



## Blog posts and traditional assignments by first- and second-language writers

*Irina Elgort, Victoria University of Wellington*

### Abstract

*This study investigates differences in the language and discourse characteristics of course blogs and traditional academic submissions produced in English by native (L1) and advanced second language (L2) writers. One hundred and fifty two texts generated by 38 graduate students within the context of the same Master's level course were analysed using Coh-Metrix indices at the surface code, textbase and situation model levels. The two text types differed in their lexical sophistication, syntactic complexity, use of cohesion and agency. Overall, the traditional course assignments were more formal, lexically sophisticated and syntactically complex, while the blog posts contained more semantic and situational redundancy, resulting in higher readability, and communicated a clearer sense of agency. There were also reliable differences between the textual artefacts generated by the L1 and L2 writers, one of which was a more traditional impersonal academic style of the L2 texts. Although no interaction was observed between the two independent variables in the Coh-Metrix analyses, an additional analysis of human ratings showed that the blog posts were rated lower on the use of language than traditional assignments for the L2, but not L1, writers. Limitations of the computational text analysis and pedagogical implications of the findings are considered.*

**Keywords:** Academic Writing, Digital Texts, Discourse Analysis, Corpus Analysis

**Language(s) Learned in this Study:** English

**APA Citation:** Elgort, I. (2017). Blog posts and traditional assignments by first- and second-language writers. *Language Learning & Technology*, 21(2), 52–72. Retrieved from <http://llt.msu.edu/issues/june2017/elgort.pdf>

### Introduction

Educational blogging is increasingly used in academia as a form of writing that facilitates student engagement with concepts and ideas, encourages deeper learning through reflection and peer feedback, and creates a clearer sense of audience (Bloch, 2007; Gallagher, 2010; Halic, Lee, Paulus, & Spence, 2010; Jimoyiannis & Angelaina, 2012; Murray, Hourigan, & Jeanneau, 2007). Blog, a social software tool that allows its author to easily post online individual text (or multimedia) entries displayed in reverse chronological order, was originally implemented and used outside of the education sector. Educational blogging arose as a result of the affordances of the tool to create new forms of audience, authorship and artefact (Warschauer & Grimes, 2007).

Warschauer and Grimes (2007) emphasise agency in blogging, suggesting that successful blogs “tend to have a strong authorial voice” (p. 8). Blogs usually allow interactivity between the author and the audience, but these interactions are asymmetrical, with the blog’s author having control over topics, their interpretations, and whether or not to engage with the audience. Reader contributions to blogs tend to be limited to comments on author’s posts. Some blogs are written with a relatively small well-defined readership in mind, but in the blogosphere (i.e., blogs and their connections), a blog’s audience is essentially unlimited, “composed of anyone with a Web browser who could accidentally or intentionally come across a blog posting” (p. 8). Educational (course) blogs usually have a wider audience than traditional course submissions—that is, they are read by the students enrolled in the course as well as the teacher. The teacher

may also choose whether to keep course blogs restricted to those taking the course or to situate them in the broader blogosphere.

Studies into first language (L1) and second language (L2) educational blogging tend to report positive effects of blogging on student motivation and willingness to write and, in some cases, on the quality of writing, especially when accompanied by peer feedback (Arslan & Sahin-Kizil, 2010; Aydin, 2014; Ciftci & Kocoglu, 2012; Lee, 2010; Sun, 2010; Sun & Chang, 2012; Vurdien, 2013). Some results, however, are less straightforward. Xie, Ke, and Sharma (2008) reported a positive effect of regular course blogging by L1 undergraduate students on the development of reflective thinking, and a positive correlation between students' reflective thinking and their course grades. Interestingly, they also found that students who sent and received peer feedback on their posts tended to be less reflective than those who did not, and that peer comments tended to have a social function of communicating general support (e.g., "good job," "I agree"), rather than engaging with ideas expressed in the posts. In another L1 undergraduate study, Lester and Paulus (2011) found that students were reluctant to take ownership of their academic views in their course blogs, resisting the position of a knowledgeable student in favour of an academic identity of a novice. Blog posts created by some students adhered to the conventions of traditional objective academic discourse, avoiding personal experience and beliefs. Lester and Paulus argue that this behaviour may have been caused by the combination of students having to publically display their knowledge (while not being sure of its quality) and their awareness of the wider readership of their blog posts, which they could not control (p. 680).

Still, little is known at this stage about how written artefacts produced in academic blogging differ from those produced when students engage in traditional, essay-style academic writing. In a recent study with undergraduate-level L1 sociology students, Foster (2015) used comparative content analysis to investigate student writing in private journals and course blogs (i.e., low-stakes writing assignments), focusing on reflection depth and engagement with course materials. The main finding of the study highlights an important difference in the kind of content generated by the students in these two writing modes: students took more intellectual risks when blogging (e.g., developing a personal theory) and more personal risks when writing in private journals (e.g., reflecting on a personal misconception). This and similar studies (e.g., Gallagher, 2010; Jimoyiannis & Angelaina, 2012; Sun & Chang, 2012) show pragmatic effects of the writing mode but, to my knowledge, there are no studies that compare textual (micro) features of student writing in blogs and academic assignments, for L1 and L2 authors (Hinkel, 2011). The present study therefore examines language and text features of course blogs and traditional course submissions by L1 and L2 students in the same university course. The study also investigates whether previously observed differences between L1 and L2 academic writing (Crossley & McNamara, 2011a, 2011b; Friginal, Li, & Weigle, 2014) are more prominent in essay or blog texts.

Characteristics of the texts produced by L1 and L2 writers in blog posts and traditional course submissions in this study were obtained using language and discourse indices generated using a computation tool, *Coh-Metrix* (Graesser, McNamara, Louwerse, & Cai, 2004). Graesser and McNamara (2011) suggest that the Coh-Metrix tool can be used "to acquire a deeper understanding of language-discourse constraints" (p. 372). Coh-Metrix provides indices at the level of the text, paragraph, sentence, and word and can be used to analyse L2 writing (e.g., Crossley & McNamara, 2012) as well as L1 writing (e.g., Graesser et al., 2014; McNamara, Crossley, & McCarthy, 2010). The Coh-Metrix tool generates indices corresponding to the first three levels of the five-level taxonomy of discourse representation (Graesser, Millis, & Zwaan, 1997) that includes the surface code, explicit textbase, situation model, discourse genre or rhetorical structure, and pragmatic communication levels (Graesser & McNamara, 2011). The surface code level indices are related to the language per se, such as words and syntax. The textbase level indices are related to explicit expression of meaning and ideas, such as explicit propositions and links between them, as well as inferences that connect these propositions. The situation model refers to the subject matter of discourse and reflects causation, intentionality, time, space, and protagonists (McNamara, Graesser, McCarthy & Cai, 2014). Since its introduction, Coh-Metrix has been used in studies investigating the cohesion and linguistic sophistication of texts, comparing different types of discourse (e.g., spoken and written discourse, narrative and informational or expository texts, textbooks and research publications, integrated and independent

essays), and studies comparing L1 and L2 writing (Crossley & McNamara, 2012; Graesser, Jeon, Yang, & Cai, 2007; Guo, Crossley, & McNamara, 2013; Louwerse, McCarthy, McNamara, & Graesser, 2004).

