

# The Assessment of Higher-order Thinking Skills in Online EFL Courses: A Quantitative Content Analysis

*Evelina Johansson, University of Gothenburg*

## Abstract

Studies within second language learning indicate that higher-order thinking skills (HOTS) are important for the process of learning a new language. At the same time, previous literature indicates that assessment tasks in online courses often focus on lower-order thinking skills. Little is still known about if and how thinking skills are assessed in online EFL courses. Hence, the purpose of this study is to create a more comprehensive understanding of if and how online EFL students at Swedish universities are given opportunities to both develop and being assessed on such skills. According to the sociocultural perspective, collaboration is beneficial to students' learning. Thus, the present study will also look into the correspondence between HOTS e-assessment tasks and collaborative e-assessment tasks. E-assessment tasks used in four online EFL courses given at Swedish universities have been classified according to the revised version of Bloom's taxonomy. This has been done through a quantitative content analysis of used e-assessment tasks. The study found that the majority of courses included more e-assessment tasks focusing on higher-order thinking than on lower-order thinking. However, a significant difference was detected between literature and linguistics modules in the sense that literature modules include more HOTS e-assessment tasks. Moreover, the results suggest that collaborative e-assessment tasks are slightly more common in e-assessment tasks that focus on HOTS than on LOTS. The present study provides insight into how thinking skills are assessed and developed in online language courses.

Keywords: higher-order thinking skills; critical thinking; collaborative learning; the revised Bloom's taxonomy; sociocultural theory; EFL; e-assessment

## 1. Introduction

Among the most important goals of education is the development of students' critical thinking (CT) skills (Arum and Roksa 2010). The growing awareness of the importance of developing students' critical thinking skills has in recent years contributed to an increased interest in thinking skills in education, language learning not being an exception. In particular, the teaching of higher-order thinking skills (HOTS) in EFL education has been extensively studied and studies indicate that higher-order thinking has an important role in second language learning (Ebadi and Rahimi 2018; Li 2016; Manalo and Sheppard 2016). The relevance

Johansson, Evelina. 2020. "The Assessment of Higher-order Thinking Skills in Online EFL Courses: A Quantitative Content Analysis." *Nordic Journal of English Studies* 19(1):224-256.

of CT in L2 learning has been discussed in terms of cognitive cost (Manalo and Sheppard 2016). Both language processing and CT require the use of cognitive resources. Learners with low levels of L2 proficiency have to use more cognitive resources to process the language and may, therefore, not have sufficient cognitive resources for the execution of CT. Thus, several studies report that cognition and the ability to learn a second language are related (Yang and Gamble 2013). Moreover, it has been found that EFL students trained in higher-order thinking benefit linguistically from CT training (DeWaelche 2015).

As technological and pedagogical improvements have contributed to making online education more attractive, many students prefer to study a second language in courses delivered completely online. Moreover, methods used in computer assisted language learning (CALL) have proven to be effective for both the purpose of developing higher-order thinking skills and in second language learning (Ebadi and Rahimi 2018; Yang 2017). Li (2016) concludes that even though it may be established that cognition and language learning are related, more research is needed in the area of technology enhanced learning and thinking skills.

Furthermore, research conducted in face-2-face (f2f) EFL courses has found that the presence of assessment items that cover HOTS in these courses is highly limited and that there is a strong focus on lower-order thinking skills (LOTS) (Köksal and Ulum 2018). Similarly, studies on other subjects reveal a predominance of LOTS in assessment tasks (FitzPatrick et al. 2015; Fensham and Bellocchi 2013). The differences between higher- and lower-order thinking skills will be discussed in more detail in section 2.1. Higher-order thinking assessment in online EFL courses stands out as an area not yet researched.

The aim of the present study is to examine the use of e-assessment tasks, that is assessment tasks delivered through digital technologies, in online EFL courses at Swedish universities. Relying on the revised version of Bloom's taxonomy (Bloom 1956), e-assessment tasks used in mentioned courses have been categorised according to the thinking level they cover from HOTS to LOTS. The English subject in Swedish universities is traditionally divided into linguistics and literature modules, it was therefore seen as interesting to investigate the differences in use of HOTS e-assessment in these modules. Moreover, as previous studies have found that collaborative learning promotes higher-order thinking skills, the study will look into the correspondence between

HOTS e-assessment and collaborative e-assessment. The focus on collaborative learning is also grounded in the sociocultural theory, which relevance to the study is described in more detail in section 2.4.

The present study is part of a larger research project which investigates the presence and role of HOTS e-assessment in online L2 learning (Author in progress). While investigations of learners' responses to e-assessment tasks are interesting and worth examining, the focus of this part-study is on the e-assessment tasks developed by the teachers.

## *2. Previous Research and Theoretical Background*

Below is a brief review of literature related to the role of higher-order thinking in L2 education. Previous research on developing higher-order thinking through e-assessment is also outlined. The study is framed within sociocultural theory and considers collaborative learning important in both L2 learning and in the development of higher-order thinking. Thus, this section will include a brief discussion on collaborative learning and HOTS.

### *2.1 Higher-order thinking skills (HOTS) and Critical Thinking (CT)*

The concept of HOTS originates from the idea that there exists a hierarchy of thinking skills. Certain thinking skills are assumed to be more difficult or demanding to attain than others. There is no consensus of what constitute HOTS; however, Lee and Choi (2017: 144) mention that most researchers today agree that HOTS involve 'complicated cognitive activities such as formulating hypotheses; elaborating, interpreting, and analysing information; applying multiple criteria; constructing arguments; making comparisons and inferences; integrating and synthesizing information; and yielding multiple solutions'. As such, HOTS include thinking processes that stretch beyond mere storage of information and understanding. Other definitions of HOTS include skills such as critical thinking, problem solving, and evaluative skills (Schraw and Robinson 2011; Gorin and Dubravka 2011).

The revised version of Bloom's taxonomy, which is described in greater detail in section 3.3, is frequently used in defining HOTS (Leighton 2011; Schraw and Robinson 2011; Afflerbach, Cho, and Kim 2011). The original Bloom's taxonomy was developed as a collaborative

effort by a committee of educators (Bloom 1956). The aim was to develop a framework for measuring and categorizing educational objectives. Almost 40 years after the advent of Bloom's taxonomy, Anderson and Krathwohl perceived a need to 'incorporate new knowledge and thought into the framework' which resulted in a revision of the taxonomy (Anderson et al. 2014: XXII). According to the revised taxonomy, thinking skills are hierarchical and divided into six categories. The categories consist of: *Remember*, *Understand*, *Apply*, *Analyse*, *Evaluate*, and *Create* (Anderson et al. 2014). The three highest skills, *Analyse*, *Evaluate* and *Create*, are traditionally considered to be encompassed within the term HOTS and the three first skills, *Remember*, *Understand* and *Apply*, are considered to be LOTS.

Related to the idea about HOTS is critical thinking (CT). Since both HOTS and CT are concerned with higher cognitive functions, they are at times used interchangeably. CT is closely related to our understanding of HOTS and, by some, included in the definition of HOTS (Schraw and Robinson 2011).

### *2.2 Teaching thinking skills in EFL education*

It is well known that the presence of HOTS has a positive effect on students' general academic achievement. Ghanizadeh (2016) found that the interaction between CT and reflective thinking contributed to students' academic achievement. While these findings are interesting, it should be noted that students' CT skills were measured using the Watson-Glaser Critical Thinking Appraisal which has been criticised for its low validity (Possin 2014). The Watson-Glaser Critical Thinking Appraisal is a multiple-choice test. A more qualitative approach could have provided insights not only on the interrelation between CT, reflective thinking and academic achievement, but also on the actual development of these. The author of the mentioned study presumes that higher-order thinking skills 'provide learners with sophisticated and complex competency to engage in effective learning strategies, to be more committed to their studies, and to be more reflective and organized in their planning and organization' (Ghanizadeh 2016: 110). Hence, the presence of HOTS is a necessity in all higher education.

