

Development of critical thinking and language proficiency among undergraduate Swedish online EFL students

Evelina Johansson

ORCID number <https://orcid.org/0000-0002-9578-0526>

evelina.johansson@sprak.gu.se

Critical thinking (CT) is considered important in second language (L2) learning. Previous studies have investigated and confirmed a correlation between critical thinking and L2 learning in face-to-face (f2f) learning. The present study aimed to investigate the correlation between CT and L2 proficiency in an online learning environment. The participants in this study were undergraduate Swedish students taking part in an online EFL course at a Swedish university. Students' CT skills were measured by the California Critical Thinking Skills Test (CCTST). As previous studies confirm that vocabulary size correlates with general language proficiency, it was seen as a suitable measure of students' L2 proficiency. Hence, the Vocabulary Size Test (VST) was used to measure students' vocabulary size. The present study was designed as a pre-test post-test study aiming to investigate the correlation between the processes of CT development and L2 learning during one semester of four months. Contrary to previous studies, no correlation between CT and L2 proficiency could be detected. Moreover, students in this study made no statistically significant improvement of CT and L2 proficiency. The results must be interpreted in light of students' already high levels of CT and L2 proficiency and the relatively short testing period.

Keywords: critical thinking, English as a foreign language, online learning, vocabulary size.

1. Introduction

The ability to think critically is generally considered important in second language (L2) learning. Previous studies have established that students trained in critical thinking (CT) perform better on different L2 tasks (Alcón, 1993; Chen, 2016b; Lin, 2018) and that there is a general relationship

between students' levels of CT and L2 proficiency (Moore, 1995; Soodmand Afshar, Movassagh, & Radi Arbabi, 2017). A possible explanation for this relationship could be that language processing as well as CT require the use of cognitive resources. Manalo and Sheppard (2016) explain that "if a considerable amount of those resources has already been expended on utilising a language in which proficiency is low, there may not be adequate resources remaining for the satisfactory execution of critical thinking" (p. 42). Considering both the empirical evidence for the importance of CT in L2 learning and the theoretical discussions about cognitive cost, it becomes difficult to question the importance of CT in L2 learning.

Many undergraduate students today choose online learning over traditional face-to-face (f2f) learning. A previous study on thinking skills in online learning shows that Swedish online EFL instructors largely believe that higher-order thinking skills, such as CT, are important in L2 learning (Johansson, 2020b). Moreover, the majority of investigated assessment tasks used in online EFL courses at four Swedish universities focus on developing students' higher-order thinking skills (Johansson, 2020a). This aside, very little is known about students' development of CT in online L2 courses. The lack of studies contributed to an interest in investigating this topic. Hence, the present study aims to find out whether Swedish students taking part in a one-semester online EFL course develop CT skills and whether there is a correlation between the processes of CT development and L2 learning, which is measured in terms of vocabulary size in the present study. The strong correlation between vocabulary size and different forms of language proficiency makes it suitable as a measure of general L2 proficiency (Schmitt, 2010).

2. Background

2.1. Development of critical thinking in higher education

Students' development of critical thinking has been singled out as the principal goal of higher education (Arum & Roksa, 2010; Glen, 1995). Nevertheless, studies show that many students fail to develop these skills. One of the most cited of these studies is that of Arum and Roksa (2010). In their longitudinal study of 2,322 college students' learning, they found that from the beginning of the first college year to the end of the second year, students improved their CT skills by 0.18 standard deviation on average. This is a low measure compared with similar studies carried out in the 1990s that point to an improvement of 0.50 standard deviation (Pascarella & Terenzini, 2005). That

45 per cent of the students in Arum and Roksa's study did not show any statistically significant gains in CT is perhaps of even more interest.

Adding to the discussion on CT development in higher education, Huber and Kuncel (2016) carried out a meta-analysis of the average development of CT in higher education. The meta-analysis included 71 studies conducted between the years 1963 and 2011, and it confirms that more recent studies provide significantly smaller effect sizes than older studies. The authors mention that the predicted four-year gain for a study published in 1963 was 1.22 standard deviation, while it was only 0.33 for a study published in 2011. Moreover, the overall CT gain from college studies was estimated to be 0.59 in Huber and Kuncel's (2016) meta-analysis, which is higher than that found in Arum and Roksa's (2010) study.