## Literature Review

### Coh-Metrix Analyses of L1 and L2 Writing

Crossley and McNamara (2012) investigated features of L2 English academic essays related to cohesion and linguistic sophistication at the surface, textbase, and situation model levels in order to identify variables that distinguish between higher- and lower-proficiency L2 writers. The essays were sourced from the Hong Kong Advanced Level Examination corpus; they were written by graduating high school students in Hong Kong and graded by groups of trained raters from the Hong Kong Examinations and Assessment Authority (p. 119). From the Coh-Metrix indices (for descriptions of individual indices, see [the Coh-Metrix website](#) or McNamara et al., 2014), organised into banks relevant to text cohesion and linguistic sophistication at each of the three levels of analysis, Crossley and McNamara first selected 14 indices that were most highly correlated with human scores. After performing a linear regression analysis, they identified the following five variables, the combination of which accounted for 30% of the variance in the essay evaluation training set (and 21% in the test set): lexical diversity, content word frequency, word meaningfulness, average of word familiarity and aspect repetition.

Similar to previous research (e.g., Ferris, 1994; Grant & Ginther, 2000), the surface code and textbase variables were the main predictors of L2 essay scores in this study, with highest graded L2 texts characterised by a higher degree of lexical diversity and lower word frequency, meaningfulness, and familiarity values. On the other hand, highly scored L2 essays in this study had less aspect repetition (i.e., a cohesion variable at the situation model level). This finding is counter to the claim that aspect (and tense) repetition is used by L2 writers to increase text cohesion (Duran, McCarthy, Graesser, & McNamara, 2007). Similarly, the use of two other cohesion variables (given vs. new information and positive logical connectors, e.g., *and*, *also*, *then*) was negatively correlated with essay ratings, suggesting that less cohesive academic texts were scored higher by trained raters. In fact, linguistic indices related to cohesion were not found to be good predictors of human judgements of essay quality across a number of L2 and L1 studies (for an overview, see Crossley & McNamara, 2011b).

In an L1 English study, McNamara et al. (2010) reported that Coh-Metrix indices of syntactic complexity, lexical diversity and word frequency were most predictive of writer proficiency for argumentative essays written by undergraduate university students. Essays written by L1 students were scored higher when they had more syntactically complex constructions (with more words before the main verb), more lower-frequency words, and a greater range of lexical diversity.

### Coh-Metrix Analyses of Different Types of Academic Writing

The Coh-Metrix tool was used to compare writing quality of integrated and independent essays in the Test of English as a Foreign Language (TOEFL iBT) (Guo et al., 2013). Based on previous studies that examined linguistic features predictive of L2 writing proficiency in independent and integrated essays (Cumming et al., 2005; Ferris, 1994; Grant & Ginther, 2000), Guo et al. (2013) used computational indices of lexical sophistication, syntactic complexity, cohesion, and text length in their analysis of TOEFL iBT essays written by 240 English as a Second Language (ESL) and English as a Foreign Language (EFL) test-takers. The Coh-Metrix indices of linguistic features successfully predicted human scores for both types of essays. Text length (in words) was the strongest predictor of the human scores for both text types, with longer essays attracting higher ratings. Two other significant predictors of the human ratings common to the integrated and independent essays were the use of past participle verbs (indicative of the use of passive voice) and word frequency (measured by linguistic indices of frequency, familiarity, and average syllables per word). Essays with more instances of past participle verbs were graded higher due to passive voice being associated with more formal academic writing style (Hinkel, 2002). Higher-rated essays were also characterised by the use of lower-frequency words—a sign of lexical sophistication.

Guo et al. (2013) also found differences in the predictive features for the two types of essays. The integrated essays (written as compare and contrast texts in response to two formal sources of information) were scored higher when they were written in a more detached style, marked by greater semantic similarity between sentences and the use of third person singular present forms. The quality of the integrated essays was also judged higher when they contained more instances of referential cohesion (semantic similarity, noun overlap, and tense repetition), presumably due to the need to integrate and compare information from two separate sources. This was not the case for the independent essays that were argumentative texts based on test-takers' prior knowledge and personal experiences and views. The latter were graded higher when fewer cohesive devices were used (i.e., fewer conditional connectives, less aspect repetition, and less content word overlap). Thus, studies, in which the Coh-Metrix tool was used to generate linguistic and discourse indices, show that computational measures can be used to predict L1 and L2 writing proficiency and distinguish between different writing tasks.

## **The Study**

The goal of the present study is to compare language and text (micro) features of the two kinds of written artefacts produced by the same students enrolled in the same postgraduate course: academic blog posts and academic assignments. Since blogs are gaining momentum as a learning and teaching tool, their educational affordances for L1 and L2 academic writing need to be understood and built upon. By comparing textual features of blogs and assignments while controlling for the author and topic variables, differences (if any) between the two academic writing modes can be better understood. This understanding can then be used to predict how educational blogging may affect student writing practices and assist in making informed decisions on whether or not to use course blogs in tertiary education. So far, it appears that research into L1 educational blogging (e.g., Foster, 2015) focuses primarily on the macro-level of discourse (ideational, epistemological, pragmatic, and socio-cultural content). L2 writing research, on the other hand, tends to compare the effect of blogging and more traditional forms of ESL or EFL written assignments on student achievements in criterion-based writing tests, scored by human markers or raters (e.g., Lin, Li, Hung, & Huang, 2014). Conversely, the present study puts the focus of the investigation squarely on the nature of the textual artefact.

Linguistic and discourse aspects of student-generated texts are examined in order to address the following research questions (RQs):

1. Do academic blog posts and traditional assignments generated by students taking the same university course differ in their language and discourse characteristics?
2. Do academic texts generated by L1 and advanced L2 writers within the same university course differ in their language and discourse characteristics?
3. Are differences associated with the two text types more apparent in the artefacts produced by L1 or L2 writers?

## **Participants**

The study was conducted with students taking a 10-week post-graduate course in Computer-assisted Language Learning (CALL), as part of the Master's programme in Applied Linguistics and Teachers of English to Speakers of Other Languages (TESOL). The texts included in the study were collected over three years, from 38 consenting students: 11 high-proficiency L2 speakers (from countries in East and Southeast Asia) and 27 L1 speakers of English. Each year's cohort consisted of on-campus and distance students.

## **Data Collection and Analysis**

### **Materials**

The course included a number of assessed written tasks: (1) regular reflections on course readings in the form of blog posts open to the course instructor and students (about 500 words per entry), and (2) traditional

essay-style written assignments submitted as document files to the instructor only. One of these assignments was a literature-informed software resource evaluation, and the other was a project design document that included a literature-informed project justification section (used in this analysis). The course blog was created using the *Campus Pack* blog tool integrated into the university online course web site. Students could create and edit their own blog entries and comment on entries from other students. They could subscribe to updates via email and RSS feeds.

Two samples of the two types of writing by each student (152 texts) were used in the study: 76 traditional assignments (ASGN) and 76 blog posts (BLOG). The two blog posts chosen for each student were those closest in length to 500 words. Essay-style texts used in the analysis were one or more consecutive whole sections of the assignments, as close as possible in length to 500 words. The selected texts (across the two types) were on average 565 words long (Min = 354; Max = 778), with a standard deviation of 81 words (less than one half of average length, as recommended in McNamara et al., 2014). For one-paragraph excerpts from sample L1 and L2 student submissions, see [Appendix A](#).

### **Text Preparation**

All texts were anonymised and tagged for language (L1 or L2) and text type (ASGN or BLOG). The data cleaning procedure was based on that used by McNamara et al. (2014). Spelling errors were corrected. The following elements were deleted from the text: sentences with long citations (i.e., quoted text), in-text citations in parentheses, tables, and figures. Bulleted lists were converted into one sentence with a semicolon as a separator. Course-specific abbreviations were replaced with full text (e.g., A2 was replaced with *assignment 2*).