A review of the literature confirms that EFL scholars and teachers find thinking skills important in the L2 classroom. Yang and Gamble's

(2013) study further strengthens our understanding of the importance of and relationship between cognition and language learning. In this study, Taiwanese EFL students were offered an English course with CT integrated instruction. To investigate the effect of the intervention, an experimental research method was used which included a control group. The study found that the experimental group significantly outperformed the control group in both CT skills and overall English proficiency as well as listening and reading proficiency. The collaborative features imbedded within the CT integrated instructions, combined with greater instructor support and self-regulated learning, are assumed to have facilitated students' L2 learning. Even though the study does not claim to prove that higher cognitive functions can be correlated with language learning, it indicates that L2 learners' linguistic proficiency benefits from CT integrated instructions.

Similar to Yang and Gamble's study, other studies provide evidence for the relationship between interventions aimed at developing students' HOTS and certain L2 skills. Chen (2010) has researched the effect of implementing higher-order questioning in Taiwanese ESL education and found that this approach had a positive effect on students' L2 speaking proficiency. A similar study among Korean university students majoring in English found that students believed that an intervention aimed at helping them develop the ability to ask HOTS questions had a positive effect on their conversational English (DeWaelche 2015). Likewise, Alcón (1993), who was among the first to conduct research on HOTS and L2 learning, discovered that L2 students trained in asking high cognitive questions wrote more semantically and syntactically complex texts than students who had not undergone the same training. It may be concluded from the studies mentioned so far that high levels of CT and CT integrated instruction provide students with skills and learning opportunities which facilitate learning in general, including L2 learning.

There is, however, research that indicates that higher cognitive functions do not relate with L2 learning. Davidson and Dunham (1997) studied English L2 proficiency and critical thinking skills among Japanese English as a second language (ESL) students who had undertaken a year-long course on CT. Students' critical thinking skills were measured using the Ennis-Weir Critical Thinking Essay Test, which is one of the few CT tests that does not have a multiple-choice format. Participating students' results were compared with those of a control

group. The treatment group scored significantly higher on the Ennis-Weir test than the control group; however, no statistically significant relationship between critical thinking skills and English proficiency could be found. While this may prove that students' results on the Ennis-Weir test were not affected by their English proficiency, it also indicates that higher levels of L2 English had little influence on students' ability to think critically. Similarly, Toyoda (2015) studied the use of HOTS and L2 proficiency among learners of Japanese at an Australian university. The students in this study were part of a video-sharing project aimed at supporting HOTS. An analysis of students' videos, reflective diaries and teacher notes did not provide evidence for a relationship between HOTS and L2 performance (Toyoda 2015). It is interesting to note that both studies which failed to prove a clear relationship between thinking skills and L2 learning are among the few studies in which CT is measured through qualitative methods. This could perhaps indicate that measuring CT is not as straightforward as promised in various multiple-choice tests.

Based on this review of literature on HOTS and L2 learning, we may conclude that there exist indications that HOTS have an important role in the process of learning a second language. However, to clearly understand how HOTS affect the language learning process, more research within the field is needed. Several scholars confirm that there is a need for more research within the field (Soodmand Afshar and Movassagh 2014; Chen 2010; Liaw 2007). All the above-mentioned studies focus on developing HOTS in regular f2f learning, thus not taking into account the e-learning environment in language learning. To the best of my knowledge, no previous research exists that investigates if and how HOTS are assessed in online EFL courses nor on the relationship between HOTS and L2 learning in an online context.

### *2.3 e-Assessment and HOTS*

Assessment has often been described as the main driver of students' learning, meaning that students will learn what they think they will be tested on (Biggs and Tang 2007). This also includes the learning of thinking skills and as such assessment tasks have an important role in the teaching of students' HOTS. Bates and Sangrá (2011: 21) explain that 'if we are setting examinations (or other forms of assessment) that do not explicitly assess problem solving, critical thinking, digital literacy, and

communications skills within the subject domain, then students will not focus on developing these’.

As e-learning has become a well-established alternative to traditional forms of learning, researchers have shown interest in the possibility of developing and assessing HOTS in an e-learning environment. A review of recent literature shows that the e-learning environment may have positive effects on students’ ability to develop these skills. Several studies indicate that the collaborative aspects often imbedded in e-assessment tasks facilitate the development of HOTS. Student discussion forums have proven to be especially beneficial for this. In line with this, Yang et al. (2005) discovered that the use of Socratic questioning in asynchronous online discussion forums may help students develop CT skills. Students’ CT skills were measured with the California Critical Thinking Skills Test (Facione and Facione 1994). The study found that students benefitted from online discussions and developed higher levels of CT skills. In section 2.2, we confirmed the efficiency of interventions aimed at fostering students’ CT skills and the study by Yang et al. proves that such interventions can also be effective in an e-learning setting.

Much of the research on e-assessment focuses on multiple choice questions (MCQs) and the effectiveness of MCQs for the development and testing of HOTS. Falchikov and Thompson (2008) argue that MCQs are not beneficial for students’ learning progress. This, they reason, is because MCQs focus on low level cognitive processes instead of HOTS. The same stance is taken by Boitshwarelo, Reedy and Billany (2017) who carried out a literature review of studies on computer-assisted assessment in higher education with the aim of investigating if these align with 21<sup>st</sup> century learning. MCQs were found to be the most commonly used alternative among online tests. These tests are primarily used to assess the first three levels, Remember, Understand and Apply, of the revised version of Bloom’s taxonomy. (Boitshwarelo, Reedy, and Billany 2017). There are, however, those who argue that well-designed MCQs can help students develop HOTS (Brookhart 2014). To conclude, the perceived limitations of computer-assisted assessment do not seem to be an inherent feature, but a result of poorly designed tests.

Even though it seems possible to support the development of and assess HOTS through suitable e-assessment tasks, this has proven a challenge (McNeill, Gosper, and Xu 2012). McNeill et al. (2012) studied the use of technologies to support higher-order learning at an Australian

university. Through interviews with convenors of online units, they found that technology supported assessment tasks focused mainly on lower-order outcomes (McNeill, Gosper, and Xu 2012). The results were partially explained by higher education being slow to change. Despite efforts made to ensure that assessment tasks support the learning of HOTS, knowledge and comprehension are more commonly assessed than higher levels of thinking (Bryan and Clegg 2006).

Reviewing the current literature, we find that developing e-assessment tasks that focuses on HOTS are not impossible, although, challenging. Caplan and Graham (2008: 252) mention that ‘the unique possibility inherent in web-based instruction originate not from the Web itself, but from the instructionally innovative ways in which it may be used’. To sum up the discussion on e-assessment and HOTS, it may be concluded that if designed correctly, e-assessment tasks can facilitate the development of higher-order thinking skills.

While assessment is considered one of the most important parts of education, previous studies that investigate HOTS and LOTS assessment tasks in different subjects are limited. To the author’s knowledge, few systematic studies have been carried out with the aim to investigate if and how HOTS are assessed in language courses. Köksal and Ulum (2018), as mentioned above, found that exam questions in English courses at Turkish universities solely focus on LOTS. Moreover, a few studies in other subjects indicate a predominance of LOTS assessment tasks in most courses. For example, Fensham and Bellochi (2013) found that the majority of the examination tasks in Australian high school chemistry courses are devoted to LOTS. Similarly, a study on cognitive processes in assessment tasks used in pharmacy education in Canada revealed that only 33.7 percent of the tasks require HOTS (FitzPatrick et al. 2015). Even though, as discussed above, there are indications from teachers working in online learning that HOTS is not sufficiently assessed in online courses, there is a lack of studies that investigate the frequency of HOTS and LOTS e-assessment in online courses.

#### *2.4 A sociocultural perspective on thinking skills and language learning*

According to the sociocultural theory, humans take control over higher mental functions by using cultural and symbolic artifacts as mediational means (Lantolf, Thorne, and Poehner 2015). The technology used in e-

learning, therefore, has an important role in students' learning. Online collaborative activities need to be designed so that the activities afford and mediate interaction and participation (Khoo and Cowie 2011).