2.2 Critical thinking and L2 learning

Given the importance of CT, many interventions aimed at supporting students' development of it have been implemented in EFL classes (e.g. Chen, 2016a; DeWaelshche, 2015; Ebadi & Rahimi, 2018; Lin, 2018; Nguyễn & Nguyễn, 2017; Yang & Gamble, 2013; Y.-T. C. Yang, Chuang, Li, & Tseng, 2013). The vast majority of these studies prove that CT interventions help students develop both CT and different types of L2 proficiency. For example, in a study conducted by Ebadi and Rahimi (2018), Iranian EFL students who took part in a WebQuest-based classroom focused on CT skills outperformed students in a regular face-to-face (f2f) classroom in both CT and academic writing. In a similar study, Yang and Gamble (2013) found that Taiwanese students participating in an EFL course with CT-enhanced instructions outperformed a control group in both CT and reading and writing. While these and similar studies indicate that CT and L2 learning are related, other factors may explain the simultaneous growth of these skills. The authors in both studies pointed out that EFL courses that aim at supporting students' CT skills are often characterised by traits previously known to support L2 learning, such as instructor support, collaboration, self-regulated learning (Yang & Gamble, 2013), and a high focus on certain L2 skills (Ebadi & Rahimi, 2018). Accordingly, drawing conclusions on the relationship between L2 learning and CT based on intervention studies may be problematic.

However, the relationship between CT and L2 proficiency has been investigated in more detail in other studies. In an attempt to investigate the relation between CT and English proficiency among

Malaysian students, Moore (1995) found that there was a significant moderate correlation between general language proficiency and CT both before and after a 16-week CT course. Nonetheless, overall L2 proficiency did not predict CT development. Similarly, Manalo and Sheppard (2016) found that Japanese EFL students' L2 proficiency correlated with the number of evaluative statements in written assignments; the number of evaluative statements was used as a measure of CT because evaluation is a CT skill. Moreover, several studies indicate that high levels of CT correlate with certain L2 skills, such as reading (Din, 2020; Kamali & Fahim, 2011) and writing (Liaw, 2007; Soodmand Afshar et al., 2017).

Since this study employed a vocabulary size test to measure students' L2 proficiency, previous studies that investigate the relationship between CT and vocabulary knowledge are of special interest. For example, Faramarzi et al. (2016) sought to study Iranian EFL students' vocabulary knowledge and CT. Students in this study were divided into groups of high, medium, and low CT levels based on their results on the Critical Thinking Ability Questionnaire. Using the IELTS as a measure of vocabulary knowledge, the study found that there was a significant difference between the three groups. Faramarzi et al. (2016) concluded that "the level of critical thinking of learners significantly affected their lexical knowledge" (p. 882).

The above-mentioned studies, as well as other studies within the field, indicate a reciprocal relationship between CT and L2 proficiency. However, the research area has not yet investigated the relationship between the processes of CT development and L2 learning. As seen above, the majority of the studies explore the correlation between CT and L2 proficiency at a certain time. The few studies that use a pre-test post-test design have not examined the correlation between CT gains and L2 development. Hence, little can be said about whether CT and L2 proficiency develop simultaneously.

It should be mentioned that some studies do not confirm a relationship between CT and L2 proficiency. For instance, Davidson and Dunham (1997) undertook a CT-intervention study among Japanese EFL students. While the treatment group scored significantly higher on the Ennis-Weir CT Essay Test, no statistically significant relationship between English proficiency and CT was detected. Similarly, in a study that investigated the relationship between higher-order thinking skills

Evelina Johansson

(HOTS) and L2 performance in an advanced Japanese course at an Australian university, Toyoda (2015) did not find a clear relationship between HOTS and L2 proficiency. However, the method used in these two studies is based on qualitative measures of CT. In contrast, most studies that confirm a relationship between CT and L2 proficiency use quantitative measures of CT. This indicates that the relationship between CT and L2 proficiency is far from straightforward.

The findings and conclusions of the previous studies on CT and L2 proficiency mentioned here indicate that more studies are needed to fully understand the complex relationship between the two aspects. Moreover, very little attention has been given to the development of CT and L2 learning in online EFL courses. It is therefore important to investigate L2 learning and CT development and the correlation between them in online learning, which the present study aims to do.

2.3 The relationship between vocabulary size and L2 proficiency

The rationale behind testing participants' vocabulary size in this study is that vocabulary knowledge can give a relatively quick insight into learners' general L2 proficiency. Schmitt (2010, p. 3) mentions that there are "high correlations between vocabulary and various measures of language proficiency". Alderson's (2005) development of the DIALANG tests is described as "one of the most systematic explorations of the relationship between vocabulary knowledge and language proficiency" (Schmitt, 2010, p. 4). Schmitt's (2010) study found that vocabulary size correlates with L2 skills such as reading, listening, writing, and grammar. Furthermore, vocabulary size accounted for as much as 37–62 per cent of the variance in language proficiency scores.

Moreover, previous studies confirm that vocabulary knowledge provides indications about general language proficiency. For example, Meara and Jones (1988, p. 80) noted that in general, speakers with a large vocabulary perform better on a wide range of linguistic indicators than speakers with a more limited vocabulary. A recent study conducted by Miralpeix and Muñoz (2018) found a high correlation between general EFL proficiency and vocabulary knowledge among Catalan and Spanish EFL students. Furthermore, according to Miralpeix (2020), "Different studies have shown that vocabulary size is closely related to proficiency; the bigger one's vocabulary is, the more proficient s/he is in a language" (p. 189). Considering the empirical evidence, it is safe to conclude that vocabulary knowledge is a reliable determiner of general language proficiency.

