

Syntactic and lexical development in an intensive English for Academic Purposes programme

**Author final version published in Journal of Second Language Writing
2015 October**

Authors: Diana Mazgutova and Judit Kormos

Affiliation Department of Linguistics, Lancaster University

Abstract

This study investigates how the lexical and syntactic characteristics of L2 learners' academic writing change over the course of a one-month long intensive English for Academic Purposes (EAP) programme at a British university. The participants were asked to produce two argumentative essays, at the beginning and at the end of the EAP course, which were analysed using measures that are theoretically motivated by previous research in corpus linguistics, systemic functional linguistics, and developmental child language acquisition. The results indicate improvements, with regard to lexical diversity, both for intermediate-level students who were preparing for undergraduate university studies in the UK and upper-intermediate level participants who were planning to continue their studies at postgraduate level. The academic argumentative texts of the students in the lower proficiency group also demonstrate development in noun-phrase complexity and in the use of genre-specific syntactic constructions. The findings suggest that despite no explicit focus on lexis and syntax in the EAP programme, by the end of the course the students' writing exhibited a developmentally more advanced repertoire of lexical and syntactic choices that are characteristic of expository texts in academic contexts.

Keywords: Second language writing, EAP writing, lexical diversity, lexical sophistication, syntactic complexity, writing development

Introduction

Learning to write effectively in an academic context is very important, not only because it is often the only means by which students' content knowledge is assessed in a large number of disciplines, but also because producing academic texts helps students to become members of a discourse community as well as to gain new knowledge through writing (Hirvela, 2011; Hyland, 2011). The development of L2 learners' academic writing ability has mostly been investigated in terms of improvements in various assessment criteria, such as cohesion, coherence and organisation, as well as overall grades (see, for example, Green & Weir, 2002). It is only recently that writing research and studies in the field of English for Academic Purposes (EAP) have started to focus on the linguistic features of students' writing and how they improve along with developments in proficiency in various instructional contexts (see, for example, the collection of studies introduced in a recent special issue of the *Journal of Second Language Writing*, guest edited by Connor-Linton and Polio, 2014). The development of the syntactic complexity of students' writing has been at the centre of a number of studies in recent years (e.g., Byrnes, 2009; Byrnes & Sinicrope, 2008; Crossley & McNamara, 2014; Shaw & Liu, 1998; Vyatkina, 2013), but only a few studies have considered lexical development in conjunction with syntactic changes in students' written production (for exceptions see Bulté & Housen, 2014; Storch & Tapper, 2009; Verspoor, Lowie, & van Dijk, 2008; Vyatkina, 2012). In our study, we investigated how the lexical and syntactic characteristics of L2 learners' writing changed during the course of an intensive EAP programme which aims to prepare international students for university studies at undergraduate and postgraduate levels in the UK. This research helps us to understand how

key linguistic features of academic writing develop and thereby contribute to supporting the more effective and efficient expression of L2 writers' thoughts and arguments.

Our research specifically focuses on linguistic features that have been shown to be typical of academic writing among L1 writers and that exemplify advanced and experienced writers' texts (Biber & Gray, 2010; Byrnes, 2009; Byrnes & Sinicrope, 2008; Halliday & Martin, 1993/1996). In this respect, we have taken a novel approach to analysing complexity since, in line with Biber, Gray, and Poonpon (2011), we argue that the complexity of learners' output should be considered with reference to the mode, genre, and communicative demands of the particular task to be performed. This position is somewhat different from that of Bulté and Housen (2014), who describe absolute complexity as “*objective* inherent properties of linguistic units and/or systems thereof” and relative complexity as “the cost and difficulty of processing” (p. 43, italics in original). In this paper we propose that in addition to these two theoretically and empirically useful aspects of complexity, the operationalisation of complexity in written and spoken performance should also consider the linguistic characteristics of the given genre or task-type. In the field of corpus linguistics, Biber and Gray (2010) and Biber et al. (2011), and, in systemic linguistics, Halliday and Martin (1993/1996) have convincingly shown that different linguistic features characterise speech and writing. Academic writing relies more on phrasal embedding than speech and is typically characterised by complex nominalisation and the use of abstract and compound nouns (Fang, Schleppegrell, & Cox, 2006; Norris & Ortega, 2009). The complexity demands of writing and speech do not only differ across modes but also across genres. This is seen in the findings of Nippold (2004) and Berman and Nir-Sagiv (2007), in the field of developmental child language acquisition, which show that in expository texts one can find a higher number of relative clauses and passive constructions and more complex noun phrases than in narratives. In line with these arguments, our study uses measures of syntactic and lexical complexity that

are theoretically motivated by previous research in corpus linguistics, systemic functional linguistics, and developmental child language acquisition.

Some recent longitudinal studies of L2 writing have examined the development of syntactic complexity, but these were mostly conducted with university learners of German at beginning (Vyatkina, 2012, 2013) and advanced levels (Byrnes, 2009; Byrnes & Sinicrope, 2008). The studies of Byrnes (2009) and Byrnes and Sinicrope (2008) reveal that – parallel with developments in proficiency – increase in nominalisation, the use of relative clauses, and grammatical metaphor can be observed. Vyatkina (2012) found that beginner learners' writing in German became lexically more varied and was characterised by longer sentences and finite verb-units and more frequent subordination as they progressed in their language studies. Verspoor et al. (2008) conducted a longitudinal case study on the development of the lexical and syntactic features of academic writing in English. Their participant, a Dutch university student, demonstrated development in terms of word and sentence length in his writing, but the growth in these features was complementary; when one of them increased, the other decreased. Vyatkina's (2012) longitudinal study showed parallel development in lexical and syntactic features of students' writing at a lower level of proficiency.

The findings concerning linguistic development in EAP programmes have been mixed. Most of these courses are relatively short and are not, or only indirectly, focused on the syntactic and lexical aspects of writing; hence, it is understandable that limited development in these areas can be observed (see e.g., Ortega, 2003). In a study conducted in a UK university context, Shaw and Liu (1998) found that linguistic accuracy in terms of the frequency of errors and complexity with regards to nominalisation and subordination did not change. Nevertheless, their research revealed that, in a number of other areas, such as the level of formality and impersonality, students' writing exhibited increased use of the characteristic features of academic genres at the end of the programme. In contrast, the

results of Polio, Fleck, and Leder (1998) and Storch and Tapper (2009) indicated significant improvement in the accuracy of students' writing in an EAP programme. Storch and Tapper's results also revealed that there were significantly more academic words found in the students' essays at the end of the EAP course.

In a recent series of analyses of writing development in the Michigan State University (MSU) corpus of descriptive essays (for a description of the corpus see Connor-Linton and Polio, 2014), Bulté and Housen (2014) found that students' writing exhibited increased use of less frequent words and a higher complex sentence ratio, compound complex sentence ratio, and subclause ratio at the end of the academic writing programme. They also observed that finite clauses, sentences, and T-units became significantly longer in students' descriptive essays. In Crossley and McNamara's (2014) computational analysis of the same dataset, "longer noun phrases, less syntactic similarity between sentences, fewer verb phrases, more words before the main verb, and more negation" (p. 73) were the differentiating features of students' writing at the beginning and at the end of the academic writing course. A multidimensional analysis of the linguistic characteristics of the texts also showed that the learners' essays were increasingly characterised by a nominal writing style and elaborated description by the end of the programme (Friginal & Weigle, 2014). However, no parallel improvement in accuracy could be observed (Polio & Shea, 2014).