The present study is framed within the sociocultural perspective as collaboration is considered important for development of students' HOTS. The zone of proximal development (ZPD) has become the most popular and famous component of sociocultural theory. Vygotsky's (1978: 86) well-known definition of the ZPD is often misunderstood as referring to the range of tasks performed in collaboration. A more accurate definition is that it refers to levels of students' development. Furthermore, the ZPD is made up of 'the area of immature, but maturing processes' (Vygotsky, 1998, p. 202). Chaiklin (2003: 52) explains that the ZPD refers to 'those intellectual actions and mental functions that a child is able to use in interaction, when independent performance is inadequate'. Thus, in collaboration students will be given opportunity to develop processes that are not yet fully mature, but in the process of maturing. It is not, as sometimes misunderstood, a process of developing previously non-existing functions and processes.

Collaborative activities have previously proven to be helpful in developing students' HOTS. Gokhale (1995) investigated the effectiveness of collaborative learning in promoting both basic thinking skills and CT skills among American undergraduate students in industrial technology. The students were divided into one control group and one experimental group. Both groups were given the same assignment; however, the experimental group was allowed to complete their assignment collaboratively. Students' basic thinking skills and critical thinking skills were tested through a pre-test post-test design study which was developed based on Bloom's taxonomy. Students who had collaborated on the given assignment scored significantly higher on CT items on the post-test. Gokhale (1995: 28) views the results in light of a sociocultural perspective and explains that 'the peer support system makes it possible for the learner to internalize both external knowledge and critical thinking skills and to convert them into tools for intellectual functioning'. It is, thus, assumed that collaborative learning activities are beneficial for the development of CT.

Collaboration, however, should 'not be understood as a joint, coordinated effort to move forward, in which the more expert partner is always providing support' (Chaiklin 2003: 54). On the contrary, as

explained by Chaiklin, it is any situation in which the learner is offered interaction with another person, peer or teacher, to solve a problem. Thus, it is not the competence of the more capable peer that is of most importance for students' development, but the actual assistance itself (Chaiklin 2003: 43). As the ZPD is well-known among educationalists, the aim of the present study was to investigate if collaborative e-assessment tasks in online EFL courses had a higher concentration of HOTS than non-collaborative tasks. The purpose of the investigation was not to study how students respond to the collaborative tasks used in the investigated courses; rather, it was to look into opportunities created for collaboration.

### *2.5 Research questions*

To sum up, HOTS seem to have an important role in L2 learning. However, previous research indicates that higher-order thinking is often not assessed in online courses (McNeill, Gosper, and Xu 2012). Even though many students choose to study a foreign language through online courses, little is still known about the e-assessment tasks used in these courses. The present study is based on the hypothesis that e-assessment tasks that are carried out collaboratively develop students' HOTS to a higher degree than non-collaborative e-assessment tasks, and seeks to investigate if collaborative e-assessment tasks include more HOTS instructions than non-collaborative tasks. Since the English subject at Swedish universities includes both linguistics and literature modules, as explained in section 3.2, investigating the differences in HOTS e-assessment in these modules stood out as something important. More specifically, the following research questions are addressed:

1. To what extent do e-assessment tasks used in online EFL courses at Swedish universities focus on higher-order thinking?
2. How do linguistic and literature modules in Swedish universities differ in their e-assessment of HOTS?
3. To what extent does the use of HOTS e-assessment correspond to the presence of collaborative e-assessment tasks?

### *3. Method, Data and Procedure*

#### *3.1 Content analysis*

For the purpose of this study, e-assessment tasks used in online English intermediate courses at four Swedish universities were investigated. The aim was to classify these e-assessment tasks according to the thinking levels in the revised Bloom's taxonomy (Anderson and Krathwohl 2001). In order to do that, a quantitative content analysis of the e-assessment tasks was conducted. The mentioned method was considered suitable for this study as the method aims at systematically classifying parts of a text to draw conclusions about the content (Rose, Spinks, and Canhoto 2015). Content analysis has been used in previous studies investigating the frequency of higher- and lower-order thinking skills (e.g. Ulum 2016).

#### *3.2 Unit of analysis and participants*

The units of analysis in the present study were the e-assessment tasks. To collect these, Swedish universities offering the English intermediate online course were invited to participate in this study. The single requirement placed on the invited universities was that they offered an online intermediate course in English equivalent of English 31-60 credits. This is a second semester course and requires that the students have completed a one semester (30 credits) English course before. There were eight universities giving this course at the time of investigation and half of them participated in this study. As the author could not be granted access to the course pages, the instructors working with these courses agreed to send the e-assessment tasks to the author. As mentioned in section 2.5, the English subject at Swedish universities is traditionally divided into both literature and linguistics modules. All four courses consisted of a 15 credits linguistic module and a 15 credits literature module. See Table 1 for a detailed description of each course's content and assessment tasks. Due to changes in staff during the time of data collection, the material from one 4 credits module in linguistics from University 3 could not be collected. All e-assessment tasks in these courses, which totalled 500 tasks, were collected and categorized.

In the term *e-assessment task* everything that the students had to complete in order to pass the course was included. Compulsory seminars and forum discussions were also included in this definition. The investigated e-assessment tasks in this study included seminar questions,

mid-course and final exams, essays, oral presentations, take-home exams, forum discussions, and compulsory study questions.

The results are at times presented in credits of HOTS and LOTS e-assessment per course. This was done through multiplying the percentage of HOTS/LOTS e-assessment tasks with the course credits to account for the number of credits assessed by each thinking level. Some courses had clear instructions for the credit value of each e-assessment task, this was in particular true for written essays and exams. In these cases, HOTS and LOTS were calculated per credit. However, when there was no specific mention of credit, HOTS and LOTS were calculated per course module. The following example illustrates how this was done: A 7.5 credit module contained one written exam, one essay and seven seminar questions. The written exam and the essay were given two credits each by the teacher responsible for the course. Three and a half credits remained for the seven seminar questions; hence, each of these had a value of 0.5 credits. The exam included ten questions of which six of them were HOTS questions and four of them were LOTS questions. Thus, the exam gave 1.2 credits of HOTS and 0.8 credits of LOTS. The essay consisted of instructions that were categorized as HOTS; hence, the essay gave 2 HOTS credits. As for the seminar questions, five of them were HOTS questions and two were LOTS questions. Therefore, 2.5 credits were in the category of HOTS and 1 credit in the category of LOTS. To sum up, this 7.5 credit module contained 5.7 HOTS credits and 1.8 LOTS credits. The reasoning behind this method was to adjust both for the unequal proportions of e-assessment tasks in investigated courses and between differences in credit value between e-assessment tasks. This was done in order to ensure a more valid representation of the results.

Table 1. Participating courses

	Credits	Modules	Content	Assessment tasks
<b>University 1</b>				
Literature	15	1	British and American literature	Written exam, discussion forum questions, written essay
Linguistics	15	1	Linguistics	Seminar questions, written exam, written essay
<b>University 2</b>				
Literature	15	1	British and American literature	Written essays, written exam, oral presentation, seminar discussions
Linguistics	15	2	Linguistics, language history	Written assignments, oral presentation, seminar questions, written exam
<b>University 3</b>				
Literature	15	2	British and American literature	Seminar questions, written assignments, written exam
Linguistics	15	3	Linguistics, translation	Written exams, discussion forum questions, oral presentation, seminar questions
<b>University 4</b>				
Literature	15	2	British and American literature	Written exams, oral presentation, compulsory study questions
Linguistics	15	2	Linguistics	Discussion forum questions, written exam, written assignments, peer-reviewing, oral presentation

### 3.3 *The revised Bloom's taxonomy*

Bloom's taxonomy (Bloom 1956) has since its publication in 1956 been widely used in measuring and classifying educational objectives. The taxonomy consists of six categories in the cognitive domain; *Knowledge*,

*Comprehension, Application, Analysis, Synthesis, and Evaluation* (Krathwohl 2002). Almost 40 years after the advent of Bloom's taxonomy, Anderson and Krathwohl perceived a need to revise it according to new knowledge (Anderson et al. 2014). The revised version of Bloom's taxonomy distinguishes between the Cognitive process dimension and the Knowledge dimension. The Knowledge dimension consists of *Factual, Conceptual, Procedural, and Metacognitive Knowledge*. The Cognitive Process dimension still includes six categories; however, a few important changes have been made within the framework. Anderson and Krathwohl's revised cognitive process dimensions include the thinking categories: *Remember, Understand, Apply, Analyze, Evaluate, and Create* (Anderson et al. 2014). These categories are assumed to be arranged in a continuum of cognitive complexity where *Remember* is the least cognitively complex category and *Create* the most complex.