3. Aim and research questions

The literature review reveals that very little is known about CT development and L2 learning in online contexts. As several studies confirm the correlation between L2 proficiency and CT in f2f learning, investigating this relationship among students participating in online EFL courses becomes important. Aside from considering the correlation between levels of CT and vocabulary size, the present study seeks to explore the correlation between the development of these over one semester. Vocabulary size was chosen as a measure of language proficiency due to its strong relationship to other L2 skills and to overall L2 proficiency. The study aims to test the assumption that the development of L2 proficiency and CT are related. Given the small number of participants (see section 4.3), the study serves as a motivation for larger scale research in the future by addressing the following research questions:

1. Do undergraduate students taking part in a one-semester online EFL course at a Swedish university develop critical thinking?
2. Do undergraduate students taking part in a one-semester online EFL course at a Swedish university develop vocabulary size?
3. Is there a correlation between the development of CT and language proficiency as measured by vocabulary size among students participating in a one-semester online EFL course at Swedish universities?

4. Research Methodology

This section includes a description of the research procedure, tests, participants, data analysis, and ethical considerations.

4.1 Procedure

Students who took an intermediate online English course at four different universities were invited to participate in this study. The participants are described in more detail in section 4.3. Invitations to participate were sent out by the course coordinators. The California Critical Thinking Skills Test (CCTST), described in more detail in section 4.2.1, is used in this study to measure students' CT. A username and password are required to access the test. Hence, students who were willing to participate were asked to contact the author to give their consent and receive the login details.

Evelina Johansson

While this was far from the most ideal research design as it added additional burden and responsibility on the students, it was considered the only possible solution to reach these students. On the other hand, the study design ensured students' voluntary participation in the study. After contacting the author, the students received login details to the CCTST and a link to the vocabulary size test (VST) available on lextutor.com. The testing was conducted during two course occasions: in the spring and fall semesters of 2019. The students took both tests at the beginning and at the end of the semester, that is, in either September and December (for students taking the spring course) or January and May (for those taking the fall course).

4.2 Tests

4.2.1 California Critical Thinking Skills Test

The California Critical Thinking Skills test (CCTST) was used to measure participants' CT. This is a multiple-choice test developed by Facione and Facione (1994) and provided by Insight Assessment. The test consists of 34 questions aimed at targeting the following skills: induction, deduction, analysis, inference, and evaluation. In order for the test to be valid, ensuring an appropriate difficulty level is crucial; hence, Insight Assessment provides tests with different levels of difficulty. This study used the CCTST that is suited for undergraduate and graduate level students. The following is an example of how a question is formulated:

Three graduate school friends – Anna, Barbara, and Carol – graduated successfully. Being in the same program, the three often worked as a team on group assignments. Anna earned the special recognition of “pass with distinction” when she graduated. Carol and Barbara, although receiving their degrees, did not earn this special honour. A fourth student in the same graduate program, Deirdre, often said that the graduate program was poorly designed and not difficult at all. Deirdre did not graduate, instead she was advised by the faculty to withdraw from the program because her work was below acceptable standards. Given this information only, it follows that

A = Carol and Barbara deserved to receive “pass with distinction” like Anna.

B = Barbara's work in the program was superior to Carol's.

C = Barbara was jealous of the academic success her friend, Anna, enjoyed.

D = Deirdre's work in the program was below the quality of Carol's work.

E = Anna, being successful, will decide to enrol in another advanced graduate program.
(Insight Assessment, 2020)

Insight Assessment analyses the test results and automatically provides the test scores. While lack of information about how the tests are scored could be seen as a limitation, the benefit of ensuring that the test is scored correctly by a program developed by experts in testing CT is considered greater. In addition to providing an overall score and individual scores for the aforementioned skills, the CCTST presents participants' percentile in relation to four-year college students. That is, the participant's score is compared to other four-year college students' scores. According to Insight Assessment, the CCTST has a strong internal consistency reliability ($\alpha > .80$). To ensure that the test measures students' CT skills and not their English proficiency, the CCTST was translated to Swedish by the author of this study in collaboration with Insight Assessment.

4.2.2 The Vocabulary Size Test

As mentioned in section 2.3, vocabulary size is a reliable measure of general language proficiency. Hence, the present study used a vocabulary size test as an indicator of language proficiency. Vocabulary size tests are commonly used to measure receptive vocabulary knowledge. Words in these tests are divided based on frequency. According to Zipf's law, "the frequency of a word is roughly inversely proportional to its rank in the frequency table" (Milton, 2009, p. 45). Hence, there are few very frequent words, a lot of words that are infrequent, and a medium number of words that are in-between. Previous research suggests a low- and a high-frequency distinction. However, Nation (2013, p. 18) proposes four frequency groups instead: high-frequency words, mid-frequency words, low-frequency words, and specialised vocabulary. According to this classification, the high-frequency words consist of the first 3,000-word families; the mid-frequency words of 6,000 to 7,000 words ranging from the third 1,000 to the ninth 1,000; and the low-frequency words consist of words beyond the first 9,000.