### **Statistical Analysis**

First, computational indices at the surface, textbase, and situation model levels were generated for each text using the Coh-Metrix tool, and theoretically and empirically motivated indices (based on the Coh-Metrix studies outlined in the introduction; see also McNamara et al., 2014) were selected for the initial analysis (see [Appendix B](#)). After this, averages per index were calculated for individual writers for each submission type, so that each student had two values for each index: one for the BLOG type and one for the ASGN type. Where non-normal distributions were observed, index values were logarithmically transformed to bring them closer to normality. Highly-correlated indices ( $r > .7$ ) within each level of analysis were identified, and only one index per correlated set was used in the analysis, in order to avoid collinearity (Crossley & McNamara, 2012).

A separate Multivariate Analysis of Variance (MANOVA) was conducted for the surface, textbase and situation model levels, with relevant Coh-Metrix indices used as dependent variables in each analysis. To investigate differences between the two text types and the L1 and L2 writers, the independent variables of interest in this study were text type (BLOG or ASGN) and language (L1 or L2). The Wilks' Lambda test statistic was used in all MANOVAs. The Least Significant Difference (*LSD*) statistic was used in post-hoc pairwise multiple comparison tests (Crossley & McNamara, 2011a). An alpha level of .05 was used for all reported analyses.

### **Human Ratings**

The texts used in the study were also rated by two human raters with prior experience in teaching or marking academic writing and with advanced university degrees in applied linguistics. The texts were rated on a 7-point scale on three criteria corresponding to the levels of the Coh-Metrix analysis: (1) use of language, (2) explicit textbase, and (3) situation model (see [Appendix C](#)). Each rater was allocated half of the texts ( $n = 76$ ) using a pseudo-random approach, such that they rated an equal number of the two text types written by L1 and L2 writers.

## Results

### Descriptive Statistics

The total number of paragraphs index was not included in the analysis due to a high correlation with the average number of sentences in each paragraph. The final analysis of variance included the following descriptive indices: total number of sentences in the text, total number of words in the text, the average number of sentences in each paragraph and the standard deviation for number of sentences per paragraph.

The MANOVA showed no reliable difference between L1 and L2 texts ( $F_{(4,69)} = 1.42, p = .24, \eta^2_p = .08$ ) across all indices, although the L1 texts used in the analysis were on average about 30 words longer than the L2 texts. The difference between the two text types (BLOG and ASGN) in the MANOVA was reliable ( $F_{(4,69)} = 6.25, p < .001, \eta^2_p = .27$ ). On average the ASGN texts were longer than BLOG texts by 46.34 words ( $SE = 14.26; p < .01$ ) and by about 4 sentences ( $SE = 1.15; p < .01$ ). The BLOG texts had on average about one sentence more per paragraph than the ASGN texts ( $SE = 0.09; p < .05$ ). There was no interaction between the two variables ( $F < 1$ ).

### Surface Code

Nine indices were used in the surface code analysis. Three indices indicated syntactic complexity: words before the main verb of the main clause in sentences, modifiers per noun-phrase, and syntactic structure similarity between all adjacent sentences. Six indices represented lexical sophistication: average word frequency for content words, age of acquisition (i.e., at what age a word tends to be first used by a native speaker; e.g., the word *smile* has a lower age of acquisition index than the word *dogma*), concreteness, imagability (i.e., how easy it is to construct a mental image of the word), meaningfulness (defined in terms of the number of associations a word has with other words, e.g., the word *people* has a high Coh-Metrix meaningfulness score of 612 and the word *abbess*, a low meaningfulness score of 218), and hypernymy for nouns and verbs. Hypernymy indices indicate a relative location of a word within a conceptual hierarchy of superordinate and subordinate words. For example, the word *entity* (that has a low hypernymy value) is at the top of a conceptual hierarchy and is more abstract; the word *chaise* (high hypernymy value) is at the bottom of the hierarchy (with many levels above it, i.e., *chair, seat, furniture, artefact, object*) and is more concrete and specific. High hypernymy values may reflect more *formal* words, for example, *Golden Retriever* sounds more formal compared to *dog* or *animal* in a sentence, “I drove the Golden Retriever to the doctor's office” (Crossley, Salsbury, & McNamara, 2009, p. 311).

The MANOVA showed reliable effects of the text type ( $F_{(9,64)} = 7.32, p < .001, \eta^2_p = .51$ ) and language ( $F_{(9,64)} = 3.67, p < .01, \eta^2_p = .34$ ) predictors, but no significant interaction between them ( $F < 1$ ). For the text type variable, the post-hoc tests showed that essay-style assignments were more syntactically complex than blogs posts, based on the mean number of modifies per noun-phrase (see Table 1). The syntactic structure similarity between adjacent sentences was greater in essay-style assignments than in blog posts. Lexical sophistication was higher in the traditional assignments than in the blogs; in the ASGN texts, the average word frequency for content words was lower and more content words that a native speaker would be expected to acquire later in life were used. On the other hand, the average ratings for concreteness and imagability were higher in the ASGN than in the BLOG texts. In addition, the hypernymy values for nouns and verbs were higher in the essay-style submissions.

For the language variable, the post-hoc tests showed that L2 writers produced more syntactically complex texts than L1 writers, based on the mean number of words before the main verb and the mean number of modifies per noun phrase. The L2 writers, however, used more similar syntactic structures compared to the L1 writers (Table 2). The L2 writers produced texts that were less lexically sophisticated: they used words that scored higher on meaningfulness, concreteness, imagability, and noun and verb hypernymy.

Table 1. *The Post-hoc Tests for the Text Type Variable, Surface Level Analysis*

Dependent Variable	Mean Diff. ASGN–BLOG	SE	p	95% Confidence Interval		
				Lower Bound	Upper Bound	
Words before the main verb of the main clause (log) <sup>a</sup>	0.03	0.06	.63	-0.10	0.16	
Modifiers per noun-phrase	0.13	0.03	4.E-06	0.08	0.18	***
Syntactic structure similarity (all adjacent sentences)	0.01	0.04	1.E-03	0.01	0.02	***
Average word frequency for content words	-0.10	0.02	1.E-04	-0.15	-0.05	***
Age of acquisition	8.61	4.03	.04	0.57	16.65	*
Concreteness	10.50	2.91	1.E-03	4.70	16.30	***
Imagability	9.86	2.68	5.E-04	4.51	15.20	***
Meaningfulness	2.14	2.79	.45	-3.42	7.69	
Hypernymy for nouns and verbs <sup>b</sup>	0.14	0.04	5.E-04	0.06	0.21	***

Notes. Here and henceforth: \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$ , \*\*\* indicates  $p < 0.001$ ; <sup>a</sup> (log) indicates that values have been logarithmically transformed, <sup>b</sup> higher hypernymy values may suggest more specific and formal words (see McNamara et al., 2014).