Among the purposes of developing a taxonomy of this kind was the desire to help educators consider the possibilities within education (Anderson et al. 2014). This, in turn, will support educators in developing instructions and test items that tap higher-order thinking. Krathwohl (2002: 213) mentions that the taxonomy has frequently been used 'to classify curricular objectives and test items in order to show the breadth, or lack of breadth, of the objectives and items across the spectrum of categories'. Thus, the categorization of educational objectives does not only aim at deciding the content of what should be taught, but also at supporting and guiding the assessment of these skills. As previously mentioned, most definitions of HOTS still rely on Bloom's taxonomy (Schraw and Robinson 2011). The three highest levels, i.e. Analyze, Evaluate, and Create, of the taxonomy are commonly considered to be encompassed within the term higher-order thinking. Several studies have been carried out with the aim of identifying higher-order thinking tasks in the EFL classroom that rely on the top three levels of Bloom's taxonomy or the revised taxonomy to define HOTS (Riazi 2010; Ali 2012; Mohammadi et al. 2015).

### 3.4 Coding scheme

The coding scheme used in this study was based on the revised version of Bloom's taxonomy. The classification scheme (Table 2) includes the

entire coding scheme with definitions of the thinking levels and action verbs that have been used as a guide in deciding on the appropriate thinking level (Anderson and Krathwohl 2001). The revised taxonomy has been used in similar studies previously and is commonly used to determine the thinking level of assessment tasks (Köksal and Ulum 2018; FitzPatrick and Schulz 2015). Each e-assessment task was categorized according to the thinking level it covers. As thinking is hierarchical, several tasks were categorized into more than one thinking level. However, the highest thinking level was always the one that was counted.

As can be seen in table 2, the same action verbs occasionally occur in several categories. The author of this study is well-aware of this and as there are clear definitions to rely on in case of uncertainty, this has not affected the categorization of e-assessment tasks. Furthermore, the categorization of e-assessment tasks was primarily based on the definitions of the cognitive categories. Table 3 provides examples of categorization of e-assessment tasks in this study. The universities which agreed to participate in this study did so with the condition that all material they shared was to be confidential and not published. Hence, the examples of e-assessment tasks in table 3 have been modified with that in mind. As can be seen in table 3, the action verbs corresponding to the thinking level are often mentioned in the assessment tasks. These action verbs are italicized. However, sometimes actions verbs were not mentioned in the question. As the categorization was primarily based on the given definitions of each thinking level, this did not constitute a problem. For example, in the e-assessment task given for the thinking level *Evaluate* no specific mention of evaluate or any other related action verb is made. Still, based on the definition of Evaluate, it was understood that this is a case of 'detecting the appropriateness of a procedure'. That is, detecting the appropriateness of studying such texts today.

Table 2. Coding scheme with definitions and action verbs. Based on (Anderson et al. 2014; Center for University Teaching)

Level	Definition
<b>1. Remember</b>	<b>Retrieve relevant knowledge from long-term memory</b>
1.1 Recognize	Locating knowledge in long-term memory that is consistent with presented material
1.2 Recalling	Retrieving relevant knowledge from long-term memory when given a prompt to do so
Action verbs	Choose, define, find, how, label, list, match, name, omit, recall, relate, select, show, spell, tell, what, when, where, which, who, why.
<b>2. Understand</b>	<b>Construct meaning from instructional messages, including oral, written, and graphic communication</b>
2.1 Interpreting	Changing from one form of presentation to another
2.2 Exemplifying	Finding a specific example or illustration of a concept or principle.
2.3 Classifying	Determining that something belongs to a category.
2.4 Summarizing	Abstracting a general theme or major point.
2.5 Inferring	Drawing a logical conclusion from presented information.
2.6 Comparing	Detecting correspondences between two ideas, objects, and the like.
2.7 Explaining	Constructing a cause-and-effect model of a system.
Action verbs	Classify, compare, contrast, demonstrate, explain, extend, illustrate, infer, interpret, outline, relate, rephrase, show, summarize, translate.
<b>3. Apply</b>	<b>Carry out or use a procedure in a given situation</b>
3.1 Executing	Applying a procedure to a familiar task.
3.2 Implementing	Applying a procedure to an unfamiliar task.
Action verbs	Apply, build, choose, construct, develop, experiment with, identify, interview, make use of, model, organize, plan, select, solve, utilize.

<b>4. Analyze</b>	<b>Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose.</b>
4.1 Differentiating	Distinguishing relevant from irrelevant parts or important from unimportant parts of presented material.
4.2 Organizing	Determining how elements fit of function within a structure.
4.3 Attributing	Determine a point of view, bias, values, or intent underlying presented material.
Action verbs	Analyze, assume, categorize, classify, compare, conclusion, contrast, discover, dissect, distinguish, divide, examine, function, inference, inspect, list, motive, relationships, simplify, survey, take part in, test for, theme.
<b>5. Evaluate</b>	<b>Make judgements based on criteria and standards</b>
5.1 Checking	Detecting inconsistencies or fallacies within a process or product; determining whether a process or product has internal consistency; detecting the effectiveness of a procedure as it is being implemented.
5.2 Critiquing	Detecting inconsistencies between a product and external criteria, determining whether a product has external consistency, detecting the appropriateness of a procedure for a given problem.
Action verbs	Agree, appraise, assess, award, choose, compare, conclude, criteria, criticize, decide, deduct, defend, determine, disprove, estimate, evaluate, explain, importance, influence, interpret, judge, justify, mark, measure, opinion, perceive, prioritize, prove, rate, recommend, rule on, select, support, value.
<b>6. Create</b>	<b>Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure.</b>
6.1 Generating	Coming up with alternative hypotheses based on criteria.
6.2 Planning	Devising a procedure for accomplishing some task.
6.3 Producing	Inventing a product.
Action verbs	Adapt, build, change, choose, combine, compile, compose, construct, create, delete, design, develop, discuss, elaborate, estimate, formulate, happen, imagine, improve, invent, make up, maximize, minimize, modify, original, originate, plan, predict, purpose, solution, solve, suppose, test, theory.

Table 3. Examples of e-assessment tasks per thinking level

Thinking level	Examples of e-assessment tasks
<b>Remember</b>	In the study of discourse analysis, <i>what</i> are hedges?
<b>Understand</b>	What are direct speech acts? What are indirect speech acts? <i>Illustrate</i> both notions with examples from your own experience.
<b>Apply</b>	<i>Apply</i> a test of any kind to show whether the underlined parts of the following sentences are independent or dependent clauses.
<b>Analyze</b>	Give examples of two texts that we have studied in this course and which illustrate trends you associate with literary realism. <i>Compare</i> and <i>contrast</i> these texts
<b>Evaluate</b>	Should this poem, and other historical texts, be studied today, at universities and in schools? What arguments can you think of for doing so and against?
<b>Create</b>	<i>Formulate</i> a rule for forming verbs in the passive voice in words like these in English.