To measure participants' vocabulary knowledge, the present study employed the Vocabulary Size Test (VST) (Nation & Beglar, 2007). The VST is a multiple-choice test that aims to measure test takers' receptive vocabulary knowledge. For this study, the 14,000-version, containing 140 multiple-choice items, was used. This version was selected because the study's participants were

Evelina Johansson

presumed to have relatively high levels of English proficiency. The VST consists of 140 test items, with 10 words per 1,000-word family level. Hence, the test includes word families from all three frequency levels mentioned above. Each word in the test represents 100 words; accordingly, the test covers the 14,000 most frequent word families. Through using the Lextutor webpage (<https://www.lexutor.ca>), the participants were able to take the test at their own convenience. The VST takes approximately 30 minutes to complete, and the mean time for the participants in this study was 27 minutes (standard deviation 16.5). As the students took the VST online without supervision, ensuring that the participants did not look up the answers was achieved through checking that participants did not deviate more than 16.5 minutes (i.e. the standard deviation) from the mean time. Hence, students who took longer than 43.1 minutes to complete the test were not included in the study.

4.3 Participants

As mentioned in section 4.1, the participants in this study were students who took an online intermediate level English course at Swedish universities. The universities were chosen for this study based on their previous participation in a study on higher-order thinking e-assessment conducted by the author (Johansson, 2020b). The course is the English 31–60 credits course. This is a second semester course that requires the students to have already studied one semester (30 credits) of English before. In addition to studying one of the mentioned courses, the only requirement on the participants was having Swedish as their first language (L1). The participant details are given in Table 1. These details are discussed in relation to the results in section 5.1.

Thirteen students decided to participate in the present study. The low participation is a limitation. In order to get more participants, the data was collected over two semesters. Nonetheless, it was difficult to get students to participate. Given the small data set, the results of this study should be interpreted with caution.

4.4 Data analysis

As the data was quantitative, it was analysed with statistical procedures using SPSS. Descriptive statistics for students' CCTST and VST scores were calculated. Since there were relatively few participants in this study, parametric tests could not be used. Thus, the data was analysed with non-

parametric tests. To investigate whether the students developed CT and vocabulary size, the Wilcoxon signed-rank test, which is the non-parametric alternative to the t-test, was used. One of the purposes of this study was to investigate the correlation between the development of CT and the development of vocabulary size over one semester. Therefore, the difference between pre- and post-test scores was first calculated then correlated using Spearman's Rho.

Table 1. Participants' details

	Age	Gender	Highest degree
1	49	Female	Bachelor
2	26	Male	High school
3	52	Female	Bachelor
4	35	Female	Master
5	26	Female	High school
6	44	Male	Bachelor
7	26	Female	High school
8	35	Female	High school
9	50	Female	Bachelor
10	31	Female	High school
11	51	Female	Master
12	30	Female	High school
13	44	Female	High school

4.5 Ethical considerations

Precautions have been taken to ensure the ethics of the study. As mentioned in section 4.1, students who wished to participate had to contact the researcher to get access to the tests. This ensured voluntary participation as the participants could easily decide whether they wanted to participate or not.

Moreover, the study prioritised ensuring confidentiality through protecting the participants' integrity and anonymity in accordance with the Swedish Research Council's guidelines for research ethics (Vetenskapsrådet, 2017, p. 41). In the collected data, the participants' personally-identifiable

information has been removed, and personal names have been replaced by numbers for identification. Thus, no information which can reveal the identity of the participants is revealed in this or other studies.

5. Results

The results of the investigation of undergraduate Swedish online students' L2 proficiency and CT skills will be presented in relation to each research question. The first section will present the results pertaining to the participants' CT levels. The second section will focus on the participants' vocabulary size as a measure of L2 proficiency. Finally, the last section will present the data analysis pertaining to the correlation between CT development and L2 learning.

5.1 Research question 1: Do undergraduate students taking part in a one-semester online EFL course at a Swedish university develop critical thinking?

Table 2 provides information about the students' scores on the CCTST pre- and post-tests. The CCTST overall scores in the range 0–7 are labelled as “not manifested” CT skills, 8–12 as “weak”, 13–18 as “moderate”, 19–23 as “strong”, and 24 or higher as “superior”. As can be seen in Table 2, all participants except two had moderate to superior CT skills. In the pre-test, six students had moderate CT skills, three students had strong CT skills, and four students had superior CT skills. In the post-tests, one student had weak CT skills, six students had moderate CT skills, one student had strong CT skills, and five students had superior CT skills.