Although previous studies offer useful insights into lexical and syntactic development in L2 writing in general, their findings might not apply to expository and argumentative academic texts, which have specific genre and linguistic characteristics. Most previous longitudinal research has used a variety of task types, prompts, and genres for data elicitation at different time points, which makes it difficult to separate the effect of linguistic development on students' output from potential task and genre effects (for an exception see the series of studies using the MSU corpus reviewed above). No previous studies have been

conducted in the contexts of the short and highly intensive pre-sessional EAP courses, which are frequently offered to students in the UK who fall short of the language requirements for university by a small margin. Furthermore, no research has investigated the impact of EAP courses on students' development when the participants show differences in the level of proficiency they have when they enrol in these courses. Therefore, little is known about the effectiveness of these courses and to what extent they contribute to the development of students' linguistic expression in writing. Our study aims to fill this gap by answering the following research question: How do the lexical and syntactic features of argumentative writing change in an intensive EAP programme (a) in the case of higher proficiency level students who have already completed their undergraduate degree in their home country, and (b) in the case of lower proficiency students who intend to undertake undergraduate studies in the UK?

Proposed framework for analysing lexical and syntactic development in academic writing in our study

In our study, we take two complementary approaches to the analysis of lexical and syntactic features of learners' writing. On the one hand, we adapt Bulté and Housen's (2012) definition of system complexity, which "refers to the degree of elaboration, the size, the breadth, the width, or richness of the learners' L2 system" (p. 25). From this perspective, we focus on what Bulté and Housen call "grammatical diversity," and which has been operationalised in task-based studies as syntactic complexity. At the level of lexis, we consider "systemic lexical complexity" (Bulté & Housen, 2012, p. 28) modified to incorporate Jarvis' (2013) recent work on lexical diversity. In line with Jarvis, the terms of "density, diversity and sophistication" in Bulté and Housen's framework are all subsumed under the construct of lexical diversity, which Jarvis sees as consisting of rarity, volume, variability, evenness, disparity, and dispersion. On the other hand, we also take a developmental perspective and,

as we argued above, not all elements of systematically complex language are relevant for the investigation of development within and across specific genres. This framework of the study of linguistic development is followed in the field of systemic linguistics (Halliday & Martin, 1993/1996) and in developmental studies of L1 writing (e.g., Berman & Nir-Sagiv, 2007).

The operationalisation of lexical and syntactic development in our study

In line with the two frameworks outlined above, a number of variables were selected to evaluate the lexical and syntactic development observed in our participants' academic writing (for a summary of the variables used in the current study see Tables 1, 2, and 3). A sizable amount of research has demonstrated the ability of English as a Second or Foreign Language (ESL/EFL) writers with higher language proficiency to produce texts with greater lexical diversity (e.g., Engber, 1995; Grant & Ginther, 2000; Jarvis, 2002; Reppen, 1994). A range of lexical diversity and sophistication measures have been proposed in the second language literature, and a number of developmental studies have investigated the extent to which these measures can be used as a reliable indication of L2 proficiency (Bulté & Housen, 2014; Wolfe-Quintero, Inagaki, & Kim, 1998). As mentioned above, Jarvis (2013) recently argued that the most important six facets of lexical variety include rarity, volume, variability, evenness, disparity, and dispersion. Jarvis suggests that volume (i.e., the total number of words produced in the text), evenness (i.e., "how evenly word tokens are distributed across types," p. 23), and dispersion ("the mean distance between different tokens of the same type," p. 30) are highly intercorrelated. As in our research, the number of words the students had to produce was specified in the task prompt; for the analysis of lexical diversity we decided to focus on the rarity, variability, and disparity dimensions.

The most common measure of lexical variability used in Second Language Acquisition (SLA) research is Type-Token Ratio (TTR) (Templin, 1957). TTR can be defined as the ratio of the number of different “lexical items,” i.e., words, to the total number of words. However, there have been severe criticisms of the use of TTR. As pointed out by Engber (1995), there are two primary limitations when using this measure. Firstly, it is difficult to decide on a fine-grained definition of a lexical item. Secondly, TTR is highly dependent on the total number of words in the text, i.e., “the ratio of tokens tends to decrease as essay length increases” (Engber, 1995, p. 145). To avoid these problems with the use of TTR, the *measure of textual lexical diversity* (MTLD), the most recent variant of the TTR, which has been found to be least affected by text length (Jarvis, 2012; McCarthy, 2005), was chosen as an index of lexical variability. MTLD was computed with the aid of Coh-Metrix 3 (Graesser, McNamara, & Kulikowich, 2011; Graesser, McNamara, Louwerse, & Cai, 2004).

Analysing texts for lexical variability also involves evaluation of the distribution of various parts of speech, such as nouns, adjectives, adverbs, and verbs in the text (Vajjala & Meurers, 2013). Although lexical verbs are less frequent in academic writing than in conversation (Biber, Johansson, Leech, Conrad, & Finegan, 1999), they play an important role in “expressing personal stance, reviewing the literature, quoting, expressing cause and effect, summarising and contrasting” (Granger & Paquot, 2009, p. 193). Therefore, we applied an additional measure of lexical variability in the study, i.e., *squared verb variation*, computed by Synlex (Lu, 2010, 2011, 2012), which was previously found to be an appropriate predictor of oral language proficiency by Lu (2012).

In order to assess lexical rarity, we used the *log frequency of content words*, estimated by Coh-Metrix 2.0 (Graesser et al., 2004) and based on the CELEX lexical database corpus, which contains the frequencies of words as different parts of speech (Baayen, Piepenbrock, & van Rijn, 1993). This measure can be considered to reflect the rarity of words used in the text

(Jarvis, 2013), and similar counts based on the British National Corpus (BNC) have been applied in previous research as an index for lexical richness (see, e.g., Edwards & Collins, 2013; Laufer & Nation, 1995). The use of the log frequency measure instead of the raw frequency value was motivated by Davis' (2005) suggestion that the log statistics can differentiate among the frequency values of rare words better than the lemmatised count.

In order to assess disparity, that is the “degree of differentiation between lexical types in a text” (Jarvis, 2013, p. 13), we applied the measure of the *latent semantic analysis* (LSA) index calculated with the help of Coh-Metrix 2.0. This computerised tool establishes the relevance of ideas to the topic and determines the similarity of meaning between words, sentences and paragraphs. Responses that more specifically address the prompt tend to show higher latent semantic analyses values (Crossley & McNamara, 2013). The index of LSA has recently been proposed by Jarvis (2013) as a potentially useful measure for the operationalisation of lexical disparity.

