantically related words such as synonyms for nouns and verbs. At the text level, cohesion is measured by the incidence of determiners, demonstratives and type-token ratios (TTR) of various parts of speech. The latter, lexical diversity, signifies referential cohesion.

3.6 Statistical analyses

Robust bootstrapped correlations (Larson-Hall, 2016, 213–214) were computed to examine the strength of connections between TAACO indices and essay ratings. Robust tests provide confidence intervals (CI), which indicate that the actual correlation coefficient is within the bounds of the CI with a probability of 95%. Although wider in smaller samples, the CI indicates that the correlation is statistically significant if it does not pass through zero. The values for the effect (R^2) are interpreted according to Plonsky and Oswald's (2014) guidelines: $R^2 = 0.06$ is small, $R^2 = 0.16$ is medium and $R^2 = 0.36$ is large. Robust multiple linear regression analyses were run to identify which cohesive indices were predictive of human-rated essay scores. Robust tests do not assume normal distribution, but the assumptions of regression, including linearity, normal distribution of errors, homogeneity of variances and multicollinearity were checked (Larson-Hall, 2016, 251). The analyses were conducted using SPSS, version 27.

4. Results

The digital analyses were based on 46 essays comprising 10.835 words. On the Finnish national assessment scale (c.f., 3.2.), the mean score approached the upper-intermediate level (mean 77.7; SD 9.2). Regarding the CEFR descriptors (Council of Europe, 2018a), the essay scores ranged between C1 and B1 so that 15% reached the level C1, 46% were at the level B2, while 39% remained at B1. Regarding the *VLT* results, the mean score (77.48) was 65% of the maximum points, which is on par with multiple results among the same age group (e.g., Peters, 2018). However, the standard deviation (SD 30.18) was exceptionally wide, as 30% of the participants scored 50% or less, while 33% of them scored 80% or more.

4.1 Which cohesive features characterize Finnish upper-secondary school EFL essays? How do the essays fulfill the CEFR expectations for cohesion at level B2? (RQ 1)

The essays were analysed for the 123 indices introduced above (c.f., section 3.5), and robust bootstrapped correlations (c.f., section 3.6) were conducted to examine whether the essay scores (dependent variable) were related to the cohesive features (independent variables). The analyses showed that 28 (roughly 23 %) variables were significantly correlated with the essay scores including 25 paragraph-level measures and three sentence-level measures, while text-level features, such as the incidence of determiners and demonstratives, did not correlate significantly with writing quality.

Fifteen indices (c.f., Appendix) with at least medium effect sizes ($R^2 \geq .160$) were chosen for further analyses. After checking for multicollinearity ($r \leq .700$), a series of multiple linear regression analyses with two independents were conducted using robust tests. The calculations were checked for outliers, homogeneity of variances, and normality and independence of residuals. The results indicated that two measures, *Adverb 2 paragraph normed* and the incidence of *Conjunctions and/but*, collectively reported the best significant model (F (3.42) = 14 008, $p < .001$, $R^2 = .37$). The coefficients are provided in Table 1. The confidence intervals inform that the association is statistically significant. The other combinations with two independents yielded effect sizes from 30% to 34%, whereas the analyses with three independent variables did not report significant equations.

Table 1. Coefficients in the robust test. N = 46

	B	sig. 2-tailed	BCa 95 % Confidence Interval	
			Lower	Upper
(Constant)	84.583	<.001	77.32	92.04
Adverb 2 paragraphs normed	3.359	<.001	1.86	4.83
Conjunctions *and/but*	–256.958	.001	–380.06	–133.05

To answer RQ 1, the digital analyses indicated that the best predictors of writing quality were using a rich array of adverbs and adverbials as referential devices across paragraphs and a wide range of connective devices. In terms of the CEFR descriptors for cohesion, EFL writers at level C1 knew how to employ referential and lexical cohesion, substitution, and a wide range of connectives, B2 writers were able to avoid errors disrupting readability but used fewer types of connectives, while B1 writers overused the conjunctions and/but and often forgot to structure the text in paragraphs, which diminished the clarity of the text.

4.2 Under what conditions can using online affordances enhance cohesion? (RQ 2)

Based on the video-recorded data of authentic writing sessions of thirty-one upper-secondary school EFL students (cf., 3.3), this subsection reports how the participants employed online dictionaries and informational sites, what they searched, what they found, and what problems they encountered during consultations. We also discuss how the online findings were related to

the cohesive indices identified by TAACO and what role the writer's lexical knowledge played in searching for words, expressions, and information online.

Number of consultations, sources, and the search language

The writers consulted online sources 341 times (Table 2). Eight writers conducted from one to five queries, thirteen consulted from six to fifteen times and nine searched from sixteen to 26 times, which was the maximum number of individual queries. Those who queried frequently returned to the same items several times. One participant did not consult any online sources.