While the purpose of the present study was not to investigate whether and how education level, age, and gender affect students' development of CT, these could possibly explain the results. As presented in Table 1, only two participants were men; hence, drawing conclusions about gender and development of CT is impossible. In the present study, age does not work as a predictor of high levels of CT skills or the development of such skills. Of the thirteen participants in this study, seven had a high school diploma, four had a bachelor's degree, and two had a master's degree. While the two students with the lowest CT skills had high school diplomas as their highest attained degree, levels of CT varied across the different educational levels, making it difficult to conclude that a higher degree predicts high levels of CT skills.

As seen in Table 3, the mean overall score of both the pre- and post-tests was 19.54, which is considered strong. Of the thirteen participants, six students improved their CT skills over the testing period, two made no improvement, and five showed weaker CT skills in the post-test than in the pre-test. Based on the distribution of the overall score percentiles for the test takers, compared to an aggregate sample of CCTST scores of four-year college students, the average percentile of the participants was 71. This implies that out of a total of 100 four-year college students, 70 will score lower than the average test taker in this study. There were significant differences both in students' percentile scores and overall scores, with the percentile ranging from 26 to 96 and the overall post-test score from 12 to 26.

Descriptive statistics for the CCTST is presented in Table 4. The Wilcoxon test was used to calculate the differences between pre- and post-test scores, and it showed that there was no statistically significant difference between these: $T = 35$, $p = .857$, $r = .035$. There were no statistically significant differences between the pre-test and post-test skills of induction ($T = 26.5$, $p = .918$, $r = .020$), deduction ($T = 8.5$, $p = .783$, $r = .054$), analysis ($T = 13.5$, $p = .248$, $r = -.226$), inference ($T = 45.5$, $p = .252$, $r = .225$), and evaluation ($T = 32$, $p = .571$, $r = -.111$).

Taken together, the results indicate that the students participating in the present study in general had relatively high CT skills, with the exception of one student. However, the students did not develop higher levels of CT skills during the one-semester course in English.

Table 2. Students' scores on the CCTST pre- and post-tests (max 30)

Student	Pre-test	Post-test
1	24	26
2	24	24
3	24	22
4	16	17
5	25	25
6	22	24
7	21	24
8	15	18

9	16	15
10	21	16
11	17	18
12	15	13
13	14	12

Table 3. CCTST pre- and post-tests: descriptive statistics

	N	Minimum	Maximum	Mean	Std. deviation
Pre-test	13	14	25	19.54	4.115
Post-test	13	12	26	19.54	4.841

5.2 Research question 2: Do undergraduate students taking part in a one-semester online EFL course at a Swedish university develop vocabulary size?

Table 4 presents the results of students' scores on the VST pre- and post-tests. Of the thirteen participants, seven students improved their vocabulary size, one student made no improvement, and five students showed lower levels of vocabulary size on the post-test compared with the pre-test. The mean vocabulary size improvement was 84 words, and the median was 100 words.

Descriptive statistics for the CCTST are presented in Table 5. The Wilcoxon test showed that there was no statistically significant difference between students' pre- and post-test scores on the VST: $T = 45.5$, $p = .511$, $r = .119$. The results indicate that seven of thirteen students developed vocabulary size during a one-semester online EFL course. However, there were surprisingly many students who made no development or who even scored lower on the post-test. Likewise, the statistical analysis shows no development of vocabulary size on a group level.

Table 4. Students' results on the VST pre- and post-tests

Student	Pre-test vocabulary size	Post-test vocabulary size
1	11500	10700
2	11800	12000
3	10900	10900
4	12300	12000

5	11000	10600
6	11800	12800
7	10700	10800
8	11800	12600
9	10800	11400
10	10600	9900
11	12000	11900
12	12600	12700
13	10200	10800

Table 5. Vocabulary size: descriptive statistics

	N	Minimum	Maximum	Mean	Std. deviation
Pre-test	13	10200	12600	11384	732
Post-test	13	9900	12800	11469	928

5.3 Research question 3: Is there a correlation between the development of CT and language proficiency as measured by vocabulary size among students participating in a one-semester online EFL course at Swedish universities?

A Spearman's rank-order correlation was used to determine the relationship between students' development of CT and vocabulary. There was no statistically significant correlation between these: $r_s = .351$, $p = .240$. Moreover, there was no statistically significant correlation between post-CT scores and post-vocabulary size scores: $r_s = -.175$, $p = .568$. Hence, the results indicate that the development of CT did not correlate with the development of vocabulary size over a one-semester online EFL course. Moreover, high scores on the post-CT test did not correlate with high scores on the post-VST test, indicating that CT and vocabulary size are not related.

Evelina Johansson

As CT did not prove to correlate with students' vocabulary size, other factors were tested to see if these better explained students' CT scores. There was no statistically significant difference on post-CT scores based on students' highest education level ($H(2) = .928, p = .629$), nor based on age ($H(8) = 9.39, p = .339$).