From a developmental perspective, we also found it important to investigate how the lexical characteristics of students' writing reflect genre-relevant lexical choice. For this purpose, the percentage of academic words in written texts was estimated by means of the *academic word list* measure using the computer program Vocabprofiler BNC (Cobb, 1994; Heatley & Nation, 1994). The academic word list constitutes a group of lower frequency words which are typically found in academic texts. It is derived from a corpus of academic texts drawn from the sub-corpora of arts, commerce, law, and science (Coxhead, 2000; Storch & Tapper, 2009).

The operationalisation of syntactic complexity has proved to be a highly complex enterprise (for a discussion of issues, see Bulté & Housen, 2012, 2014) and multiple indices have been proposed in SLA research (Ellis & Yuan, 2004; Larsen-Freeman, 2006; Lu, 2011; Nelson & Van Meter, 2007; Norrby & Hakansson, 2007) to assess the development of

syntactic complexity in second language writing. In SLA studies, the two most frequently used measures to investigate sentence and clausal complexity are the *mean length of T-units* and the *mean number of dependent clauses per T-unit*. These were also applied in our study to assess clausal elaboration and embedding. Both of these measures were retrieved via Synlex.

Table 1. Summary of the lexical measures used in the study

Measure	Definition
Measure of textual lexical diversity (MTLD)	MTLD is a measure of lexical diversity, which is calculated as the mean length of sequential word strings in a text that maintain a given TTR value (McNamara, Graesser, Cai, & Kulikowich, 2011).
Log frequency of content words	The mean of the log frequency of all content words in the text established using the CELEX corpus (Graesser, et al. 2004).
Latent semantic analyses (LSA)	LSA computes how conceptually similar each sentence is to every other sentence in the text (Graesser et al., 2004). It considers meaning overlap between explicit words and also words that are implicitly similar or related in meaning (McNamara & Graesser, in press).

Academic word list A list of 570 frequently occurring words in an academic context (Coxhead, 2000).

Squared verb variation An estimation of lexical diversity as the ratio of the squared number of verb types to the total number of verbs in the text (Lu, 2010).

At the level of phrasal complexity, academic writing is characterised by the use of complex noun phrases and nominalisation (Halliday, 1989). Thus, a further measure of syntactic complexity adopted in this study was the *mean number of modifiers per noun phrase*, computed by Coh-Metrix 2.0. Noun phrase modifiers might appear either before the head noun, i.e., “premodifiers,” or after the head noun, i.e., “postmodifiers” (Biber et al., 1999). *Mean number of complex nominals*, a measure argued to reflect syntactic complexity in academic writing at phrasal level (Biber et al., 2011; Lu, 2010), was also utilised in our research. The computer program Synlex was used to obtain the mean number of complex nominals in subject position per essay in our study.¹

Bulté and Housen’s (2012) definition of syntactic complexity makes reference to the variety of syntactic structures in L2 learners’ knowledge repertoire. The *syntactic structure similarity* index, defined by Crossley and McNamara (2009) as a measure of the consistency of syntactic structures in the text, helps to evaluate syntactic similarity by taking into consideration different parts of speech. In order to be able to account for changes in the variety of syntactic constructions used by our participants, we applied this measure and computed it with the aid of Coh-Metrix 3.0 (see Table 2 for a summary of general measures

of syntactic complexity). It was expected that the syntactic similarity index decreases as syntactic variety increases in students' writing.

Table 2. Summary of the general measures of syntactic complexity

Measure	Definition
Mean length of T-unit	This measure of syntactic complexity is calculated by dividing the total number of words by the total number of T-units. A T-unit is characterised as one main clause plus any subordinate clause or non-clausal structure that it is attached to or is embedded within it (Vajjala & Meurers, 2013).
Mean number of dependent clauses per T-unit	This measure of syntactic complexity is estimated as the ratio of dependent clauses to a T-unit (Wolfe-Quintero, Inagaki, & Kim, 1998).
Mean number of modifiers per noun phrase	This measure of syntactic complexity is calculated as the ratio of modifiers, such as adjectives and prepositional phrases, to noun phrase (Graesser et al., 2004).
Mean number of complex nominals in subject position	This measure of syntactic complexity comprises nouns plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive; nominal clauses; and gerunds and infinitives in the subject position (Cooper,

1976; cited in Lu, 2010).

Syntactic structure similarity This measure of syntactic complexity compares the syntactic tree structures of sentences and identifies the proportion of intersection tree nodes between all adjacent sentences (Graesser et al., 2004).

The global measures used in the study can potentially provide useful information about syntactic changes in students' writing. However, complementing these measures with the analysis of some more specific features of the academic register was also deemed necessary. The genre-specific syntactic constructions were selected based on Biber et al. (1999), who provide a detailed description of clausal and phrasal level structures that are significantly more frequent in academic genres than in conversation, fiction, and news, based on the analysis of the BNC. Our analyses were also guided by Biber et al.'s (2011) recent work, which compared the frequency of a number of grammatical features in conversation and academic texts in the BNC. Following corpus-based studies, our analysis was motivated by the assumption that the development of academic writing abilities of L2 learners would move towards exhibiting the specific syntactic characteristics of the academic genre.

Based on these considerations, one of the syntactic structures we focus on is the frequency of *conditional clauses*. As pointed out by Warchal (2010), conditional clauses can perform a wide range of functions, and they are especially important in academic writing tasks that require logical argumentation and problem solving. As noted by Biber et al. (1999), *prepositional phrases* are the most common type of noun postmodifiers in academic discourse and their frequency in students' essays was also assessed in our research. Prepositional phrases can sometimes be replaced by *relative clauses*. Although this type of

postmodification is not as common as prepositional phrases, it also appears frequently in academic writing (Byrnes & Sinicrope, 2008). Relative clauses are one of the most explicit types of noun modification, and their frequency is often used as one of the indices of syntactic complexity (Jucker, 1992). *Infinitive clauses* represent another type of postmodifiers that are found more often in written than conversational registers. In sum, the following specific indices of syntactic complexity were selected: the ratios of conditional clauses, relative clauses, prepositional phrases, and infinitive clauses as noun postmodifiers to the total number of words, and the ratios of *simple postmodifiers*, i.e., modified by one clause or phrase of any type, and *complex postmodifiers*, i.e., modified by two or more consecutive phrases or clauses, to the total number of words. Following Biber et al. (2011), a normed rate of occurrence for the features of syntactic complexity in each text was counted, and each of the measures was standardised to 1,000 words.

Table 3. Summary of the syntactic measures specific to the academic genre

Measure	Example from student essay
Conditional clauses	<i>If students were dismissed directly</i> , their parents would be really disappointed.
Prepositional phrases as postmodifiers	Serious punishment can be a warning <i>for all students</i> .
Relative clauses	Cheating has become a widespread problem <i>which bothers professors and even degrades school's reputation</i> .
Infinitive clauses as postmodifiers	Universities should give students more opportunities <i>to correct their mistakes</i> .
Simple postmodifiers (one postmodifier per NP)	The advantages <i>of exams</i> cannot be ignored.
Complex postmodifiers (more than one postmodifier per NP)	They have to spend a great amount of time (1) <i>to prepare for them in case of failure</i> (2) <i>in the exams</i> .