Table 2. *The Post-hoc Tests for the Language Variable, Surface Level Analysis*

Dependent Variable	Mean Diff. L2–L1	SE	p	95% Confidence Interval		
				Lower Bound	Upper Bound	
Words before the main verb of the main clause (log) <sup>a</sup>	0.15	0.06	.02	0.02	0.28	*
Modifiers per noun-phrase	0.07	0.03	8.E-03	0.02	0.12	**
Syntactic structure similarity (all adjacent sentences)	0.01	4.E-03	6.E-03	3.E-03	0.02	**
Average word frequency for content words	-0.03	0.02	.19	-0.08	0.02	
Age of acquisition	-5.79	4.03	.16	-13.83	2.24	
Concreteness	7.53	2.91	.01	1.73	13.33	*
Imagability	6.17	2.68	.02	0.83	11.51	*
Meaningfulness	7.55	2.79	8.E-03	1.99	13.10	**
Hypernymy for nouns and verbs <sup>a</sup>	0.11	0.04	4.E-03	0.04	0.19	**

Notes. <sup>a</sup> higher hypernymy values may suggest less-developed lexical networks (see Crossley et al., 2009).

### Textbase

Seven indices were used in the textbase analysis: one index of lexical diversity (*vocd* lexical diversity measure for all words) and six cohesion indices. One of the cohesion indices, global noun overlap, was related to referential coherence (or co-reference), that is, the noun overlap of each sentence with every other sentence in the text. The other five cohesion indices were related to connectives-based cohesion: causal connectives (e.g., *because*, *so*), logic connectives (e.g., *and*, *or*), temporal connectives (e.g., *first*, *until*),

positive connectives (e.g., *also*, *moreover*), and negative connectives (e.g., *however*, *but*).

The MANOVA showed a reliable effect of the language variable ( $F_{(7,66)} = 3.04$ ,  $p < .01$ ,  $\eta^2_p = .24$ ), but only a trend towards the text type effect ( $F_{(7,66)} = 1.82$ ,  $p = .098$ ,  $\eta^2_p = .16$ ), and no reliable interaction between the two variables ( $F < 1$ ). For the text type predictor, the post-hoc tests showed that two indices had a trend toward difference that did not reach conventional reliability (see Table 3). Referential cohesion via global noun overlap was higher in the essay-style assignments than in the blog posts. On the other hand, somewhat fewer temporal connectives were used in the ASGN than in the BLOG texts.

**Table 3.** *The Post-hoc Tests for the Text Type Variable, Textbase Analysis*

Dependent Variable	Mean Diff. ASGN-BLOG	SE	p	95% Confidence Interval	
				Lower Bound	Upper Bound
Lexical diversity measure for all words	-4.77	4.33	.28	-13.41	3.87
Global noun overlap	0.05	0.03	.06	-2.E-03	0.11
Causal connectives a	0.49	1.87	.79	-3.23	4.21
Logic connectives	2.93	2.29	.21	-1.63	7.48
Temporal connectives	-2.22	1.18	.06	-4.57	0.14
Positive connectives	-1.88	2.46	.45	-6.79	3.02
Negative connectives	0.30	1.16	.80	-2.01	2.62

Note. <sup>a</sup> The score for causal connectives has been shown to discriminate between the high and low cohesion texts (see McNamara et al., 2014).

For the language variable, the post-hoc tests showed that compared with L1 writers, L2 writers produced text that was higher in explicit cohesion, as a result of using more causal and logical connectives (see Table 4).

**Table 4.** *The Post-hoc Tests for the Language Variable, Textbase Analysis*

Dependent Variable	Mean Diff. L2–L1	SE	p	95% Confidence Interval		
				Lower Bound	Upper Bound	
Lexical diversity measure for all words	-7.50	4.33	.09	-16.14	1.14	
Global noun overlap	0.05	0.03	.08	-0.01	0.11	
Causal connectives	7.36	1.87	2.E-04	3.64	11.08	***
Logic connectives	6.37	2.29	.007	1.81	10.93	**
Temporal connectives	-2.19	1.18	.07	-4.54	0.16	
Positive connectives	3.79	2.46	.13	-1.12	8.69	
Negative connectives	0.45	1.16	.70	-1.87	2.76	

### Situation Model

10 indices were used in the situation model analysis. Two of them were semantic overlap indices based on Latent Semantic Analysis (LSA; Landauer et al., 2007): the mean of the LSA cosines between adjacent paragraphs (measure of semantic similarity in the high-dimensional word-by-context LSA matrix) and the average givenness of each sentence (information available for recovery from the preceding discourse). One index was that of word part-of-speech (i.e., the number of personal pronouns per 1000 words). Four were syntactic pattern density indices: the incidence score of preposition phrases, the incidence score of agentless passive voice forms, the incidence score of gerunds, and the incidence score of infinitives. Three were

situation model indices: the temporal cohesion index based on the repetition score for tense and aspect, the causal content index based on the incidence of causal verbs and causal particles in text, and the causal cohesion index based on a ratio of causal particles to causal verbs.

The MANOVA showed reliable effects of the language ( $F_{(10,63)} = 4.18, p < .001, \eta^2_p = .40$ ) and text type ( $F_{(10,63)} = 13.61, p < .001, \eta^2_p = .68$ ) predictors, but no significant interaction between them ( $F_{(10,63)} = 1.16, p = .33, \eta^2_p = .16$ ). The post-hoc tests for the text type variable (see Table 5) showed that the BLOG texts had significantly more instances of personal pronouns per 1000 words and gerund than the ASGN texts, and more semantic overlap in adjacent paragraphs. The ASGN texts also had more incidents of *agentless* passive voice forms and higher repetition scores for tense and aspect.

Table 5. The Post-hoc Tests for the Text Type Variable, Situation Model Analysis

Dependent Variable	Mean Diff. ASGN–BLOG	SE	p	95% Confidence Interval		
				Lower Bound	Upper Bound	
LSA semantic similarity (adjacent paragraphs)	-0.04	0.02	.04	-0.08	-3.E-03	*
LSA average givenness of each sentence	0.01	0.01	.27	-0.01	0.02	
Number of personal pronouns per 1000 words (log)	-0.60	0.08	2.E-10	-0.76	-0.44	***
Incidence of preposition phrases	-2.79	2.85	.33	-8.47	2.90	
Incidence of agentless passive voice forms	5.24	1.10	9.E-06	3.05	7.43	***
Incidence of gerunds	-4.20	1.76	.02	-7.71	-0.70	*
Incidence of infinitives	1.30	1.73	.45	-2.15	4.75	
Tense and aspect repetition	0.03	0.02	.03	2.E-03	0.06	*
Incidence of causal verbs and particles in text (log)	0.08	0.05	.12	-0.02	0.19	
Ratio of causal particles to causal verbs (log)	-0.20	0.14	.14	-0.48	0.07	

For the language variable, the post-hoc tests (see Table 6) showed that the L2 writers produced texts that were higher in referential cohesion than those produced by L1 writers: the L2 texts had more semantic overlap in adjacent paragraphs and a higher proportion of given (vs. new) information. The texts generated by the L2 writers had more instances of agentless passive voice forms. The L2 writers used more gerunds, infinitive forms, causal verbs, and particles in their writing than L1 writers.