6. Discussion

The first research question sought to investigate whether undergraduate Swedish students develop CT during a one-semester online EFL course. As there was no statistically significant difference between students' pre- and post-test scores on the CCTST, whether the students developed CT skills cannot be confirmed. Students' development of CT skills is singled out as one of the main goals of higher education (Arum & Roksa, 2010). That participants did not develop these skills over a semester-long course is problematic and raises questions about how CT is best taught in online EFL courses. As discussed above, EFL learning environments that help students develop CT skills are characterised by instructor support, collaboration, self-regulated learning, and a high focus on certain L2 skills (Ebadi & Rahimi, 2018; Yang & Gamble, 2013). Incorporating these learning situations into online EFL courses is perhaps more challenging than it is in f2f courses. Hence, EFL instructors need to employ pedagogical tools known for helping students develop CT skills and adapt these to the online learning environment.

Moreover, the results of this study are in line with previous studies on students' development of CT in higher education. As mentioned above, Arum and Roksa (2010) discovered that as many as 45 per cent of the students in their study did not improve CT skills over two years of college studies. Even though the sample in this study is much smaller than that of Arum and Roksa's, similar results were found. A possible explanation for the non-significant results could be the relatively short time between the pre-test and the post-test. Students' CT skills were measured at the beginning and the end of the one-semester course. The results may have been different had students' CT skills been measured over a longer period. However, this was not possible as the English course that the students took was a separate course and not part of a programme.

The results of this study become interesting when taking Johansson's (2020a) previous study into consideration. In the mentioned study, assignment tasks in the studied courses were mapped with regards to HOTS and LOTS. Johansson (2020a) found that the courses have a high degree of

HOTS e-assessment tasks. Moreover, Swedish online EFL instructors consider HOTS and CT to be important in EFL courses (Johansson, 2020b). Nonetheless, students showed no statistically significant gains in CT. Previous research that proves that students can develop CT skills in EFL courses is often based on intervention studies. The mere intention of developing HOTS e-assessment and the use of a relatively high number of HOTS e-assessment tasks are perhaps not sufficient to help students develop these skills. Unlike the results of intervention studies that prove gains in CT skills, the results of this study indicate that there is need for more focused training in CT. Moreover, participating in online EFL courses does not in itself have a positive effect on students' development of CT.

While more research is required to understand why students did not develop CT, it is worth remembering that this topic has rarely been studied in the context of online courses before. Hence, it is possible that for online students to develop higher levels of CT, certain methods and learning environments are required. Moreover, Swedish students' relatively high levels of CT should be taken into consideration in the interpretation of the results. The mean CCTST overall score was 19.54 for both the pre-test and the post-test. Compared to an aggregate sample of CCTST four-year college students, the average percentile of the students was 71. This means that of 100 students, 70 students scored lower than the average student in this study. As shown in Table 2, most students already had relatively high levels of CT in the beginning of the course and made only minimal improvement during their studies. It is natural that students with high levels of CT skills make less improvement.

The second research question aimed at finding out whether Swedish students participating in online EFL courses develop vocabulary size as a measure of L2 proficiency over one semester. As mentioned in section 2.2, vocabulary size is in this study and in many other studies used as a measure of general language proficiency. Hence, the development of students' vocabulary size should be interpreted as an improvement of language proficiency in general. As seen in the results section, there was no statistically significant difference between students' pre- and post-test scores on the VST. Hence, the results indicate that students in these courses did not develop vocabulary size and general language proficiency. While it could be argued that students need more than one course of English studies in order to attain higher scores on the VST, the results raise questions about how

Evelina Johansson

to most effectively help students develop vocabulary size and language proficiency in general in online courses.

Notably, all students scored high on the VST. The mean for the pre-test was 11,384, and the mean for the post-test was 11,469. These scores indicate that the students in this study have mastered high-frequency words, mid-frequency words, and a substantial amount of low-frequency words (Nation, 2013). Low-frequency words are words that are not encountered as often as words in the two other categories. Hence, students with high levels of vocabulary size have more limited opportunities for development. Thus, taking the students' high VST scores into consideration, it is perhaps not surprising that no statistically significant difference between the pre- and post-tests was detected. Moreover, considering that vocabulary size has been found to correlate with L2 proficiency (Meara & Jones, 1988; Schmitt, 2010), it can be assumed that the students in this study had relatively high levels of English proficiency.

That five students had lower scores on the CCTST and the VST on the post-test compared with the pre-test is surprising. While it is difficult to explain these results, possible reasons could be lack of motivation or absence of supervision. However, this was not investigated in the present study. These factors could negatively affect the study's internal validity. As previously mentioned, a time limit was set on the VST, and students who deviated more than one standard deviation were excluded from the study. This ensured that the included students did not look up the meaning of words or randomly selected answers without reading the questions.