Method

Research context

The study was carried out on a pre-sessional EAP/Study Skills programme during the summers of 2012 and 2013. The EAP programme is an intensive four-week course offered by a university in the UK. The major aims of the programme are to develop students' use of English in an academic context, to foster the critical and analytical thinking skills they will need for academic study, and to cultivate an awareness of the learning skills and strategies they might use whilst studying in a British university environment.

The EAP programme is primarily targeted at students with IELTS (International English Language Testing System) scores of 5.5 to 6.5 (B1 to B2 on the Common European Framework of Reference [CEFR], Council of Europe, 2001) and who received only a conditional offer from their university because their current level of English proficiency did not meet the minimum entry requirements. The course is also open for students with higher IELTS scores (i.e., IELTS score of 7, C1 level on the CEFR) who wish to improve their academic writing skills (see below the higher scores of students in Group 1). During the EAP course, students are offered 15 hours of in-class teaching per week and the opportunity to attend one-to-one tutorials. The programme adopts a task-based approach and comprises three modules, which are: (1) academic reading and writing, (2) academic listening, reading, and discussion, and (3) oral presentations. Importantly, academic reading and writing are emphasized as core elements of the programme because these skills are thought to be the most difficult for students to master and yet have the greatest impact on their performance on a degree programme.

There is no summative assessment on the EAP course; therefore, students' performance is evaluated formatively by means of weekly assignments, which take the form

of argumentative essay writing tasks. Students are expected to perform a more detailed analysis of source material, to use more in-depth evaluation and critical thinking, and produce texts which approximate academic writing standards more closely each week. After they complete an assignment, students receive written feedback from their academic reading and writing module tutor and are invited to attend an individual tutorial where the specific strengths and weaknesses of their writing are discussed and suggestions for further improvement are made.

Linguistic improvement is not the primary focus of the EAP curriculum, and students do not receive any explicit language instruction. Nevertheless, linguistic errors, such as recurring grammar, word choice, and spelling errors, are generally highlighted in the feedback provided on students' essays and discussed in the one-to-one tutorials.

Participants

Two groups from two consecutive cohorts of students in the academic years of 2012 and 2013 who enrolled on the 4-week intensive pre-session EAP programme participated in the study. The majority of learners in both groups were females (Group 1: 21 female and 4 male students; Group 2: 12 female and 2 male students) of Chinese L1 background. The students in both groups were planning to study one of the following disciplines upon completion of the EAP course: Business Studies, Economics, Accounting and Finance, or Media and Cultural Studies. Group 1 consisted of 25 postgraduate students, whose ages ranged from 21 to 34, with a mean age of 23.2. Their level of overall language proficiency ranged from IELTS 6 to 7, with an average score of 6.7, and on the IELTS writing component from 5.5 to 6.5 (mean score of 6.3). In terms of the learners' EFL background, they had all studied English at secondary and high school in their home country for 10 to 12 years. The mean length of students' stay in the UK was approximately one month at the time of the study.

Even though they had already completed an undergraduate degree in their home country, all participants acknowledged having had only limited experience of academic writing at university level.

Group 2 consisted of 14 undergraduate students, ranging in age from 18 to 21 years, with a mean age of 19.4. None of the participants in this group had completed an undergraduate degree prior to applying for undergraduate studies in the UK. The English language proficiency of this group was slightly lower than that of the previous group in terms of both their general IELTS scores and specific writing scores in the IELTS exam. The overall IELTS scores and writing scores of Group 2 ranged between 5.5 and 6.5 (mean score of 5.9), and between 5.5 and 6 (mean score of 5.8), respectively. Similar to Group 1, the students in Group 2 had all studied English at school in their home country and had no prior experience of living in any English-speaking country at the time of the research. According to the language proficiency test results, the participants of Group 1 and Group 2 could be defined as “proficient users of the language” (C1 level on the CEFR scale) and “independent users of the language” (B2 level on the CEFR scale), respectively. All 39 students took part in the research voluntarily and were each awarded a £10 Amazon gift voucher in return for their participation. Table 4 summarises the background data of the participants of both groups.

Table 4. Learner profiles

		Group 1	Group 2
Gender	Male	4	2
	Female	21	12
Age	Mean	23.2	19.4
	Range	21-34 years	18-21 years
L1 background	Chinese	17	14

	Japanese	3	0
	Thai	5	0
L2 learning experience	Length of learning English	11 years	10 years
	Length of staying in the UK prior to the EAP programme	2 weeks	2 weeks
English language proficiency	Mean IELTS listening	6.4	6.3
	Mean IELTS reading	6.8	6.2
	Mean IELTS speaking	6.3	5.9
	Mean IELTS writing**	6.3	5.8
	Mean IELTS overall*	6.7	5.9

**Statistically significant difference between the two groups: $t(39) = 2.7645$, $p = 0.00885$

* Difference between the two groups approaching significance: $t(39) = 1.3598$, $p = 0.06518$

Instruments

Each participant in both Groups 1 and 2 was asked to complete two argumentative writing tasks as part of the study, one at the very beginning, week one, and the other in the final week, week four, of the EAP/ Study Skills programme. Both writing sessions were conducted in a computer lab, where students were required to write an essay of between 300 and 400 words using word-processing software. In order to control for topic difficulty, the essay prompts were selected from the general field of education, which was assumed to be relevant and familiar to all participants. The task prompts used in the study were as follows:

Topic A: Exams cause unnecessary stress for students. *How far do you agree?*

Topic B: Any student caught cheating in school or college exams should be automatically dismissed. *How far do you agree?*

The order of tasks was counterbalanced, so that half the students completed the task on topic A in the first session, and on topic B in the second session. The other half of the participants started with topic B and wrote about topic A at the end of the study. To check for significant differences due to the effect of the topic, a Mann Whitney U test was applied to the data set. No significant differences were found between the groups on any of the linguistic variables analysed in our study (see Tables 1-3 above) with regard to the topic they wrote about.

Data collection procedures

Data collection took place over a period of four weeks, that is, from the beginning to the end of the EAP programme. Although this period might seem relatively short to observe linguistic development, the intensity of the programme is very high as it provides 60 hours of instruction, which is commensurate with a semester-long (15 weeks) course that offers 4 hours of instruction per week. Two writing sessions were set outside the regular class hours of the EAP course. Each participant was given a prompt and asked to complete the task by typing an essay in no more than 45 minutes. The students were instructed to work individually, and the use of a dictionary or any other reference materials was prohibited in order to judge the participants' current level of linguistic development without the use of external resources. The tasks used in the study had been previously piloted on a similar population and proved to be manageable within the allocated time.