Table 6. *The Post-hoc Tests for the Language Variable, Situation Model Analysis*

Dependent Variable	Mean Diff.	SE	p	95% Confidence Interval		
	L2–L1			Lower Bound	Upper Bound	
LSA semantic similarity (adjacent paragraphs)	0.05	0.02	.02	0.01	0.09	*
LSA average givenness of each sentence	0.02	0.01	.02	0.00	0.03	*
Number of personal pronouns per 1000 words (log)	-0.01	0.08	.91	-0.17	0.15	
Incidence of preposition phrases	-0.32	2.85	.91	-6.00	5.37	
Incidence of agentless passive voice forms	2.55	1.10	.02	0.36	4.73	*
Incidence of gerunds	4.16	1.76	.02	0.66	7.66	*
Incidence of infinitives	3.68	1.73	.04	0.23	7.12	*
Tense and aspect repetition	-0.01	0.02	.38	-0.04	0.02	
Incidence of causal verbs and particles in text (log)	0.21	0.05	2.E-04	0.10	0.32	***
Ratio of causal particles to causal verbs (log)	0.20	0.14	.14	-0.07	0.48	

### Analysis of the Human Ratings

The ratings of the use of language, textbase, and situation model were used as dependent variables, while the text type (ASGN or BLOG) and writer language (L1 or L2) were used as primary-interest predictors (i.e., independent variables). For each rater, the ratings were first standardised and centred using the *scale()* function in R.

Linear mixed effects modeling using the *lmer* function (Bates, 2011; Bates & Sarkar, 2010) of the interactive programming environment R was used in the analyses of the ratings. In the regression analyses of the data, *t*-values larger than 2 indicated a significant effect at  $p < .05$  (Baayen, 2008). The study participants were included in the model as random effects. A separate model was constructed for each rating criterion. A minimally adequate statistical model was fitted to the scaled ratings data, using a stepwise variable selection and the likelihood ratio test for model comparisons (Baayen, Davidson, & Bates, 2008).

On the use of language criterion (corresponding to the surface code level in the Coh-Metrix analysis), the analysis showed reliable effects of both text type and writer language, and a reliable interaction between the two predictors (see Table 7, section A). The blog posts were rated lower than traditional assignment for L2 writers, but this was not the case for the L1 writers (Figure 1). The texts produced by the L1 writers were rated higher than those produced by the L2 writers on this criterion. The findings for the human ratings of language use were mostly aligned with the Coh-Metrix analysis at the surface code level: the text type and writer language were found to be reliable predictors in both analyses. However, the interaction between the two predictors was only observed in the analysis of human ratings.

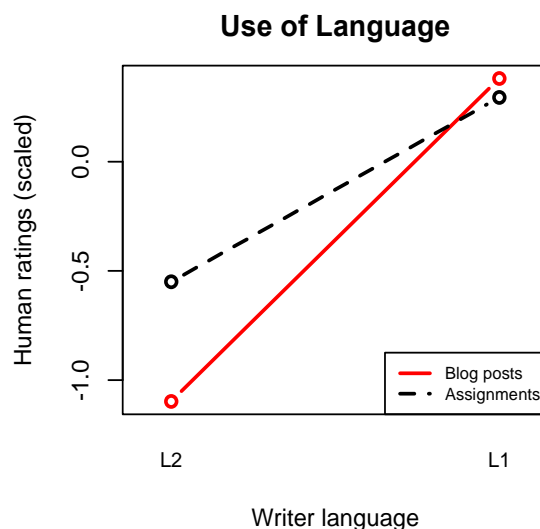


Figure 1. Analysis of human ratings of language use: an interaction between text type and writer language.

On the textbase criterion, the analysis of ratings (see Table 7, section B) showed a reliable effect of writer language (the L1 texts were rated higher than the L2 texts), but the effect of text type was not statistically reliable, and there was no interaction between the two predictors. A similar pattern was observed in the Coh-Metrix analysis at the textbase level, where the effect of writer language was reliable, but only a weak trend was observed for the text type variable. Finally, the analysis of human ratings on the situation model criterion showed a reliable effect of writer language (in favour of the L1 texts), but not text type (Table 7, section C), and no interaction.

Table 7. Analysis of Human Ratings

Rating criterion:	A. Use of Language			B. Textbase			C. Situation Model		
	$\beta$	<i>SE</i>	<i>t</i>	$\beta$	<i>SE</i>	<i>t</i>	$\beta$	<i>SE</i>	<i>t</i>
(Intercept)	-0.55	0.21	-2.67	-0.49	0.19	-2.64	-0.31	0.18	-1.71
Text type = BLOG	-0.55	0.21	-2.58	0.15	0.15	1.01	-0.21	0.15	-1.43
Writer_language = L1	0.84	0.24	3.47	0.59	0.20	2.89	0.59	0.20	2.99
Text type = BLOG: Writer_language = L1	0.64	0.25	2.51	–	–	–	–	–	–

Notes. Intercept levels: Text type = ASGN; Writer language = L2.

## Discussion

The analysis of the written artefacts produced by the students taking the same graduate course showed that educational blogging and traditional essay-style course assignments differ in some language and discourse characteristics (RQ1). The traditional academic texts were more lexically sophisticated, that is, they included more lower-frequency vocabulary and words that are acquired by native speakers later in life.<sup>1</sup> The assignment submissions were also more syntactically complex than the blog posts (as indicated by the number of modifiers per noun-phrase). Since Coh-Metrix linguistic indices of lexical sophistication and syntactic complexity are predictive of human judgements of essay quality (Crossley & McNamara, 2012; McNamara et al., 2010), the essay-style assignments in the present study may be expected to attract higher human ratings than the blogs (because the assignments were shown to be more lexically sophisticated and syntactically complex). This was the case for L2, but not L1 writers (see Figure 1) presumably because the

quality of a blog post is judged differently than that of an academic essay due to the nature of the two text types.

Guo et al. (2013) reported that there were differences in the language and discourse features predictive of human ratings for integrated and independent essays. The integrated (but not independent) essays were scored higher when they had a higher degree of referential cohesion (through noun overlap and tense repetition). Similarly, compared to the blogs, the assignments in the present study were characterised by greater referential cohesion via global noun overlap and more repetition of grammatical tense and aspect. Furthermore, the integrated essays were scored higher when they were written in a more formal, detached style. The assignments in the present study were also more formal than the course blog posts, as indicated by the use of agentless passive voice and higher hypernymy values for nouns and verbs (i.e., preference for more specific, formal words). They also had more similar syntactic structures, characteristic of more formal texts (Graesser et al., 2014).

The independent essays that incorporated the writers' personal experiences and points of view (Guo et al., 2013) were scored higher when less content overlap and less aspect repetition was used. Similarly, the blog posts (across the two groups of writers in the present study) were characterised by less referential cohesion and less repetition of grammatical tense and aspect. On the other hand, the blogs included more instances of semantic and situational redundancies that contribute to text readability and ease of understanding and communicated a clearer sense of agency, containing more personal pronouns and fewer passives.

These language and discourse differences between the two text types were observed in the artefacts produced by L1 or L2 writers (RQ3). However, the L2 writers were less able than the L1 writers to achieve the same quality of writing in their course blogs as in traditional academic assignments; L2 writing was scored lower by human raters on blog posts than on assignments (Figure 1). This suggests that the L2 writers may have found moving away from the familiar impersonal academic writing style to a relatively unfamiliar style of education blogging challenging.

In summary, the analysis of text types shows that, compared to the essay-style assignments that adhered to a more formal, complex, impersonal, and repetitive academic style, educational blogging was less formal, less formulaic, less semantically and syntactically complex, and easier to understand. Blogging also created a clearer representation of agency. Thus, the written artefacts generated by the students on similar topics within the same course differed as a result of the medium used to create these texts and an expectation of a broader readership. Moreover, these differences were observed in the Coh-Metrix analysis for both the L1 and L2 writers, suggesting that the effect of writing medium occurs when writing in the first and second language. However, the blog posts generated by the L2 writers were rated lower by the human raters than their traditional course assignments, suggesting that the quality of writing suffered when the L2 writers were writing outside of the more familiar, formal academic writing mode. This effect did not occur for L1 writers, whose blog texts were rated at about the same quality as their essay-style course submissions, in terms of language use.