#### 4. Results

A total number of 500 e-assessment tasks were coded in this study of which 177 were from literature modules and 323 from linguistics modules. The results will be presented in relation to each research question.

##### 4.1 Research question 1: To what extent do e-assessment tasks used in online EFL courses at Swedish universities focus on higher-order thinking?

The 500 e-assessment tasks analysed in this study were categorized according to the highest level that these assessments are relevant (Table 4). The thinking level *Understand* took up the majority of the questions, with 50 percent of all questions, followed by *Analyze*, *Evaluate*, *Remember*, *Apply*, and *Create*. The usage of both *Apply* and *Create* questions was very limited.

Table 4. Total number of tasks per thinking level

Level	Assessment tasks
<b>Remember</b>	34
<b>Understand</b>	250
<b>Apply</b>	14
<b>Analyze</b>	152
<b>Evaluate</b>	43
<b>Create</b>	7
<b>Total</b>	500

The results per university reveal that three out of four universities had almost the same frequency order of questions per thinking level (Table 5). E-assessment tasks focusing on *Understand* had the highest frequency for University 1, 2, and 3. *Analyze* was the second most frequent thinking level in these three universities, followed by *Evaluate* in University 1, *Remember* in University 2, and *Apply* in University 3.

While the frequency of e-assessment tasks (Table 4 and 5) indicates a predominance of LOTS (298 LOTS tasks and 202 HOTS tasks), the results shift to the advantage of HOTS when calculated per credit as explained in section 3.2. (Figure 1). The data analysis revealed that three of four universities investigated in this study included more HOTS e-assessment tasks than LOTS e-assessment tasks. Only University 3 had a lower rate of HOTS than LOTS e-assessment tasks. It should, however, be noted that one 4 credits (of 30 credits) module in linguistics is missing from University 3, as mentioned in section 3.2.

Table 5. Total number of tasks/thinking level per university

Level	Uni 1	Uni 2	Uni 3	Uni 4	Total
<b>Remember</b>	6	27	1	0	35
<b>Understand</b>	49	125	35	41	250
<b>Apply</b>	3	0	5	6	14
<b>Analyze</b>	36	50	10	56	152
<b>Evaluate</b>	8	12	4	19	43
<b>Create</b>	4	3	0	0	7
<b>Total</b>	106	217	55	122	500

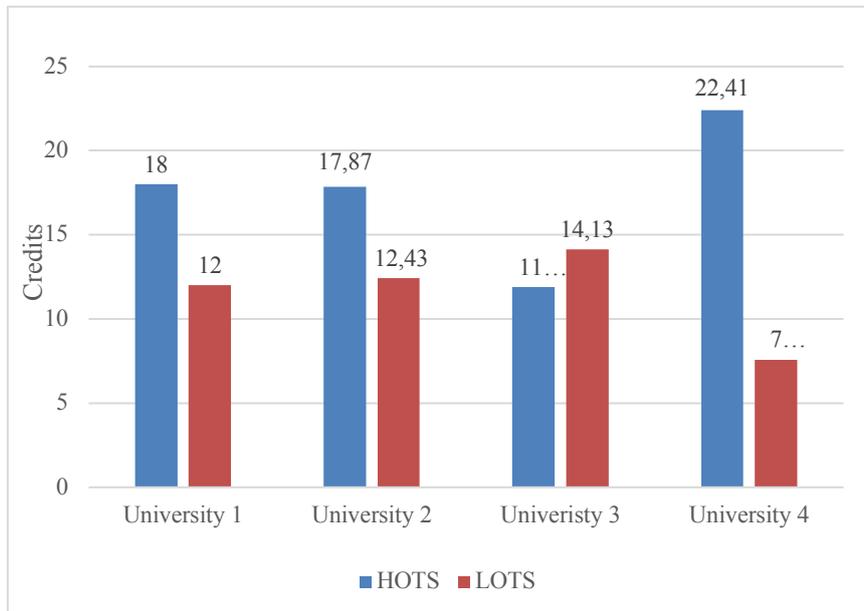


Figure 1. HOTS/LOTS e-assessment tasks per university

#### 4.2 Research question 2: How do linguistic and literature modules in Swedish universities differ in their e-assessment of HOTS?

Further analysis of the data was carried out centering on the division between linguistics and literature modules. Language courses at Swedish universities are traditionally divided equally between linguistics and literature modules, which was also the case with the courses in the present study. A Chi-square test of independence was calculated comparing the frequency of HOTS e-assessment in linguistic and literature modules. A significant interaction was found ( $\chi^2(1) = 84.85, p = 0.00$ ). The effect size measured by Cramer's V was moderate, 0.46. The higher frequency of HOTS e-assessment tasks in literature modules is also illustrated in Figure 2 and Figure 3.

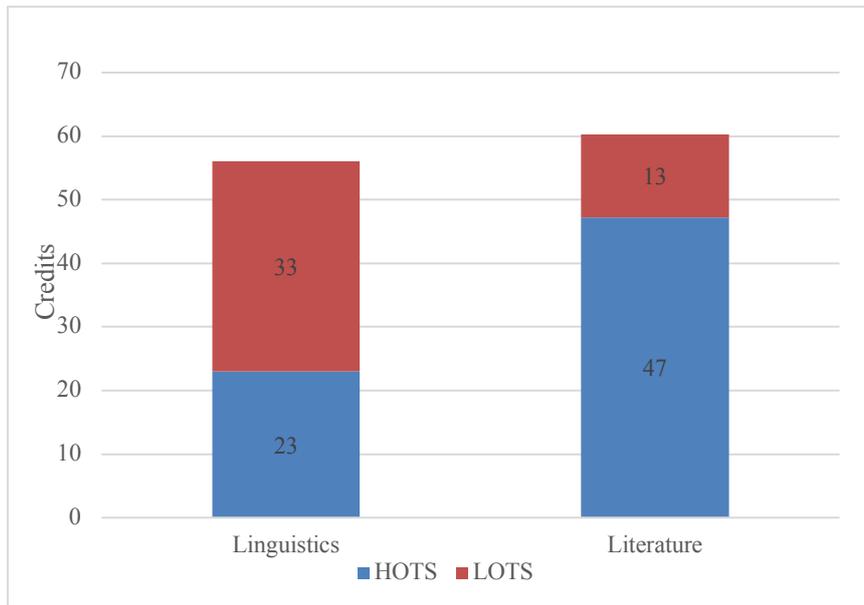


Figure 2. HOTS/LOTS in literature and linguistics modules

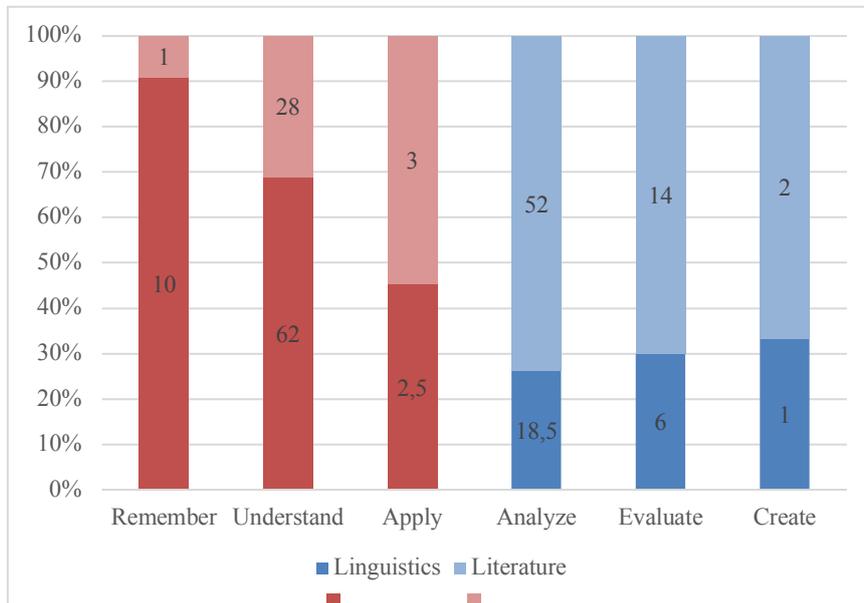


Figure 3. Distribution of thinking levels in linguistics and literature modules in percent

The data analysis revealed that the predominance of HOTS e-assessment tasks in literature modules compared with linguistics modules could perhaps be explained by the nature of the e-assessment tasks used in each module (Figure 4 and Table 6). While written exam questions, seminar questions, and compulsory study questions were the most common assessment forms in both linguistic and literature modules, the frequency of these differs greatly. Although compulsory study questions were much more frequently used in literature modules, the linguistics modules included more written exam questions and seminar questions. Furthermore, it was suggested that the high number of written essays in literature modules contribute to an explanation of the results. The literature modules in this study included 15 essays, while the linguistic modules included 8 essays. In comparison to the number of e-assessment tasks, these numbers may seem insignificant. However, the essays often formed an important and major part of the e-assessment tasks given to the students. As HOTS were calculated per credit, and not per frequency, these essays probably contributed to the results.