The third research question sought to investigate whether there is a correlation between students' scores on the CCTST and the VST. In order to do this, the correlation between post-test CCTST and VST scores and the correlation between the differences between pre- and post-test scores on the CCTST and the VST were calculated. Investigating the correlation between CT and different L2 skills is commonly done through comparing CT scores with L2 scores. While this provides a good understanding of whether these are correlated, it does not reveal anything about the correlation between the process of CT development and L2 learning. As the purpose of this study was to investigate whether these developed at a similar pace over time, it was important to calculate the correlation between the development of both CT and vocabulary size. However, the results of this study did not show a statistically significant correlation between the two.

Despite the many studies that prove that CT and L2 proficiency are related (Din, 2020; Manalo & Sheppard, 2016; Moore, 1995; Soodmand Afshar et al., 2017), this could not be established in the present study. Many of the studies in which a correlation between CT and language proficiency is established are intervention studies. As previously discussed, it is possible that CT interventions also support students' development of L2 skills. Hence, the detected correlation between CT and L2 proficiency in previous studies could be due to beneficial learning environments. It is possible that ambitious students in these intervention studies develop both CT and L2 skills, but this may have more to do with these students' efforts than with the correlation between the two skills. The present study is novel in that it sheds light on the correlation between CT and L2 proficiency in a formal learning environment without CT interventions. That no correlation could be found in the present study speaks for the complex relation between CT and L2 proficiency.

A possible explanation of the lack of correlation is students' already high levels of both CT and language proficiency. In the discussion and comparison between this study and other similar studies, it should be remembered that the correlation between CT and L2 proficiency in online learning has not, to this author's knowledge, been studied before. In order to determine whether students make less progress in CT in online courses, more research is needed. Future studies would ideally compare two equivalent or similar courses taught f2f and online to see whether and how the online learning environment affects students' development of CT.

Moreover, the results of this study could be interpreted to support the cognitive cost theory (Manalo & Sheppard, 2016). As the students had relatively high levels of vocabulary size and thus presumably high levels of English proficiency, they had enough cognitive resources available for the execution of CT skills. This could partially explain the difference between the results of this study and studies in which the participants had lower levels of English proficiency.

7. Conclusion

This study aimed to investigate CT development and L2 learning in online EFL courses over one semester (four months) due to a lack of studies in this context. Contrary to studies on f2f EFL courses, no statistically significant gains in CT or L2 proficiency could be established. The present

Evelina Johansson

study is novel in that it investigates the correlation between the actual development of CT and vocabulary size over one semester; however, no correlation between these aspects was discovered. Moreover, the Swedish online EFL students in this study already had high levels of both CT and L2 proficiency at the beginning of the semester. While high proficiency is a positive factor, it may have contributed to the lack of CT and L2 development. Thus, more research is needed in order to fully understand how an online learning environment affects students' CT development and L2 learning. Future studies would ideally investigate this topic with a less homogenous group to see whether varying levels of CT and L2 proficiency affect their development in online EFL courses.

To conclude, the results of the present study indicate that online EFL teachers need to take students' development of HOTS and CT into consideration. Studying online is limiting in that students miss out on important learning activities imbedded in traditional forms of learning, such as student-teacher and student-student interaction. It is possible that the lack of such activities has a negative effect on students' development of HOTS and CT. Hence, designing assessment tasks and online learning environments that aid students' development of these skills is crucial. Students' low development of CT skills in the present study implies that more work is needed within this area.

The present study is limited in that the sample is relatively small. Finding online students willing to participate in the study proved to be challenging. Hence, the results of this study should only be seen as a small contribution to the discussion about the development of CT and L2 learning in online contexts. Moreover, investigating this topic in an online context proved to be more difficult than anticipated. The absence of supervision as the students took the CCTST and the VST should be considered in the evaluation of the results.

Acknowledgements

I would like to thank the students who participated in this study. This research project would not have been possible without your collaboration. I would also like to show gratitude to Christina Lindqvist, Christine Fredriksson and the reviewers for providing comments that greatly improved the manuscript. A special thanks to Tom Cobb for all the help with the VST.