The present study also explored differences between the writing of L1 and L2 postgraduate students, across academic writing and academic blogging (RQ2). Friginal et al. (2014) showed both similarities and differences in the use of linguistic and discourse features (and clusters of features) in highly rated L1 and L2 texts (for overviews, see Crossley & McNamara, 2011b; Hinkel, 2011). However, more research is needed in this area, as results so far are limited and inconclusive (Friginal et al., 2014). Hinkel (2011) also points out that L2 writers' language proficiency is a common confounding factor in research comparing L1 and L2 writing. The present study shows a number of differences in the use of language and discourse structures in the written artefacts generated by the L1 and advanced L2 writers.

Across both text types, the L2 writers produced more syntactically complex texts than the L1 writers, but used more similar syntactic structures and more meaningful, concrete, and imagable words (i.e., words that were less lexically sophisticated). They also used words that had higher hypernymy values, associated with more formal discourse (Ordonez, Carlo, Snow, & McLaughlin, 2002). Furthermore, previous studies have

shown that lexical development is associated with an increased depth of lexical networks and ability to use abstract language, corresponding to a decrease in hypernymy values (Crossley et al., 2009). Therefore, the finding that texts of L2 writers had higher hypernymy values may also indicate their lower lexical proficiency compared to the L1 writers. This conjecture is supported by the finding that, similar to previous studies (Crossley & McNamara, 2011b), the L2 texts were reliably less lexically sophisticated than L1 texts.

The L2 writers produced texts that were higher in textual and situational cohesion in comparison with the L1 writers. They used more causal and temporal connectors, more causal verbs and causal particles, and more semantic overlap in adjacent paragraphs. L2 writing also contained a higher proportion of given (vs. new) information. Higher values of cohesion indices in previously published Coh-Metrix analyses have been negatively correlated with human raters' scores of academic writing (Crossley & McNamara, 2012, 2011b). Thus, the texts produced by the L2 writers in the present study were predicted to attract lower ratings than those of the L1 writers. This was indeed the case in the analysis of human ratings on the textbase and situation model criteria. Finally, the finding that the L2 writers' texts had more incidents of agentless passive voice forms suggests that they were more inclined to use impersonal academic writing style than the L1 writers, across both text types (Hinkel, 1997). This result corroborates the proposed reasons why L2 student blog posts were rated lower than their assignments.

## Limitations

The Coh-Metrix analysis of differences between course blogs and traditional assignments has produced interesting and novel findings, but an additional linguistic analysis of the two text types by a human researcher may be able to further explain and interpret some of these differences. For example, the different hypernymy values may be interpreted in a number of ways relative to the basic category word (Brown, 1958; Murphy, 2004). That is, in relation to the basic level word *car*, the word *vehicle* will have a lower hypernymy index and be considered a more abstract concept, and the word *sedan* will have a higher hypernymy index and be considered a more specific, formal word (Crossley et al., 2009). In the present study, higher hypernymy values are interpreted to represent a more formal style. However, hypernymy indices used in Coh-Metrix are based on the WordNet database (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) and provide values in absolute taxonomic hierarchies rather than values relative to the basic category word in each hierarchy. Therefore, the findings and conclusions based on the Coh-Metrix hypernymy indices can only be tentative. Another example of a limitation in the ability to interpret a computational-only analysis is the syntactic complexity measure—words before the main verb. This is because syntactic complexity may indicate a more sophisticated manner of academic writing (i.e., higher essay quality) or a less skilled use of syntactic structures in communicating ideas and arguments (i.e., lower writing quality). The relationship between computational measures of syntactic complexity and human ratings of writing quality depends on the genre and mode of writing, chosen topic, and task (Crossley & McNamara, 2014; Crowhurst, 1983; Lu, 2011; Yanga, Lub, & Cushing Weigle, 2015). However, when combined with the results for other indices within the same level of analysis and conceptual groupings of linguistic and discourse features, valid interpretations of computational indices can be achieved.

## Pedagogical Implications

This study has shown that the choice of writing environment affects the resulting written artefact, for L1 and L2 university students, even when academic writing on similar topics is undertaken within the same course. Educational blogging encouraged a less formal approach to academic writing with a clearer sense of agency and increased text readability (Gallagher, 2010; Sun & Chang, 2012). Language teachers and instructional designers may therefore benefit from carefully considering consequences of using blogs in academic courses. Using traditional assignments is likely to better prepare students for situations when they need to comply with academic or professional certification standards or meet university admission requirements. On the other hand, if the goal is to broaden the students' academic writing repertoire, adjusting the style of writing to the target audience and utilising affordances of different communication

modes, providing opportunities to engage in educational blogging is clearly worthwhile (Bloch, 2007; Foster, 2015; Jimoyiannis & Angelaina, 2012; Lee, 2010). This is particularly critical for L2 writers, who tend to use more formal, impersonal language in their academic writing (Hinkel, 1997; Hyland, 2002).

## Conclusions

This study examined language and discourse features of course blogs and traditional assignments by L1 and advanced L2 university students completing an advanced degree in applied linguistics and TESOL. Students enrolled in the same postgraduate course were instructed to write blog posts and traditional essay-style assignments on the topics covered in the course. This design made it possible to compare and contrast micro-features of the two text types generated by the same students at the surface code, explicit textbase, and situation model levels of discourse representation (Graesser et al., 1997). The computational analyses revealed reliable differences between the two text types at the surface and situation model levels. The textual artefacts of course blogging were less lexically sophisticated and syntactically complex than those of the assignments, were written in a less formal style, and had higher readability and a clearer sense of agency. These findings suggest that educational blogging creates opportunities for students to develop an academic writing style they can use to communicate complex ideas in a more informal manner to wider audiences—a valuable attribute of a university graduate. Based on the quality ratings of the two text types by human raters, it appears that the educational blogging task is more challenging for L2 students, whose native academic writing norms and L2 training may encourage more formal impersonal writing style. This conjecture should be investigated in future research into L2 blogging.

## Acknowledgements

I am grateful to Robin Reid and Mark Toomer for their assistance with text ratings. I am also grateful to three anonymous reviewers for their comments and suggestions. Finally, I would like to thank my MA students who gave me permission to use their course submissions in this research project.

## Notes

1. The assignments had more concrete and imagable words than the blogs. Although the latter finding may appear counter-intuitive, it makes sense if we consider that the two traditional assignments required the students to design or evaluate a learning task, tool, or programme (a more concrete subject matter), while the blog posts were discussions of course readings, primarily journal articles.