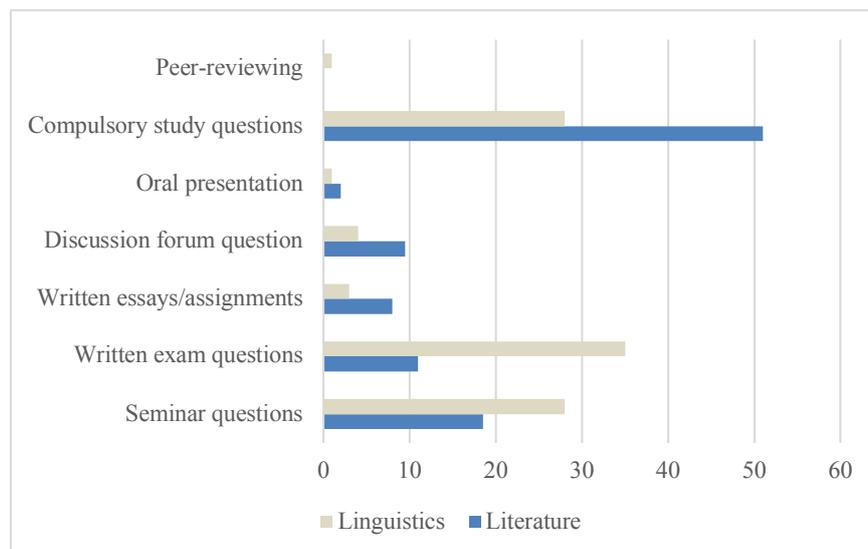


Figure 4. Type of e-assessment items in linguistic and literature modules in percent

Table 6. Type of e-assessment items in linguistic and literature modules

	Literature	Linguistics
<b>Seminar questions</b>	33	89
<b>Written exam questions</b>	19	114
<b>Written essay/assignment</b>	15	8
<b>Discussion forum question</b>	17	14
<b>Oral presentation</b>	4	3
<b>Compulsory study questions</b>	90	92
<b>Peer-reviewing</b>	0	2
<b>Total</b>	178	322

*4.3 Research question 3: To what extent does the use of HOTS e-assessment correspond to the presence of collaborative e-assessment tasks?*

A chi-square test of independence was calculated to compare the frequency of collaborative e-assessment tasks among HOTS and LOTS e-assessment tasks. A statistically significant interaction was found ( $\chi^2(1) = 26.99, p = 0.00$ ). The effect size measured by Cramer's V was small to moderate, 0.23. The proportion of collaborative e-assessment tasks among HOTS e-assessment tasks (27%) was slightly greater than the expected proportion (21%), whereas the proportion of non-collaborative e-assessment tasks among HOTS e-assessment tasks (16%) was lower than the expected proportion (22%). Regarding collaborative e-assessment tasks in LOTS e-assessment tasks, the proportion of these (22%) was lower than the expected proportion (28%), whereas the proportion of non-collaborative e-assessment tasks among LOTS e-assessment tasks (35%) was higher than the expected proportion (29%). Overall, these results suggest that collaborative e-assessment tasks are slightly more common in e-assessment tasks that focus on HOTS than on LOTS.

*5. Discussion*

The results of the present study provide a glimpse of the type of e-assessment tasks that are used in online language courses at Swedish universities. Investigating the first research question, which asked to what extent e-assessment tasks used in online EFL courses at Swedish

universities focus on HOTS, the overall tendency indicates that higher-order thinking skills are assessed in these courses. As no previous research has been carried out on HOTS e-assessment in language courses, the results can only be discussed in light of what is known about the implementation of thinking skills in e-assessment and assessment in general. The findings do not entirely align with previous studies that show that assessment tasks are often formed to tap LOTS rather than HOTS (FitzPatrick et al. 2015; Fensham and Bellocchi 2013; Köksal and Ulum 2018). While previous literature does not reveal much about if and how HOTS is assessed through e-assessment tasks, there are indications that HOTS e-assessment is rather uncommon in higher education (Boitshwarelo, Reedy, and Billany 2017; McNeill, Gosper, and Xu 2012). The high frequency of HOTS e-assessment in the participating EFL courses suggests that online language teachers are aware of the importance of higher-order thinking skills in language education.

As previously discussed, multiple choice questions are a common e-assessment alternative and much of the previous research conducted within the field explores the use of MCQs. Boitshwarelo et al.'s (2017) literature review of studies on online tests in higher education showed that test items were often in the format of MCQs and that these were mostly appropriate for assessing lower-order thinking skills. However, the universities investigated in the present study made very little use of MCQs, which may partially explain the difference between the results of the present study and previous studies.

Furthermore, the e-assessment tasks in this study encourage HOTS to a higher extent than discovered in previous studies on assessment in EFL courses. For example, Köksal and Ulum (2018) found that exam questions used in Turkish university EFL courses entirely lacked higher-order thinking questions. Likewise, studies in other subjects indicate a high focus on LOTS in assessment tasks (FitzPatrick et al. 2015; Fensham and Bellocchi 2013). Based on previous research, there seems to be a general inclination towards lower-order thinking in EFL teaching (Wu and Pei 2018; Jabr 2003). It should, however, be noted that the majority of studies carried out so far have been conducted in countries and cultures that are highly different from the Swedish context. Additionally, previous studies have mainly focused on the teaching of thinking skills in EFL courses that do not include a literature module. While it could be argued that the high focus of HOTS compared with

other studies within the field is perhaps more a result of differences in course design than in focus on thinking skills, the relatively high levels of HOTS e-assessment in linguistics modules point to another explanation. Although LOTS e-assessment tasks were more frequent than HOTS e-assessment tasks, the difference was not as great as detected in other studies (Köksal and Ulum 2018). While more research is needed to fully understand why HOTS assessment is more common in Swedish online EFL courses than in other investigated courses, possible explanations could be the course design and teachers' awareness of the importance of HOTS.

The paper's second research question was: how do linguistic and literature modules in Swedish universities differ in their e-assessment of HOTS? Investigating this question, it was found that although levels of HOTS e-assessment were higher than those of LOTS e-assessment in most universities, differences emerged when the data were further broken down between linguistics and literature modules. Literature modules consisted of significantly more HOTS e-assessment tasks than linguistics modules. As mentioned in the data analysis, the difference between e-assessment tasks in linguistics and literature modules could possibly be explained by the type of e-assessment tasks used in these courses. Assessment tasks that require discussions and reflections are more likely to assess and develop students' higher-order thinking skills. Even though most written essays and assignments were not collaborative learning experiences, the students were required to discuss and reflect on the essay topic.