Data analysis

Several software packages were used to analyse the lexical diversity and syntactic complexity of texts. These packages were Coh-Metrix 2.0 and Coh-Metrix 3.0, Synlex L2 Syntactic

Complexity Analyzer and Synlex Lexical Complexity Analyzer, and Vocabprofiler BNC. In order to avoid misinterpretation of the data by those computer-assisted tools, all essays were corrected by one of the researchers for misspellings and erroneous punctuation so that the computational programs could detect and identify the words. The syntactic structures specific to the academic genre chosen for the analyses were identified and coded manually. The coding was initially done by one of the authors, and following this a quarter of the data set was coded by a second native speaker with a PhD in Applied Linguistics. The inter-rater reliability for the coding of genre-specific syntactic structures (Cohen's kappa) was 0.75, which according to Landis and Koch (1977) signifies "excellent agreement."

Statistical analyses were carried out using SPSS (Statistical Package for Social Sciences) version 16.0 for Windows. As the analysed variables were not normally distributed, nonparametric tests were used for statistical inference. The statistical test applied to examine differences from Time 1 to Time 2 was the Wilcoxon signed-rank test, a non-parametric equivalent to the paired sample t-test. Effect size was calculated and absolute effect sizes of 0.1 to 0.29 were taken as indicating a small effect, from 0.3 to 0.49 a medium effect, and greater than 0.5 a large effect (Cohen, 1969).

Results

This section gives an overview of the findings in light of the research question that guided the study. The descriptive statistics (mean and standard deviation) for all five lexical diversity measures of the lower and higher level proficiency groups are displayed in Table 5. The descriptive statistics reveal that lexical diversity increased from Time 1 to Time 2 for four measures (with the exception of the log frequency of content words) in both groups. However, the results of the Wilcoxon signed-rank tests show statistically significant differences only for two measures of Group 1 (higher proficiency level): 1) *squared verb*

variation ($Z = -3.123, p < 0.002, r = -0.44$); and 2) *academic word list* ($Z = -2.222, p < 0.026, r = -0.31$). The effect sizes were medium for both measures. Conversely, in Group 2 (lower proficiency level), statistically significant differences were found for all five measures of lexical diversity: 1) MTLTD ($Z = -3.296, p < 0.001, r = -0.62$); 2) squared verb variation ($Z = -2.731, p < 0.006, r = -0.52$); 3) log frequency of content words ($Z = 2.166, p < 0.03, r = 0.41$); 4) *academic word list* ($Z = -2.104, p < 0.035, r = -0.4$); 5) latent semantic analyses ($Z = -2.04, p < 0.041, r = -0.39$). The significant differences all suggest improvement in lexical diversity including the log frequency of content words, for which the decrease in mean values indicates the use of less frequent words. The effect sizes can be identified as large for the first two measures, i.e., MTLTD and squared verb variation, and medium for the remaining measures of lexical diversity (see Table 6).

Table 5. Descriptive statistics for the measures of lexical diversity

Measures	Group 1				Group 2			
	Time 1		Time 2		Time 1		Time 2	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
MTLTD	73.55	19.78	86.21	17.75	72.36	11.15	87.66	12.2
Squared verb variation	20.38	5.59	25.22	4.59	18.25	3.12	22.98	5.92
Log frequency of content words	2.382	0.1	2.38	0.09	2.42	0.09	2.35	0.08
Academic word list	5.20	2.12	6.82	2.29	3.61	1.55	5.78	2.36
Latent semantic analyses	0.28	0.09	0.28	0.09	0.27	0.09	0.33	0.08

Table 6. Results of Wilcoxon signed-rank tests for the changes in lexical diversity

Measures	Group 1	Group 2
----------	---------	---------

	<i>Z</i>	<i>p</i>	<i>r</i>	<i>Z</i>	<i>p</i>	<i>r</i>
MTLD	-0.794	0.427	-0.11	-3.296	0.001**	-0.62
Squared verb variation	-3.123	0.002**	-0.44	-2.731	0.006**	-0.52
Log frequency of content words	-0.186	0.853	-0.03	2.166	0.03*	0.41
Academic word list	-2.222	0.026*	-0.31	-2.104	0.035*	-0.4
Latent semantic analyses	-0.214	0.83	-0.03	-2.04	0.041*	-0.39

** <.01, * <.05

The descriptive statistics for the general syntactic measures and the results of Wilcoxon signed-rank tests are displayed in Tables 7 and 8, respectively. The data demonstrate that, for Group 1, the use of these syntactic structures remained unchanged from the beginning to the end of the study, with the exception of one measure, *syntactic structure similarity* between adjacent sentences for which the change between Time 1 and Time 2 was significant and had a moderate effect size ($Z = -3.287, p < 0.001, r = -0.46$). In contrast, the students in Group 2 significantly increased their usage of three out of five general measures of syntactic complexity: 1) *complex nominals* ($Z = -2.045, p < 0.041, r = -0.39$); 2) *modifiers per noun phrase* ($Z = -2.542, p < 0.011, r = -0.48$); and 3) *syntactic structure similarity* ($Z = -2.355, p < 0.019, r = -0.45$). The effect sizes for all three measures were in the medium range.

Table 7. Descriptive statistics for global measures of syntactic complexity

Measures	Group 1				Group 2			
	Time 1		Time 2		Time 1		Time 2	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Mean length of T-unit	17.04	2.6	16.44	1.8	16.26	3.06	16.93	4.05
Dependent clauses per	0.77	0.21	0.74	0.3	0.72	0.29	0.84	0.25

T-unit								
Modifiers per noun phrase	0.68	0.13	0.72	0.12	0.63	0.13	0.74	0.15
Complex nominals	36.88	8.03	38.38	8.11	34.57	15.3	40.14	10.65
Syntactic structure similarity	0.08	0.02	0.12	0.02	0.08	0.03	0.10	0.04

Table 8. Results of Wilcoxon signed-rank tests for the changes in global measures of syntactic complexity

Measures	Group 1			Group 2		
	<i>Z</i>	<i>p</i>	<i>r</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Mean length of T-unit	-0.578	0.563	-0.08	-0.471	0.638	-0.09
Dependent clauses per T-unit	-0.336	0.737	-0.05	-1.319	0.187	-0.25
Modifiers per noun phrase	-1.257	0.209	-0.18	-2.542	0.011*	-0.48
Complex nominals	-0.822	0.411	-0.12	-2.045	0.041*	-0.39
Syntactic structure similarity	-3.287	0.001**	-0.46	-2.355	0.019**	-0.45

** <.01* <.05

As is clear from Tables 9 and 10, with respect to genre-specific syntactic measures, the results for Group 1 and Group 2 differ considerably. The only difference for Group 1 was found in the frequency of infinitive clauses. The students in the higher proficiency group used significantly fewer infinitive clauses in their writing at Time 2 than they did at Time 1 ($Z = -2.112$, $p < 0.035$, $r = -0.3$). As far as the performance of Group 2 is concerned, statistically significant differences were found for three syntactic measures, i.e., conditional clauses,

relative clauses, and complex postmodifiers. Specifically, the ratios of these syntactic structures increased from Time 1 to Time 2: 1) *conditional clauses* ($Z = -2.312, p < 0.021, r = -0.44$); 2) *relative clauses* ($Z = -1.978; p < 0.048, r = -0.37$); and 3) *complex postmodifiers* ($Z = -1.977; p < 0.048; r = -0.37$). The effect sizes for all these differences across time were moderate.