The students employed from one to four different sources. These included freely available multilingual dictionaries, translation tools and informational sites like Wikis and home pages of educational institutions. The most frequently consulted sources were Sanakirja.org[2] (174 queries) and Google Translate (140 queries). The majority used only one source so that fourteen chose Sanakirja.org, six writers used Google translate, and four writers employed them both. One writer consulted EUdict.com (four queries) while five writers employed different combinations of dictionaries, translation tools and informational sites, e.g., Wikipedia or home pages of educational institutions (86 queries). No expert-constructed learners' dictionaries were consulted. Roughly 91% of the consultations were conducted from L1 to L2. Sixteen students searched only from L1 to L2, while fourteen also used L2 when crosschecking meanings or consulting informational sites.

Table 2. Sources, consultations, and the search language

Number of writers**	Sources consulted*	Number of all consultations	Number of successful consultations	L1 ->L2	L2 -> L1
14	D	138	107	126	12
6	GT	70	54	69	1
4	D + GT	43 [33 + 10]	24	39	4
1	OD	4	4	3	1
1	GT + K	11 [6 + 5]	10	10	1
1	OD + GT	26 [4 +22]	20	26	0
1	GT + I	19 [13 + 6]	17	19	0
1	D + GT + I	22 [1 +19 + 2]	20	12	10
1	D + OD + I + I	8 [2 + 1 + 5]	8	5	3
Total 30**	1 to 4 sources	341	264 (77%)	309 (91%)	32 (9%)

*D = Sanakirja.org; GT = Google translate; K = Kaannos.com; OD = Other dictionaries; I = Informational sites; **One of the 31 recorded participants did not consult any online sources.

After searching, the writers chose one of the five following actions: found the item, and used it correctly, found the item but used it incorrectly, hesitated and decided not to use a word

---

[2] Information provided by Sanakirja.org and IlmainenSanakirja, and the translation tool Kaannos.com are based on Wiktionary articles.

unknown to them, did not find the item and replaced it with something else or did not find the item and discarded the topic.

What was searched frequently?

The most frequently searched item was the official term for the Finnish national school-leaving examination (*matriculation examination*) searched by twenty students. Fifteen students queried the noun (school) *subject*. Eleven queries were made for tertiary-level schools such as *polytechnic* and *university of applied sciences*.

How did the consultations succeed?

Although consulting the references was not without problems, 77% of the queries were conducted successfully. Thirteen participants found *matriculation examination* either in *Sanakirja.org*, or later by chance in *Wikipedia* when searching for something else. The word seemed unknown to the majority, which manifested as hesitation in the recorded data. For example, three students replaced it by *final test* or the verb *graduate*. Four queries were unsuccessful: one writer did not find the item, another chose *baccalaureate* but used it as a verb, while two writers accepted the literal translation from L1 (*student writings**) by *Google translate*. The noun (school) *subject* was successfully found by eleven students from *Sanakirja.org*. Four students chose the literal translation from L1 (*substance*, *material*) but changed the word after consulting other sources. However, three of them also used the inappropriate words in their texts. Instead of searching for *subject*, five students replaced it with specific terms, such as *history* or *physics*. Twenty-two students searched for adjectives to characterize school subjects or examinations. The most frequently queried adjectives were *compulsory* (3 queries), *mandatory* (7 queries), *obligatory* (1 query), and *optional* (6 queries). Apart from one case, these searches were successful.

What was not searched?

The verbs collocating with *matriculation examination* or *subject* were not searched. The verb choice was strongly affected by L1 interference, as in Finnish, you "write an examination" and "read a 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 *subject*, *Google translate* suggested *read*, e.g., *read subjects** or *read mathematics**. Connective devices were searched rarely. One successful query was made for each of the following connectives: *although*, *compared to*, *even though*, *firstly*, *howeve*r, *nevertheless*, and *unfortunately*. One student searched for the combination *on one hand* without finding it.

Problems in searching

The recordings revealed three major problems: ineffective use of online sources, lack of basic computer skills and inadequate vocabulary knowledge.

Firstly, most participants tended to rely on one source without crosschecking and evaluating the information in the entries. Neither did they consult definitions and examples, which would have provided collocates and information on register. For example, the Finnish counterpart for (school) *subject* is a polysemic word covering such meanings as *matter*, *material* and *substance.* The suggestions in *Sanakirja.org* are provided in decontextualized word lists, which resulted in choosing *material or *substance. The latter was also the primary suggestion by *Google translate*. Moreover, the users of *Google translate* often relied on the suggested literal translations. For example, aiming to express "*I am going to take the matriculation examination in the fall*," one student copied the suggested phrase "*I am going to write a fall in the matriculation examination.*" Another student intended to communicate his future plans, which *Google translate* formulated as "*I read myself a building engineer.*" There were only few cases in which changing the search term helped the writers to find the appropriate target.