However, the type of e-assessment used in each module probably does not form the entire explanation of the difference between linguistics and literature modules. The nature of the two subjects inevitably leads to different types of teaching and assessment. It may be assumed that linguistic knowledge is, in many ways, more easily assessed through LOTS assessment tasks. However, understanding a theory or a piece of literature often necessitates higher-order thinking such as analysis and evaluation. Thus, it is not surprising that literature modules include more HOTS e-assessment than linguistic modules. However, that the difference between these modules were that significant was not expected. Furthermore, as discussed in the literature review, previous studies indicate a relationship between L2 learning and thinking skills; it seems, therefore, to be in the interest of online EFL instructors to develop e-

assessment tasks that support students' HOTS. This may be especially the case in linguistics courses where there is a higher focus on language proficiency and performance than in literature courses. The scarce use of HOTS e-assessment in linguistics courses could perhaps be explained by difficulties faced by instructors in developing these tasks. To this author's knowledge there exists no research on online EFL instructors' views on the importance of HOTS e-assessment and the development of such. To fully understand the results of this study, an investigation of online EFL instructors' views and experiences is needed.

The slightly greater proportion of collaborative e-assessment tasks among HOTS e-assessment tasks than the expected proportion as well as the smaller proportion of collaborative e-assessment tasks among LOTS e-assessment aligned with the author's hypothesis of collaborative tasks being seen as more suitable for developing students' HOTS. Previous studies, as mentioned in section 2.4, have found that collaborative learning in general is beneficial for the development of critical thinking (Gokhale 1995; Li 2011). It was, therefore, assumed that collaborative e-assessment tasks would more often be designed to tap HOTS. However, considering the small to moderate effect size, the author of this paper cannot give an affirmative answer to the third research question, which asked if the presence of HOTS e-assessment in online EFL courses corresponds with collaborative e-assessment tasks.

According to the sociocultural theory, learning takes place in interaction and collaborative e-assessment tasks should thus be optimal for the purpose of developing HOTS. As mentioned in 2.4, well-designed online learning and e-assessment tasks afford opportunities for collaboration. The present study is limited in that it only investigates the format and content of the e-assessment tasks given to the students, and it does not look into students' responses and discussions. It is possible that an investigation of student discussions in seminars and forums would reveal a different result. In order to fully understand if and how students develop higher cognitive skills, there is a need to investigate students' collaborative efforts. This would ideally also reveal if collaborative e-assessment tasks are better suited for the purpose of developing HOTS than non-collaborative e-assessment tasks.

Moreover, the present study would have benefitted from comparing the results to that of similar f2f courses. Such a comparison would reveal if and how HOTS and LOTS e-assessment tasks differ from regular

assessment tasks. While no such comparison can be made at this moment, it is likely that the differences are not very great. Online courses are often designed with the f2f course as a model and similar assessment tasks can be found in both types of courses. Moreover, many f2f courses are designed so that much of students' work is completed online. Although a comparison of this kind would be interesting, the purpose of this study has been to investigate the topic in online courses. As this study is part of a larger project on HOTS e-assessment in online EFL courses, it was not possible at this time to investigate HOTS assessment in f2f courses.

### *6. Conclusion*

The present study makes a small contribution to the field of assessing thinking skills in online L2 learning by investigating HOTS e-assessment tasks used in four online English courses given at Swedish universities. This investigation revealed that HOTS e-assessment tasks generally were more common than LOTS e-assessment tasks. However, there was a significant difference in the use of HOTS between linguistics and literature modules. Literature modules included more HOTS e-assessment tasks, which seems to be a consequence of both a higher frequency of written essays and the nature of the subject matter.

The study of e-assessment tasks helped us gain a better understanding of the type of assessment tasks used in online language courses, as well as providing us with empirical evidence of the possibility of developing e-assessment tasks that assess and develop higher-order thinking skills. All universities included in this study used e-assessment tasks that focus on HOTS, which proves that the implementation of such is not impossible. Well-planned and -designed e-assessment tasks will create outcomes that are both linguistically and cognitively beneficial. It is thus important that teachers working within computer assisted language learning are given appropriate training in developing e-assessment tasks that support both of these skills.

Among the limitations of the present study is the relatively small set of data. Even though this study looks into the e-assessment tasks used in 50 percent of all online intermediate English courses given at Swedish universities, these only consist of four courses. A higher participation would have provided a more representative picture of how HOTS are

assessed. Moreover, the missing 4 credit module in linguistics from University 3 would have given a more complete picture of used e-assessment tasks.

It should be remembered that the categorizations of the e-assessment tasks are based on the author's interpretation. Assuring an acceptable interrater reliability would have been ideal; however, this was not possible in the present study. The process of categorizing the e-assessment tasks according to the revised Bloom's taxonomy as well as examples of the categorization are described in detail in sections 3.3 and 3.4. While acknowledging this limitation, it is hoped that the detailed explanation of the categorization will give sufficient support to give credit to this study.

Further studies would benefit from investigating students' responses and discussions in the given e-assessment tasks. This would hopefully provide a deeper understanding of if and how students use higher-order thinking skills in online language courses. Moreover, there is a need to investigate the relationship between higher-order thinking skills and second language proficiency in CALL. An investigation of this sort would do good in researching if students participating in courses with a high frequency of HOTS e-assessment actually develop these skills; thus, ascertaining the effectiveness of HOTS e-assessment.

#### *Acknowledgements*

I would like to thank the universities that have chosen to participate in this study. This research project would not have been possible without your help and collaboration. I would also like to show gratitude to Christina Lindqvist, Pia Köhlmyr and the reviewers for providing comments that greatly improved the manuscript.

#### *Competing Interests*

The author has no competing interests to declare.

#### *References*

Afflerbach, Peter, Byeong-Young Cho, and Jong-Yun Kim. 2011. The assessment of higher order thinking in reading. In *Assessment of*

- higher order thinking skills*, edited by Gregory Schraw and Daniel R. Robinson, 185-218, Charlotte: Information Age Publishing.
- Alcón, Eva Guzman 1993. High cognitive questions in Nns group classroom discussion: do they facilitate comprehension and production of foreign language? *RELC Journal* 24: 73-85.
- Ali, Siti Noridah. 2012. "Malaysian polytechnic lecturers' teaching practices with ICT utilization to promote higher-order thinking skills". PhD diss., Iowa State University.
- Anderson, Lorin W., and David R Krathwohl. 2001. *A Taxonomy for learning, teaching, and assessing*. Boston: Allyn and Bacon.
- Anderson, Lorin W., David R Krathwohl, Peter W Airasian, Kathleen A Cruikshank, Richard E Mayer, Paul R Pintrich, James Rath, and Merlin C Wittrock. 2014. *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Harlow: Pearson Educational Limited.
- Arum, Richard, and Jospia Roksa. 2010. *Academically adrift: limited learning in college campuses*. Chicago: University of Chicago.
- Bates, A. W. Tony, and Sangrá Albert. 2011. *Managing technology in higher education*. San Fransisco: Jossey-Bass.
- Biggs, John, and Catherine Tang. 2007. *Teaching for quality learning at university*. New York: The Society for Research into Higher Education.
- Bloom, Benjamin S. 1956. *Taxonomy of educational objectives*. New York: David McKay Company.
- Boitshwarelo, Bopelo, Alison Kay Reedy, and Trevor Billany. 2017. Envisioning the use of online tests in assessing twenty-first century learning: a literature review. *Research and Practice in Technology Enhanced Learning* (12): 1-16. doi:10.1186/s41039-017-0055-7.
- Brookhart, Susan M. 2014. *How to design questions and tasks to assess student thinking*. Alexandria: Association for Supervision & Curriculum Development.
- Bryan, Cordelia, and Karen Clegg. 2006. *Innovative assessment in higher education*. Abington: Routledge.
- Caplan, Dean, and Rodger Graham. 2008. The development of online courses. In *The theory and practice of online learning*, edited by Terry Anderson, 245-263, Edmonton: AU Press.
- Center for University Teaching, Learning, and Assessment. *Assessment - Bloom's taxonomy action verbs*. Accessed June 15, 2020.