Table 9. Descriptive statistics for measures of syntactic complexity specific to the academic genre (frequency per 1,000 words)

Measures	Group 1				Group 2			
	Time 1		Time 2		Time 1		Time 2	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Conditional clauses	4.3	3.27	5.74	3.85	2.8	3.93	6.16	5.42
Prepositional phrases	49.08	15.92	46.51	15.75	48.39	17.51	43.64	17.39
Relative clauses	13.81	8.9	11.05	6.8	9.44	5.44	14.99	5.93
Infinitive clauses	14.12	7.26	10.88	7	15	7.6	12.33	6.19
Simple postmodifiers	35.31	11.92	34.27	11.23	33.22	8.18	30.84	12.32
Complex postmodifiers	18.37	7.64	15.27	6.6	13.72	8.26	18.05	6.56

Table 10. Results of Wilcoxon signed-ranks tests for the changes in measures of syntactic complexity specific to the academic genre

Measures	Group 1			Group 2		
	<i>Z</i>	<i>p</i>	<i>r</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Conditional clauses	-1.686	0.092	-0.24	-2.312	0.021*	-0.44

Prepositional phrases	-0.713	0.476	-0.1	-0.596	0.551	-0.11
Relative clauses	-0.996	0.319	-0.14	-1.978	0.048*	-0.37
Infinitive clauses	-2.112	0.035*	-0.3	-1.287	0.198	-0.24
Simple postmodifiers	-0.121	0.904	-0.02	-0.22	0.826	-0.04
Complex postmodifiers	-1.655	0.098	-0.23	-1.977	0.048*	-0.37

* <.05

To illustrate, only one instance of conditional clause use has been found in the essay produced by Student 1 at Time 1: (1) *“If they change their mind and find the real aim of exams, they would do better not only in next exams, but also in their future study.”* In contrast, three instances of this syntactic construction usage have been detected at Time 2 in the essay written by the same student: (1) *“If students get cheating in exams as a habit, that must be dangerous.”* (2) *“If the number keeps increasing, there is no need to have exams anymore.”* (3) *“If every student cheats in exams, schools can cancel exams for exams cannot achieve their aim.”*

With regard to differences observed in the usage of relative clauses from Time 1 to Time 2, Student 5 produced only two of these syntactic structures in their essay at Time 1: (1) *“They may do not enough time for their sleep and spare time to relax, which may lead students feel stressful.”* (2) *“On the contrary, exams are important and effective method which can reflect students' achievement of study, they directly can know where are wrong or right.”* It needs to be noted that many of these relative clauses were not adjectival clauses but sentential relative clauses, i.e., comments on a whole previous sentence or series of clauses, and they did not postmodify noun heads. Conversely, in the essay produced by the same

student at Time 2, relative clause constructions were used five times: “(1) Many schools adopt the serious measure to deal with some students *who cheat in school or college exams*, for example these students will be dismissed automatically.” (2) “Some of them may be afraid to go to school again and join other activities, *which could cause these students become timid and dissocial*.” (3) “Many researches and studies prove that someone has shadow *mentally who might have a unhappy future*, and they may do some wrong things in the future.” (4) “In some American schools, these students will be recommended to have a class that is about psychology, *which can help students to avoid cheating again*.” (5) “To sum up, students' plagiarizing is incorrect behaviour, *which should be avoided*.”

To illustrate the difference in students' use of complex postmodifiers, the two essays written by Student 7 are used. At Time 1, only one instance of complex modifiers was observed: “Everyone have different ability (1) *to face the stress* (2) *in different situations*, (3) *which happen at college or university*.”

However, at Time 2, Student 7 used four constructions of that kind in their writing:

(1) “It is more basic than goodness and evil because someone pretends to be great with dishonesty, so as cheating 1) *in the exams* 2) *to get better results*.”

(2) “Nowadays, the exam has become one of the most important ways 1) *to test students' ability* 2) *on learning* 3) *in school or college* not only the Asian countries but also other foreign areas.”

(3) “In his final result (1) *of academic performance* (2) *at the university*, the student just scored D.”

(4) “I will first focus on the two different attitudes 1) *to the punishment* 2) *about cheating* 3) *whether students* who cheated in the exams should be dismissed without any other consideration.”

Discussion and Conclusion

Our results indicate four major trends with regard to lexical and syntactic development in the learner groups investigated. First, the lower level proficiency group showed improvements in considerably more areas than did the higher proficiency group. Second, changes in the linguistic features of students’ writing in both groups could be observed in the area of lexis. Third, the lower proficiency group demonstrated changes in their use of complex noun phrases and conditional clauses in the area of syntax. Finally, despite an expectation that with the development of writing skills students would use more varied syntactic structures, syntactic structure similarity increased, indicating in fact that the students applied a smaller variety of syntactic constructions in their academic writing. We will discuss each of these findings in turn below.

It is not surprising to find that students whose linguistic competence is at a relatively advanced level, as attested to by their IELTS scores, and who have already gained some literacy experience in their L1 during their university studies, make little linguistic progress in an EAP programme that does not explicitly focus on areas of syntax and vocabulary. Shaw and Liu’s results (1998) also indicate only a minor development in the syntactic complexity of students’ essays in an EAP setting similar to ours. Nevertheless, the changes that can be observed in the syntactic features of the higher level students’ writing both in terms of global syntactic complexity and specific to the academic genre are mostly moving towards the syntactic characteristics of academic writing. Specifically, the mean length of T-units and

clausal embedding can be observed to decrease and modifiers per noun phrase, complex nominal and syntactic structure similarity increase. Interestingly, complex post-modification and the frequency of relative clauses and prepositional phrases drop slightly although insignificantly in this group. Nevertheless, when compared with the corpus data in Biber et al.'s (2011) study, the frequency of these features can still be considered as approaching the frequency values observed in native speakers' writing. The most important syntactic change that took place in the writing of the higher proficiency group was a reduction in the use of infinitive clauses. Taken together with other syntactic changes in terms of the reduction in clausal complexity and in the increase of frequency of words in the academic word list, this might indicate that these students move in the direction of relying more on nominalisation in their writing than on pre- and post-modification. This can be illustrated by the case of Student 1 in whose essay at the start of the EAP programme one can find six instances of nominalisation, out of which on four occasions the student repeats the word *dismissal*. In the post-test the same student uses nominalisation nine times (e.g., in sentences such as "Although it requires *memorization* of some ideas and knowledges, in fact it also requires students to be able to truly understand and adapt it to their everyday 's *usage*.").

The lower proficiency group was found to have made significant improvements in all measures of lexical diversity and showed the largest gains on measures assessing lexical variability. Lexical variability was also found to be one of the most sensitive indicators of L2 proficiency in a study by Crossley et al. (2011). An increase was also detected in the variation of verbs used for both the lower and higher level groups. For the lower-level group the effect size was large, while for the higher level group it was considered medium. Verb variation was found to differentiate L1 and L2 writers of French in a study by Harley and King (1989) and was also shown to be a useful predictor of proficiency-related differences in oral production (Lu, 2012). The increase in variability of the words used can be explained by

the interplay of a number of factors. On the one hand, during the one-month period of the study, the learners' receptive and productive vocabulary size could have grown due to the incidental learning of new words in the target language environment and during the EAP course. This assumption could be supported by considering the increase in frequency of words from the academic word list in the students' writing. On the other hand, it is also possible that the student writers were paying conscious attention to varied lexical choice and the use of formal and academic vocabulary in their essays after completing the EAP course.