Secondly, some participants lacked basic computer skills. Instead of using the copy-and-paste function, some students navigated back and forth between the source and the task copying one word at a time, while more skilled students had the primary source opened in its own window next to the writing task allowing a quick navigation between the task and the sources. Regarding misspelt words, most writers corrected the errors when flagged. A few students corrected only some of the spelling errors, while two students ignored them all. However, the writers rarely noticed errors that changed the word meaning as the proof-reading function did not flag them, e.g., *chance* instead of *change*, or *curses* instead of *courses*. At times, the dictionary did not provide any suggestions due to spelling errors. Some spelling errors occurred even in L1 queries.

Thirdly, the students with low vocabulary knowledge could not evaluate the words and phrases suggested in the sources as demonstrated in the examples above. Regarding the *VLT* results (c.f., the beginning of section 4), the examples of incorrect choices originated from writers scoring under 64 % or less in the *VLT*. In contrast, the participants scoring 80 per cent or more seemed to know most of the topic-related key words without having to search for them, but when they searched, they knew how to crosscheck the findings.

Lastly, some students stopped querying if the item was not found immediately. This may indicate low motivation due to the problems experienced when consulting the sources. Moreover, the participants knew that their language teacher would not grade the essays.

To answer RQ 2, the video-recordings suggested that consulting online dictionaries and informational pages may help to enhance cohesion if the writer possesses adequate vocabulary knowledge and computer skills and knows how to exploit the sources efficiently. Regarding the framework of Halliday and Hasan (1976), this indicates that consulting online sources facilitates using substitution and lexical cohesion. Substitution allows the writer to exploit a greater range of topic-related lexis, for example, to replace the noun *subject* with more specific terms such as *social studies* or *psychology*. Regarding lexical cohesion, higher-scoring writers searched or checked semantically related words when presenting information (*curriculum, examination, mathematics, matriculation examination, pronunciation, skills*) as well as adjectives to specify meanings (*advanced, basic, complicated, compulsory, optional, oral*).

Regarding the role of connectives and demonstrative reference in cohesion-building, the overuse of conjunctions *and/but* correlated negatively with writing quality and made the texts resemble spoken language, e.g., "*And I have good memories.*" or "*And the second year's courses are a little bit harder*". In contrast, demonstrative reference, such as using adverbs as

modifiers across paragraphs (*especially, definitely, hopefully, luckily, nearly, personally, probably*) was a positive predictor of essay scores. However, higher-scoring writers, who employed a wide range of connectors and adverbs, did not need to search for them.

5. Discussion

The present study on cohesion in EFL writing was grounded on the seminal framework of Halliday and Hasan (1976) presenting five types of cohesive means: employing referential devices, replacing words instead of repeating them, omitting words, linking text segments with connectives, and using different but semantically related words and collocations. Multiple research results have shown that such features facilitate reading comprehension and assist noticing relationships between the ideas and information presented in the text, and that writing quality is strongly related to the writer's ability to produce cohesive text.

The study aimed, firstly, to ascertain which cohesive devices Finnish upper-secondary school EFL students employed in their compositions and analyse how these devices associated with human-rated writing quality. The present findings aligned with the results attained among tertiary-level students (e.g., Crossley et al., 2016b; Kim, & Crossley, 2018) in that features of paragraph-level cohesion were closely related to writing quality. In terms of the framework of Halliday and Hasan (1976), higher-scoring Finnish upper-secondary school writers employed a rich array of adverbs and adverbials as referential devices across paragraphs and linked sentences and paragraphs using a wide range of connectives, i.e., both younger and older EFL students tended to employ cohesive devices related to textual organization. In contrast to Crossley et al. (2016b), no connection was found between writing scores and the type-token ratios or the use of pronouns and determiners. The demonstratives are particularly problematic for Finnish EFL learners, as the Finnish language does not take articles. However, the correlation between the incidence of articles and essay quality in the Finnish essays was approaching the significance level, suggesting gradual development towards more right types of articles in the right places. Without digital analyses, such subtle change would have been difficult to detect.

In comparison with the CEFR descriptions for cohesion (Council of Europe, 2018a), approximately 60% of the essays had reached at least the level B2, which is the stage that Finnish EFL students are expected to reach in English by the national school-leaving examination. This means that over half of the writers structured their text in paragraphs, avoided errors disrupting readability, and at least most of the time employed contextually appropriate lexis. The writers at level C1 knew how to employ lexical cohesion, substitution, and also employed more referential devices and a wider range of connectives compared to the students at level B2. The students remaining at level B1 overused the conjunctions and/but and often forgot to structure the text in paragraphs, which diminished the clarity of the text, and moreover, wrote in a rather informal style.