- <https://www.cte.cornell.edu/documents/Assessment%20-%20Blooms%20Taxonomy%20Action%20Verbs.pdf>.
- Chaiklin, Seth. 2003. The zone of proximal development in Vygotsky's analysis of learning and instruction. In *Vygotsky's educational theory in cultural context*, edited by Alex Kozulin, Boris Gindis, Vladimir S. Ageyev and Suzanne M. Miller, 39-63. Cambridge: Cambridge University Press.
- Chen, Mei-Hui. 2010. "Infusing thinking skills into an L2 classroom: a case study of an innovation in a Taiwanese university." PhD diss., Newcastle university.
- Davidson, Bruce W., and Rodney A Dunham. 1997. Assessing EFL student progress in critical thinking with the Ennis-Weir critical thinking essay test. *JALT Journal* 19(1): 43-57.
- DeWaesche, Scott A. 2015. Critical thinking, questioning and student engagement in Korean university English courses. *Linguistics and Education* 32: 131-47. doi:10.1016/j.linged.2015.10.003.
- Ebadi, Saman, and Masoud Rahimi. 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): 617-651. doi:10.1080/09588221.2018.1449759.
- Facione, Peter A., and N. C. Facione. 1994. *The California critical thinking skills test: test manual*. Milbrae: California Academic Press.
- Falchikov, Nancy, and Karen Thompson. 2008. Assessment: what drives innovation? *Journal of University Teaching & Learning Practice* 5(1): 50-60.
- Fensham, Peter J., and Alberto Bellocchi. 2013. Higher order thinking in chemistry curriculum and its assessment. *Thinking Skills and Creativity* 10: 250-64. doi:10.1016/j.tsc.2013.06.003.
- FitzPatrick, Beverly, John Hawboldt, Daniel Doyle, and Terri Genge. 2015. Alignment of learning objectives and assessments in therapeutics courses to foster higher-order thinking. *American Journal of Pharmaceutical Education* 79(1): 1-8.
- FitzPatrick, Beverly, and Henry Schulz. 2015. Do curriculum outcomes and assessment activities in science encourage higher order thinking? *Canadian Journal of Science, Mathematics and Technology Education* 15(2): 136-54. doi:10.1080/14926156.2015.1014074.

- Ghanizadeh, Afsaneh. 2016. The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *Higher Education* 74(1): 101-14.  
doi:10.1007/s10734-016-0031-y.
- Gokhale, Anuradha. 1995. Collaborative learning enhances critical thinking. *Journal of Technology Education* 7(1): 22-30.  
doi:10.21061/jte.v7i1.a.2.
- Gorin, Joanna S., and Svetina Dubravka. 2011. Test design with higher order cognition in mind. In *Assessment of higher order thinking skills*, edited by Gregory Schraw and Daniel R Robinson, 121-149, Charlotte: Information Age Publishing.
- Jabr, Dua' Ahmad Faheem. 2003. Teaching critical thinking in the English language classroom: the case of Palestine. *Mediterranean Journal of Educational Studies* 8(2): 137-62.
- Khoo, Elaine, and Bronwen Cowie. 2011. A framework for developing and implementing an online learning community. *Journal of Open, Flexible and Distance Learning* 15(1): 47-59.
- Krathwohl, David R. 2002. A revision of Bloom's taxonomy: an overview. *Theory Into Practice* 41(4): 212-18.  
doi:10.1207/s15430421tip4104\_2.
- Köksal, Dincay, and Ömer Gökhan Ulum. 2018. Language assessment through Bloom's taxonomy. *Journal of Language and Linguistic Studies* 14(2): 76-88.
- Lantolf, James P., James L Thorne, and Matthew E Poehner. 2015. Sociocultural theory and second language development. In *Theories in second language acquisition: an introduction*, edited by Bill VanPatten and Jessica Williams, 207-226, New York: Routledge.
- Lee, Jihyun, and Hyoseon Choi. 2017. What affects learner's higher-order thinking in technology-enhanced learning environments? The effects of learner factors. *Computers & Education* 115: 143-52.  
doi:10.1016/j.compedu.2017.06.015.
- Leighton, Jacqueline P. 2011. A cognitive model for the assessment of higher order thinking in students. In *Assessment of higher order thinking skills*, edited by Gregory Schraw and Daniel R Robinson, 151-184, Charlotte: Information Age Publishing.
- Li, Li. 2011. Obstacles and opportunities for developing thinking through interaction in language classrooms. *Thinking Skills and Creativity* 6(3): 146-158. doi:10.1016/j.tsc.2011.05.001.

- Li, Li. 2016. Thinking skills and creativity in second language education: where are we now? *Thinking Skills and Creativity*, 22: 267-272. doi:10.1016/j.tsc.2016.11.005.
- Liaw, Meei-Ling. 2007. Content-based reading and writing for critical thinking skills in an EFL context. *English Teaching & Learning* 31(2): 45-87.
- Manalo, Emmanuel, and Chris Sheppard. 2016. How might language affect critical thinking performance? *Thinking Skills and Creativity* 21: 41-49. doi:10.1016/j.tsc.2016.05.005.
- McNeill, Margot, Maree Gosper, and Jing Xu. 2012. Assessment choices to target higher order learning outcomes: the power of academic empowerment. *Research in Learning Technology* 20: 283-96. doi:0.3402/rlt.v20i0.17595.
- Mohammadi, Elham, Gholam Reza Kiany, Reza Ghaffar Samar, and Ramin Akbari. 2015. Appraising pre-service EFL teachers' assessment in language testing course using revised Bloom's taxonomy. *International Journal of Applied Linguistics and English Literature* 4(4): 8-20. doi:10.7575/aiac.ijalel.v.4n.4p.8.
- Possin, Kevin. 2014. Critique of the Watson-Glaser Critical Thinking Appraisal Test: the more you know, the lower your score. *Informal Logic* 34(4): 393-416.
- Riazi, Mehdi A. 2010. Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using Bloom's taxonomy. *The Electronic Journal for English as a Second Language* 13(4): 1-16.
- Rose, Susan, Nigel Spinks, and Ana Isabel Canhoto. 2015. *Management research: applying the principles*. New York: Routledge.
- Schraw, Gregory, and Daniel R Robinson. 2011. *Assessment of higher order thinking skills*. Charlotte: Information Age Club.
- Soodmand Afshar, Hassan, and Hossein Movassagh. 2014. On the relationship among critical thinking, language learning strategy use and university achievement of Iranian English as a foreign language majors. *The Language Learning Journal* 45(3): 382-98. doi:10.1080/09571736.2014.914238.
- Toyoda, Etsuko. 2015. Relationship between higher-order thinking skills and L2 performance. *Electronic Journal of Foreign Language Teaching* 12(2): 200-218.

- Ulum, Ömer Gökhan. 2016. A descriptive content analysis of the extent of Bloom's taxonomy in the reading comprehension questions of the course book Q: Skills for success 4 reading and writing. *The Qualitative Report* 21(9): 1674-1683.
- Vygotsky, Lev S. 1978. *Mind in society*. Cambridge: Harvard University Press.
- Wu, Yidi, and Zhengwei Pei. 2018. An investigation of critical thinking manifested in the questions of EFL textbooks for tertiary-level English majors of China. *American Journal of Education and Learning* 3(2): 72-84. doi:10.20448/804.3.2.72.84
- Yang, Ya-Ting C., and Jeffrey Gamble. 2013. Effective and practical critical thinking-enhanced EFL instruction. *ELT Journal* 67(4): 398-412. doi:10.1093/elt/ccto38.
- Yang, Rong. 2017. The use of questions in a synchronous intercultural online exchange project. *ReCALL* 30(1): 112-30. doi:10.1017/S0958344017000210
- Yang, Ya-Ting C., Timothy J. Newby, and Robert L. Bill. 2005. Using Socratic questioning to promote critical thinking skills through asynchronous discussion forums in distance learning environments. *American Journal of Distance Education* 19(3): 163-8. doi:10.1207/s15389286ajde1903\_4.