Our study also shows syntactic changes in the writing of the lower proficiency group. Measures of clausal embedding, which are often assumed to be representative of syntactic complexity in writing (see e.g., Brown & Yule, 1983; Bulté & Housen, 2012, 2014; Hyland, 2002), were not found to change significantly during the EAP programme. In the case of higher level learners, the trend was for these measures to decrease somewhat. In contrast, noun-phrase complexity increased significantly in the lower proficiency group in terms of the frequency of complex nominals and noun-phrase modifiers, the use of relative clauses as postmodifiers and the frequency of complex postmodifiers overall. This indicates that these learners have reached a stage of development where their noun phrase constructions become embedded and elaborate (cf. Biber et al., 2011; Crossley & McNamara, 2014). As we pointed out earlier, in the higher proficiency group, noun-phrase complexity did not continue to rise; on the contrary, a negative tendency was observed. This pattern of change illustrates that it is probably at this point that students started using syntactically less complex but conceptually more abstract lexical units to express their views and opinions (Byrnes, Maxim, & Norris, 2010).

Finally, a particularly interesting finding in our study was that syntactic structure similarity had significantly increased in both groups by the end of the EAP programme. This finding is in contrast with that of Crossley and McNamara (2014) who found a decrease in

syntactic structure similarity in the MSU corpus. A comparison of the mean values of the similarity scores in the two studies reveals that the essays in the MSU dataset had higher similarity indices both at the beginning and at the end of the academic writing programme than the texts in our study. This might be due to the fact that the learners represented in the MSU corpus might have had lower proficiency than our participants and that their texts were descriptive in nature, whereas ours were argumentative. We can hypothesize that the development in the variety of syntactic constructions in learners' writing might not be linear. Initially students move from the use of similar constructions towards variety, but beyond a point variety impacts the readability of texts. Crossley, Greenfield, and McNamara (2008) argue that similar syntactic constructions "provide important links between sentences" (p. 489). Therefore, it can be hypothesized that the participants in our study might have used similar syntactic constructions, adjacent to each other, to increase the cohesion of their writing by grammatical means.

The findings with regard to lexical and syntactic development in our study have important implications for the field of second language writing, SLA, and EAP teaching. The fact that the less proficient group made substantial improvements in lexical diversity and sophistication indicates that lexical development in these areas can take place without explicit vocabulary instruction on an EAP programme. Similarly, the syntactic features of students' writing that are typical characteristics of academic genres were also found to develop, albeit modestly, in the lower-proficiency group. In line with recent studies in the EAP context represented in the MSU corpus (Bulté & Housen, 2014; Crossley & McNamara, 2014; Friginal & Weigle, 2014), this suggests that development in the syntactic domain of L2 academic writing is possible in the absence of explicit language instruction. Nevertheless, it is important to note that the period of study was relatively short compared with other studies investigating writing development on an EAP programme (e.g., Shaw & Liu, 1998; Storch &

Tapper, 2009). Hence, our research does not provide evidence for the development of underlying knowledge representations of complex syntactic structures or features of writing characteristic of academic genres, especially because we did not analyse the accuracy with which the students used these constructions. What our study does show is that students' "repertoire of choices" (Ortega & Byrnes, 2008, p. 287), of specific syntactic structures and lexical features, moved in the direction of "idealized writing profiles" (Byrnes, Maxim, & Norris, 2010, p. 91) as expected for university-level academic writing assignments.

The intensive nature of the EAP course, exposure to a variety of academic reading texts, individualised feedback on the overall quality of writing assignments, and immersion in the target language environment could all have contributed to increases in the lexical complexity of students' essays and to the changes observed in the use of syntactic features. Our results suggest that these instructional and environmental conditions can be conducive to lexical and syntactic development and should, if possible, be included in pre-sessional EAP programmes. It might also prove useful to call learners' attention explicitly to the lexical and syntactic features of academic writing, both before they embark on a writing assignment, for example, in the academic reading tasks they are set, and when they revise their essays (see Storch & Tapper, 2009, for similar suggestions). Model essays, checklists, and self-assessment tools that include descriptions of lexical and syntactic characteristic features of academic writing might also assist learners in various phases of the writing, editing, and revision process (Vyatkina, 2013).

Our study has a number of limitations and these should be followed up by further research. First, as the numbers of participants in both groups were relatively low, individual differences might have masked some patterns that could have emerged had the sample size been larger. Thus it might be useful in the future to replicate the research so that the results might be confirmed by a study with a larger sample size. A more detailed analysis of the

changes in nominalisation and noun-phrase complexity in the students' essays would also be informative. Another possible future research direction is to investigate the development of students' writing by means of a series of individual case studies, which would help to identify specific particular factors linked to L2 writing development on EAP programmes (Norris & Manchón, 2012). In terms of research methods, the application of multiple data gathering and data analysis techniques might be worthy of inclusion in further studies. Specifically, the integration of retrospective think-aloud protocols, in-depth semi-structured interviews, and self-reflection journals (Bosher, 1998; Cumming, 2006; Manchón, 2011) would undoubtedly lead to a better understanding of learners' composing behaviours. The interrelation between learners' perceived difficulties, the strategies they use, and the syntactic, lexical, and cohesive features found in their writing could also be explored. This would offer insights into the factors that appear to affect learners' perception and assessment of their own difficulties in L2 writing, both in general and when producing academic texts in particular.

Notes

¹Complex nominals can also occur in non-subject positions, but because we used a computerised tool that could only detect complex nominals in the subject position these were not analysed in this study.

References

- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (Eds.) (1993). *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium.
- Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes, 43*, 79-120.
DOI: 10.1080/01638530709336894.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *English for Academic Purposes, 9*, 2-20.
DOI: 10.1016/j.jeap.2010.01.001.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*, 5-35. DOI:10.5054/tq.2011.244483.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
- Bosher, S. (1998). The composing processes of three Southeast Asian writers at the post-secondary level: An exploratory study. *Journal of Second Language Writing, 7*, 205-241. DOI: 10.1016/S1060-3743(98)90013-3.
- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. House, F. Kuiken, & I. Vedder (Eds.). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42-65.
DOI:10.1016/j.jslw.2014.09.005.

- Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education, 20*, 50–66.
DOI: 10.1016/j.linged.2009.01.005.
- Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *The Modern Language Journal, 94*(s1), i-ii.
- Byrnes, H., & Sinicropo, C. (2008). Advancedness and the development of relativization in L2 German: A curriculum-based longitudinal study. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 109–138). New York: Routledge/Taylor & Francis.
- Cobb, T. (1994). *Web Vocabprofile*. Retrieved from <http://www.lexutor.ca/vp/>
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Connor-Linton, J., & Polio, C. (2014). Comparing perspectives on L2 writing: Multiple analyses of a common corpus. *Journal of Second Language Writing, 26*, 1-9. DOI: 10.1016/j.jslw.2014.09.002.
- Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213–238.
DOI: 10.2307/3587951.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly, 42*, 475-493.
DOI: 10.1002/j.1545-7249.2008.tb00142.x
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences

- in L1 and L2 writing. *Journal of Second Language Writing*, 18, 119–135.
- DOI: 10.1016/j.jslw.2009.02.002.
- Crossley, S. A., & McNamara, D. S. (2013). Applications of text analysis tools for spoken response grading. *Language Learning and Technology*, 17, 171-192.
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66-79. DOI: 10.1016/j.jslw.2014.09.006.
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, 45, 182-193. DOI: 10.5054/tq.2010.244019.
- Cumming, A. H. (2006). *Goals for academic writing: ESL students and their instructors*. Amsterdam: John Benjamins.
- Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behaviour Research Methods*, 37, 65-70. DOI: 10.3758/BF03206399.
- Edwards, R., & Collins, L. (2013). Modelling L2 vocabulary learning. In S. Jarvis and M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 157-184). Amsterdam: John Benjamins.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59–84.
- Engber, C.A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139-155. DOI: 10.1016/1060-3743(95)90004-7.

- Fang, Z., Schleppegrell, M. J., & Cox, B. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research, 38*, 247–273. DOI:10.1207/s15548430jlr3803_1.
- Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. *Journal of Second Language Writing, 26*, 80-95. DOI: 10.1016/j.jslw.2014.09.007.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*, 223-234. DOI: <http://dx.doi.org/10.3102/0013189X11413260>.
- Graesser, A. C., McNamara, D. S., Louwrese, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*, 193–202. DOI: 10.3758/BF03195564.
- Granger, S., & Paquot, M. (2009). Lexical verbs in academic discourse: A corpus-driven study of learner use. In C. Maggie, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 193-214). London: Continuum.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9*, 123–145. DOI: 10.1016/S1060-3743(00)00019-9.
- Green, A., & Weir, C. (2002). *Monitoring score gain on the IELTS academic writing module in EAP programmes of varying duration. Phase 1 report*. Cambridge: UCLES.
- Halliday, M. A. K. (1989). *Spoken and written language*. Oxford: Oxford University Press.
- Halliday, M. A. K., & Martin, J. R. (1993/1996). *Writing science: Literacy and discursive power*. London: Falmer Press.

- Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition*, 11, 415-440.
DOI: 10.1017/S0272263100008421.
- Heatley, A., & Nation, P. (1994). *Range* [Computer software]. Victoria University of Wellington, NZ. Retrieved from <http://www.vuw.ac.nz/lals/>
- Hirvela, A. (2011). Writing to learn in content areas: Research insights. In R. M. Manchón (Ed.). *Learning-to-write and writing-to-learn in an additional language* (pp. 37-59). Amsterdam: John Benjamins.
- Hyland, K. (2002). *Teaching and researching writing*. London: Longman.
- Hyland, K. (2011). Learning to write: Issues in theory, research, and pedagogy. In R. M. Manchón (Ed.) *Learning-to-write and writing-to-learn in an additional language* (pp.17-36). Amsterdam: John Benjamins.
- Jarvis, S. (2002). Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing*, 19, 57-84. DOI:10.1191/0265532202lt220oa.
- Jarvis, S. (2013). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 13-45). Amsterdam: John Benjamins.
- Jucker, A. H. (1992). *Social stylistics: Syntactic variation in British newspapers*. New York: Mouton de Gruyter.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27, 590–619. DOI:10.1093/applin/aml029.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written

- production, *Applied Linguistics*, 16, 307-322.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15, 474-496. DOI: <http://dx.doi.org/10.1075/ijcl.15.4.02lu>.
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96, 190-208. DOI:10.1111/j.1540-4781.2011.01232.x.
- Lu, X. & Ai, H. (2011). Synlex: Lexical complexity analyzer. L2 syntactical complexity analyzer. [Computer software]. Retrieved from <http://aihaiyang.com/software/>
- Manchón, R. M. (2011). *Learning-to-write and writing-to-learn in an additional language*. Amsterdam: John Benjamins.
- McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Unpublished doctoral dissertation). University of Memphis. Retrieved from <https://umdrive.memphis.edu/pmmccrth/public/Phil's%20papers.htm?uniq¼-xq6brv>
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381-392. DOI: 10.3758/BRM.42.2.381.
- McNamara, D. S., & Graesser, A. C. (in press). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.

- McNamara, D. S., Graesser, A. C., Cai, Z., & Kulikowich, J. M. (2011, April). *Coh-Matrix easability components: Aligning text difficulty with theories of text comprehension*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading and Writing Quarterly*, 23, 287–309. DOI:10.1080/10573560701277807.
- Nippold, M. A. (2004). Research on later language development: International perspectives. In R. Berman (Ed.), *Language development across childhood and adolescence* (pp. 1-8). Amsterdam: Benjamins.
- Norrby, C., & Håkansson, G. (2007). The interaction of complexity and grammatical processability: The case of Swedish as a foreign language. *International Review of Applied Linguistics*, 45, 45–68. DOI:10.1515/IRAL.2007.002.
- Norris, J. M., & Manchón, R. M. (2012). Investigating L2 writing development from multiple perspectives: Issues in theory and research. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 221-244). New York: de Gruyter Mouton.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555-578. DOI:10.1093/applin/amp044.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24, 492–518. DOI:10.1093/applin/24.4.492.
- Ortega, L., & Byrnes, H. (2008). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 3–20). New York: Routledge/Taylor & Francis.

- Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7, 43–68. DOI: 10.1016/S1060-3743(98)90005-4.
- Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, 26, 10-27. DOI: 10.1016/j.jslw.2014.09.003.
- Reppen, R. (1994). *Variation in elementary student language: A multi-dimensional perspective* (Unpublished doctoral dissertation). Northern Arizona University, Flagstaff.
- Shaw, P., & Liu, E. T. K. (1998). What develops in the development of second language writing? *Applied Linguistics*, 19, 225-254.
- Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes*, 8, 207-223. DOI:10.1016/j.jeap.2009.03.001.
- Templin, M. C. (1957). *Certain language skills in children: Their development and interrelationships*. Westport, CT: Greenwood.
- Vajjala, S., & Meurers, D. (2013). On the applicability of readability models to web texts. In *Proceedings of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (pp. 59-68). Sofia, Bulgaria: Association for Computational Linguistics.
- Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92, 214–231. DOI: 10.1111/j.1540-4781.2008.00715.x.

- Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*, 96, 576–598. DOI: 10.1111/j.1540-4781.2012.01401.x.
- Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, 97, 11–30. DOI: 10.1111/j.1540-4781.2012.01421.x.
- Warchal, K. (2010). Moulding interpersonal relations through conditional clauses: Consensus-building strategies in written academic discourse. *Journal of English for Academic Purposes*, 9, 140-150. DOI: 10.1016/j.jeap.2010.02.002.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii Press.