With regard to second research question, recent examinations suggest that various real-time observations are fundamental for understanding EFL students' writing processes, as they may assist teachers in introducing and instructing cohesion in the classroom (e.g., Bowen, & Thomas, 2020). Following this line of research, the present study employed video-recordings to investigate how cohesion emerged during essay writing.

The present observations concurred with previous results in that cohesion-building is closely connected to the writers' linguistic resources. The recorded data demonstrated that using online reference sources facilitated cohesion-building on the condition that the writer possessed adequate lexical knowledge and digital skills to search for lexis in online sources. For example, the present analyses indicated that writers scoring less than 64% in the VLT did not benefit from consulting online dictionaries and other reference sources in cohesion-building. In contrast, the participants scoring 80% or more in the VLT seemed to master most of the topic-related words without having to search for them, but when they searched, they were able to choose the appropriate option and crosscheck the findings. Higher-scoring writers also managed to add, change, and delete words to avoid repetition, whereas lower-scoring writers had difficulty in finding alternative lexis even if they had a chance to search for it. Efficient consultations were particularly beneficial to finding semantically related lexis. In this respect, the results supported previous observations conducted among adult students and based on different methodologies (Bowen, & Thomas, 2020; Abdel Latif, 2021; Lyashevskaya et al., 2021) and aligned with earlier findings on translation tasks (Mutta et al., 2014) as well as on indirect writing tasks (Niitemaa, & Pietilä, 2018).

6. Conclusion

As regards limitations of the present study, the small sample size restricts generalizing the findings on the use of cohesive features in EFL essays. However, the present findings seem to concur with recent results on cohesion-building. In the case of the video-recordings, the sample was even smaller due to scarcity of time and space in the computer room. Moreover, it may have been worthwhile to survey the students' individual experiences on the benefits and problems of accessing online affordances when writing. In relation to pedagogical implications, the findings demonstrate, firstly, that EFL learners need to be taught which properties different online dictionaries and translation tools provide and how they function. Although EFL learners encounter English in multiple digital environments in their free time, they mostly participate in activities which do not require consulting online dictionaries (Niitemaa, 2020). Thus, EFL students also need opportunities to practise using online reference sources in the classroom. Furthermore, emphasizing the individual nature of the writing processes, the present observations point towards searching for the optimal ratio between automated and person-to-person feedback.

In the context of observational research, future studies could benefit from recent developments in key-stroke logging such as *GenoGraphiX-LOG 2.0* (www.ggxlog.net). This tool stores the whole writing process and total navigation as a log file, calculates the writing bursts, additions, insertions, and deletions, and moreover, provides graphs of the writer's actions. For now, the tool has been used at the tertiary level to investigate, e.g., the types of collocations produced in different languages, and for feedback in consultations between the student and the teacher. Another area calling for future research is to examine more closely how lexical competence develops from recognition skills towards a large associative lexical network allowing productive use of lexis, e.g., for cohesion-building. In this regard, researchers could approach productive use of lexis from the perspective of lexical networks (Meara, 2009; Sigman, & Cecchi, 2002) comparing differences between cohesion in texts written by learners

with large but loosely organised lexicons and by learners with similar-sized but more densely organised lexicons.

References

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

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

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

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-with-new-descriptors-2018/1680787989

Council of Europe (2018b). *Descriptors of competences for democratic culture*. *Reference Framework of Competences for Democratic Culture*. Vol. *2*. https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/reference-framework-of-competences-for-democratic-culture

Crossley, 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., & 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. doi:10.17239/jowr-2016.07.03.02

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/ ezproxy.utu.fi/10.3758/s13428-018-1142-4

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., 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.

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

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

independent writing. *Assessing Writing, 37*, 39–56. https://doi.org/10.1016/j.asw.2018.03.002

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

Lintunen, P., Mutta, M., & Peltonen, P. (2020). *Fluency in L2 learning and use*. Multilingual Matters. https://doi.org/10.21832/9781788926317

Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing, 49*, 1–16. https://doi.org/10.1016/j.asw.2021.100529

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

Meara, P. (2009). *Connected Words: word Associations and second language vocabulary acquisition.* John Benjamins Pub.Co.

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

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

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*(3), 453–462.

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

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

Plonsky, L., & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*, 878–912. https://doi-org.ezproxy.utu.fi/10.1111/lang.12079

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

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

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

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

TAACO. https://www.linguisticanalysistools.org/taaco.html

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/papers/G-2020-68