## References

- Arslan, R. S., & Sahin-Kizil, A. (2010). How can the use of blog software facilitate the writing process of English language learners? *Computer Assisted Language Learning*, 23(3), 183–197.
- Aydin, S. (2014). The use of blogs in learning English as a foreign language. *Mevlana International Journal of Education*, 4(1), 244–259.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Bates, D. M. (2011). Linear mixed model implementation in lme4. Retrieved from <http://cran.r-project.org/web/packages/lme4/vignettes/Implementation.pdf>
- Bates, D. M., & Sarkar, D. (2010). lme4: Linear mixed-effects models using S4 classes. Retrieved from <http://CRAN.R-project.org/package=lme4>

- Bloch, J. (2007). Abdullah's blogging: A generation 1.5 student enters the blogosphere. *Language Learning & Technology*, 11(2), 128–141. Retrieved from <http://lt.msu.edu/vol11num2/bloch/default.html>
- Brown, R. (1958). How shall a thing be called? *Psychological Review*, 65(1), 14–21.
- Ciftci, H., & Kocoglu, Z. (2012). Effects of peer e-feedback on Turkish EFL students' writing performance. *Journal of Educational Computing Research*, 46(1), 61–84.
- Crossley, S. A., & McNamara, D. S. (2011a). Shared features of L2 writing: Intergroup homogeneity and text classification. *Journal of Second Language Writing*, 20(4), 271–285.
- Crossley, S. A., & McNamara, D. S. (2011b). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, 21(2/3), 170–191.
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The role of cohesion, readability, and lexical difficulty. *Journal of Research in Reading*, 35(2), 115–135.
- 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.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2009). Measuring second language lexical growth using hypernymic relationships. *Language Learning*, 59(2), 307–334.
- Crowhurst, M. (1983). Syntactic complexity and writing quality: A review. *Canadian Journal of Education*, 8, 1–16.
- Cumming, A., Kantor, R., Baba, K., Erdoosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in writing-only and reading-to-write prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43.
- Duran, N. D., McCarthy, P. M., Graesser, A. C., & McNamara, D. S. (2007). Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behavior Research Methods*, 29(2), 212–223.
- Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, 28(2), 414–420.
- Foster, D. (2015). Private journals versus public blogs: The impact of peer readership on low-stakes reflective writing. *Teaching Sociology*, 43(2), 104–114.
- Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, 23, 1–16.
- Gallagher, J. (2010). "As y'all know": Blog as bridge. *Teaching English in the Two Year College*, 37(3), 286–294.
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371–398.
- Graesser, A. C., Jeon, M., Yang, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal*, 15(3), 199–213.
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *Elementary School Journal*, 115(2), 210–229.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36(2), 193–202.

- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163–189.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123–145.
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Writing Assessment*, 18(3), 218–238.
- Halic, O., Lee, D., Paulus, T. M., & Spence, M. (2010). To blog or not to blog: Student perceptions of blog effectiveness for learning in a college-level course. *Internet and Higher Education*, 13(4), 206–213.
- Hinkel, E. (1997). Indirectness in L1 and L2 academic writing. *Journal of Pragmatics*, 27(3), 361–386.
- Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum.
- Hinkel, E. (2011). What research on second language writing tells us and what it doesn't. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*, Vol. 2, (pp. 523–538). New York, NY: Routledge.
- Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34(8), 1091–1112.
- Jimoyiannis, A., & Angelaina, S. (2012). Towards an analysis framework for investigating students' engagement and learning in educational blogs. *Journal of Computer Assisted Learning*, 28(3), 222–234.
- Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Lee, L. (2010). Fostering reflective writing and interactive exchange through blogging in an advanced language course. *ReCALL*, 22(2), 212–227.
- Lester, J. N., & Paulus, T. M. (2011). Accountability and public displays of knowing in an undergraduate computer-mediated communication context. *Discourse Studies*, 13(6), 671–686.
- Lin, M. H., Li, J.-J., Hung, P. Y., & Huang, H.-W. (2014). Blogging a journal: Changing students' writing skills and perceptions. *ELT Journal*, 68(4), 422–431.
- Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 843–848). Mahwah, NJ: Lawrence Erlbaum.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication*, 27, 57–86.
- McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). *Five papers on WordNet* (Tech. Rep. No. 43). Princeton, NJ: Princeton University, Cognitive Science Laboratory.
- Murphy, G. L. (2004). *The big book of concepts*. Cambridge, MA: MIT Press.

- Murray, L., Hourigan, T., & Jeanneau, C. (2007). Blog writing integration for academic language learning purposes: Towards an assessment framework. *Iberica*, 14, 9–32.
- Ordonez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology*, 94(4), 719–728.
- Sun, Y.-C. (2010). Extensive writing in foreign-language classrooms: A blogging approach. *Innovations in Education and Teaching International*, 47(3), 327–339.
- Sun, Y.-C., & Chang, Y.-j. (2012). Blogging to learn: Becoming EFL academic writers through collaborative dialogues. *Language Learning & Technology*, 16(1), 43–61. Retrieved from <http://ilt.msu.edu/issues/february2012/sunchang.pdf>
- Vurdien, R. (2013). Enhancing writing skills through blogging in an advanced English as a foreign language class in Spain. *Computer Assisted Language Learning*, 26(2), 126–143.
- Warschauer, M., & Grimes, D. (2007). Audience, authorship, and artifact: The emergent semiotics of Web 2.0. *Annual Review of Applied Linguistics*, 27, 1–23.
- Xie, Y., Ke, F., & Sharma, P. (2008). The effect of peer feedback for blogging on college students' reflective learning processes. *The Internet and Higher Education*, 11(1), 8–25.
- Yanga, W., Lub, X., & Cushing Weigle, S. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53–76.

## Appendix A. Excerpts from One L1 and One L2 Student Texts

	Text type = BLOG	Text type = ASGN
L1	If we want our students to be able to engage in CALL effectively, they need to be able to do so not only in the computer lab or at home with their PCs, but whenever they have the urge or the opportunity. The ideal as I see it is to create seamless technology - for the tools that learners use for language input, output and interaction to be easy and intuitive. It is already thus for much of social networking and therefore shouldn't be such a stretch for it to become so for the various language learning resources. Additionally, as [author] notes, a tool ought to be matched appropriately with the task at hand. In many (but not all) cases the [technology] is an ideal way to facilitate this.	This language learning tool provides authenticity in what it does. The focus is on accuracy and complexity, but fluency practice will occur outside of the program. Perhaps the biggest weakness is that it does not allow learners to engage in extended conversation outside of the discussion boards. User feedback is an important part of [software] and it is still a relatively new tool, so it can be expected that it will continue to evolve, adding components (within the translating framework) that will further enhance language learning. This program is recommended for anyone wishing to study a new language, particularly as a learning strategy tool within a largely autonomous learning approach.

L2	<p>One of the main points of [source] about computer-mediated communication (CMC) is the actual role of CMC in the second language classroom. According to the authors, one of its noticeable innovatory roles compared with traditional L2 classrooms is its ability to involve learners in a mobile classroom with either synchronous or asynchronous interaction through which to negotiate meaning. [Author], cited by [author], proved that the language produced in the electronic discussions can be lexically and syntactically more complex than that of face to face interaction. It is true that with CMC, turn-taking can be distributed more equally among learners. Also, within asynchronous tools, learners can have time to perform more thoughtfully and coherently. Sharing the same idea about CMC role, [author] suggests that it helps assisting the process of collaborative learning by creating a new domain for both teacher and learners.</p>	<p>The three aspects of "language difficulty", "program difficulty" and "content" belonging to both syllabus and learner variables have to some extent been already discussed throughout the paper. By selecting some random reading texts and transcripts to analyse in [software], the language difficulty is proven to be compatible to these learners' language competence as more than 95% coverage are within 5000 level and low frequency words are rarely found. About "program difficulty", it can be suggested that [software] can hardly cause any technical problem to users as it is free and accessible for everyone to download materials within a couple seconds. Audios in mp3 format and texts in pdf are very accessible to users. One limited point about the site is that some activities have not been updated lately. However, as some overlapped ideas can be noticed between the old activities and the newly released ones, it can be because the site designers want to replace some tasks with the newer ones.</p>
----	--	--