References

- Alcón, E. G. (1993). High cognitive questions in Nns group classroom discussion: do they facilitate comprehension and production of foreign language? *RELC Journal*, 24(1), 73-85.
- Alderson, C. J. (2005). *Diagnosing foreign language proficiency: the interface between learning and assessment*. New York: Continuum.
- Arum, R., & Roksa, J. (2010). *Academically Adrift: Limited learning in college campuses*. Chicago, IL: University of Chicago.
- Chen, M.-H. (2016a). Fostering L2 speaking and thinking through a HOT approach. *Curriculum & Instruction Quarterly*, 18(4), 193-225.
- Chen, M.-H. (2016b). Theoretical Framework for Integrating Higher-order Thinking into L2 Speaking. *Theory and Practice in Language Studies*, 6(2), 217-226. doi:10.17507/tpls.0602.01
- Davidson, B. W., & Dunham, R. A. (1997). Assessing EFL Student Progress in Critical Thinking With the Ennis-Weir Critical Thinking Essay Test. *JALT Journal*, 19(1), 43-57.
- DeWaelshche, S. A. (2015). Critical thinking, questioning and student engagement in Korean university English courses. *Linguistics and Education*, 32, 131-147. doi:10.1016/j.linged.2015.10.003
- Din, M. (2020). Evaluating university students' critical thinking ability as reflected in their critical reading skill: A study at bachelor level in Pakistan. *Thinking Skills and Creativity*, 35, 1-11. doi:10.1016/j.tsc.2020.100627
- Ebadi, S., & Rahimi, M. (2018). An exploration into the impact of WebQuest-based classroom on EFL learners' critical thinking and academic writing skills: a mixed-methods study. *Computer Assisted Language Learning*, 31(5-6).
- Facione, P. A., & Facione, N. C. (1994). *The California critical thinking skills test: Test manual*. Milbrae, CA: California Academic Press.
- Faramarzi, S., Elekaei, A., & Tabrizi, H. H. (2016). Critical Thinking, Autonomy, and Lexical Knowledge of Iranian EFL Learners. *Theory and Practice in Language Studies*, 6(4), 878-885. doi:10.17507/tpls.0604.28
- Glen, S. (1995). Developing critical thinking in higher education. *Nurse Education Today*, 15, 170-176.
- Huber, C. R., & Kuncel, N. R. (2016). Does College Teach Critical Thinking? A Meta-Analysis. *Review of Educational Research*, 86(2), 431-468. doi:10.3102/0034654315605917

- Johansson, E. (2020a). The Assessment of Higher-order Thinking Skills in Online EFL Courses: A Quantitative Content Analysis. *Nordic Journal of English Studies*, 19(1), 224-256.
- Johansson, E. (2020b). Higher-order thinking e-assessment in online EFL courses: Swedish instructors' perceptions and experiences. *Moderna Språk*, 114(2), 161-187.
- Kamali, Z., & Fahim, M. (2011). The Relationship between Critical Thinking Ability of Iranian EFL Learners and Their Resilience Level Facing Unfamiliar Vocabulary Items in Reading. *Journal of Language Teaching and Research*, 2(1), 104-111. doi:10.4304/jltr.2.1.104-111
- Lextutor. Vocabulary Size Test. Retrieved from <https://www.lextutor.ca/tests/vst/index.php?mode=test>
- Liaw, M.-L. (2007). Content-based reading and writing for critical thinking skills in an EFL context. *English Teaching & Learning*, 31(2), 45-87.
- Lin, Y. (2018). *Developing critical thinking in EFL classes: an infusion approach*. Singapore: Springer.
- Manalo, E., & Sheppard, C. (2016). How might language affect critical thinking performance? *Thinking Skills and Creativity*, 21, 41-49. doi:10.1016/j.tsc.2016.05.005
- Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. *Applied linguistics in society*, 80-87.
- Milton, James. (2009). *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.
- Miralpeix, I. (2020). L1 and L2 Vocabulary Size and Growth. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*. London: Routledge.
- Miralpeix, I., & Munoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, 56(1), 1-24. doi:10.1515/iral-2017-0016
- Moore, R. (1995). *The relationship between critical thinking, global English proficiency, writing, and academic development for 60 Malaysian second language learners*. (PhD). Indiana University, Indiana.
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). New York: Cambridge University Press.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nguyễn, T. M. T., & Nguyễn, T. T. L. (2017). Influence of explicit higher-order thinking skills instruction on students' learning of linguistics. *Thinking Skills and Creativity*, 26, 113-127. doi:10.1016/j.tsc.2017.10.004

- Pascarella, E. T., & Terenzini, P. T. (2005). How does college affect students: A third decade of research. San Francisco, CA: Jossey-Bass.
- Sample Thinking Skills Questions Retrieved from <https://www.insightassessment.com/article/sample-thinking-skills-questions>
- Schmitt, N. (2010). Researching Vocabulary A Vocabulary Research Manual: Palgrave Macmillan
- Soodmand Afshar, H., Movassagh, H., & Radi Arbabi, H. (2017). The interrelationship among critical thinking, writing an argumentative essay in an L2 and their subskills. *The Language Learning Journal*, 45(4), 419-433. doi:10.1080/09571736.2017.1320420
- Toyoda, E. (2015). Relationship between Higher-Order Thinking Skills and L2 Performance. *Electronic Journal of Foreign Language Teaching*, 12(2), 200-218.
- Vetenskapsrådet. (2017). Good Research Practice.
- Yang, & Gamble, J. (2013). Effective and practical critical thinking-enhanced EFL instruction. *ELT Journal*, 67(4), 398-412. doi:10.1093/elt/cct038
- Yang, Y.-T. C., Chuang, Y.-C., Li, L.-Y., & Tseng, S.-S. (2013). A blended learning environment for individualized English listening and speaking integrating critical thinking. *Computers & Education*, 63, 285-305. doi:10.1016/j.compedu.2012.12.012