## Appendix B. Coh-Metrix 3.0 Indices and Their Categorisation in the Present Study

Used	Coh-Metrix Index	Name and Description	Levels of Analysis	Conceptual Grouping
0	DESPC	Number of paragraphs	DS	DS
1	DESSC	Number of sentences	DS	DS
1	DESWC	Number of words	DS	DS
1	DESPL	Mean length of paragraphs	DS	DS
1	DESPLd	<i>SD</i> of the mean length of paragraphs	DS	DS
1	DESSL	Mean number of words (length) of sentences	SC	SY
1	DESSLd	<i>SD</i> of the mean length of sentences	SC	SY
1	DESWLsy	Mean number of syllables (length) in words	SC	LS
1	DESWLsyd	<i>SD</i> of the mean number of syllables in words	SC	LS
1	SYNLE	Words before main verb	SC	SY
1	SYNNP	Modifiers per NP	SC	SY
1	SYNSTRUTa	Syntactic structure similarity adjacent	SC	SY
0	SYNSTRUTt	Syntactic structure similarity all	SC	SY
1	WRDFRQc	Word frequency for content words	SC	LS
0	WRDFRQa	Word frequency for all words	SC	LS
0	WRDFRQmc	Minimum word frequency	SC	LS

1	WRDAOAc	Age of acquisition	SC	LS
0	WRDFAMc	Familiarity	SC	LS
1	WRDCNCc	Concreteness	SC	LS
1	WRDIMGc	Imagability	SC	LS
1	WRDMEAc	Meaningfulness	SC	LS
0	WRDPOLc	Polysemy	SC	LS
0	WRDHYPn	Hypernymy for nouns	SC	LS
0	WRDHYPv	Hypernymy for verbs	SC	LS
1	WRDHYPnv	Hypernymy for nouns and verbs	SC	LS
0	CRFNO1	Noun overlap (adjacent sentences)	TB	CO
0	CRFAO1	Argument overlap (adjacent sentences)	TB	CO
0	CRFSO1	Stem overlap (adjacent sentences)	TB	CO
1	CRFNOa	Noun overlap (global)	TB	CO
0	CRFAOa	Argument overlap (global)	TB	CO
0	CRFSOa	Stem overlap (global)	TB	CO
0	CRFCWO1	Content word overlap (adjacent sentences)	TB	CO
0	CRFCWO1d	<i>SD</i> of the adjacent content word overlap	TB	CO
0	CRFCWOa	Content word overlap (global)	TB	CO
0	CRFCWOad	<i>SD</i> of the global content word overlap	TB	CO
0	LDTTRc	Type-token ratio (content words)	TB	LS
0	LDTTRa	Type-token ratio (all words)	TB	LS
0	LDMTLD	MTLD lexical diversity measure (all words)	TB	LS
1	LDVOCd	<i>vocd</i> lexical diversity measure (all words)	TB	LS
1	CNCCaus	Causal Connectives	TB	CO
1	CNCLogic	Logical connectives	TB	CO
0	CNCADC	Adversative/contrastive connectives	TB	CO
1	CNCTemp	Temporal connectives	TB	CO
0	CNCAdd	Additive connectives	TB	CO
1	CNCPos	Positive connectives	TB	CO
1	CNCNeg	Negative connectives	TB	CO
0	SMCAUSv	Causal verbs	SM	CO
1	SMCAUSvp	Causal verbs and causal particles	SM	CO
1	SMCAUSr	The ratio of causal particles to causal verbs	SM	CO
1	SMTEMP	Temporal cohesion - tense and aspect repetition	SM	CO
0	LSASS1	LSA, adjacent sentence-to-sentence units	SM	CO
0	LSASS1d	<i>SD</i> of the LSA adjacent sentence-to-sentence units	SM	CO
0	LSASSp	LSA, sentence-to-sentence units (all)	SM	CO
0	LSASSpd	<i>SD</i> of the LSA, sentence-to-sentence units (all)	SM	CO

1	LSAPP1	LSA, adjacent paragraphs	SM	CO
0	LSAPP1d	SD for the LSA, adjacent paragraphs	SM	CO
1	LSAGN	LSA, the average givenness of each sentence	SM	CO
0	LSAGNd	SD of the LSA, the average givenness, each sentence	SM	CO
1	WRDPRO	Number of personal pronouns per 1000 words	SM	SY
0	WRDPRP1s	First person, single form pronouns	SM	SY
0	WRDPRP2	Second person pronouns	SM	SY
0	WRDPRP3s	Third person, single form pronouns	SM	SY
0	WRDPRP3p	Third person, plural form pronouns	SM	SY
1	DRPVAL	Agentless passive voice forms	SM	SY
1	DRGERUND	Gerunds	SM	SY
1	DRINF	Infinitives	SM	SY
1	DRPP	Preposition phrases	SM	SY

Notes. <sup>a</sup> Comprehensive documentation on the Coh-Metrix indices is available from <http://cohmetrix.com>; Levels of analysis: DS - Descriptive; SC - Surface Code; TB - textbase; SM - Situation Model; Conceptual grouping: DS - Descriptive; LS - Lexical Sophistication; SY - Syntactic Complexity; CO - Cohesion; Used Codes: 1 = used in the analysis; 0 = not used in the analysis.

## Appendix C. Rating Criteria and Descriptors for Human Raters

### 1. Use of Language (Surface Code)

*Descriptor.* Appropriate and effective use of linguistic resources: (1) appropriate and effective use of syntax (arrangement of words and phrases to create well-formed sentences); (2) appropriate, precise and effective use of vocabulary; (3) lexical sophistication.

*The rating of 7* (highest rating) is assigned when the writer uses a wide range of vocabulary and grammatical structures fluently and flexible to convey precise meaning; the language shows natural and sophisticated control of lexical and syntactic features.

*The rating of 1* (lowest rating) is assigned when the writing shows an extremely limited range of vocabulary and grammatical structures; little control of lexical, grammar and syntactic features; sentences are poorly constructed; the writing does not convey precise meaning.

### 2. Textbase

*Descriptor.* Local and global coherence: (1) effective use of cohesive devices (e.g., linkers, referents) within and between paragraphs; (2) well-formed paragraphs; (3) clear and logical structure; (4) logical sequencing of information and ideas.

*The rating of 7* (highest rating) is assigned when the writer uses cohesion in such a way that it attracts no special attention, creating logical, natural and fluent flow of written discourse. Paragraphing is skilfully managed.

*The rating of 1* (lowest rating) is assigned when a very limited range of cohesive devices is used, and those used may not convey causal or logical relationship; there is very little control over organisational features; poor use of paragraphing.

### 3. Situation Model

*Descriptor.* (1) Communication of agency, causality, intentionality; (2) quality of inferences, elaborations, idea development; (3) the writing supports the creation of a mental model of the situation or argument by

the reader.

*The rating of 7* (highest rating) is assigned when the writing creates a clear mental representation of the situation / context in the mind of the reader; presents a clear, fully developed position supported by fully extended and well developed ideas and argumentation.

*The rating of 1* (lowest rating) is assigned when the writing does not create a mental model of the situation, process or argument; does not express a position; may attempt one or two ideas but there is no development; the ideas are not logically organised; there is no clear progression.

### **About the Author**

Irina Elgort is senior lecturer in higher education at Victoria University of Wellington. Computer-assisted language learning and the use of digital technology in higher education are among her research interests. Her other research interests include second language vocabulary, mental lexicon, and reading in a second language.

**E-mail:** [irina.elgort@vuw.ac.nz](mailto:irina.elgort@vuw.ac.nz)