English Teaching, Vol. 75, No. 1, Spring 2020, pp. 25-48
DOI: https://doi.org/10.15858/engtea.75.1.202003.25
http://journal.kate.or.kr

Direct Writing, Translated Writing, and Machine-Translated Writing: A Text Level Analysis with Coh-Metrix

Yuah V. Chon and Dongkwang Shin*

Chon, Yuah V., & Shin, Dongkwang. (2020). Direct writing, translated writing, and machine-translated writing: A text level analysis with Coh-Metrix. *English Teaching, 75*(1), 25-48.

While learners may have access to reference tools during second language (L2) writing, the latest developments in machine translation (MT), such as *Google Translate* requires examination as to how using the tool may factor into the second language learners' writing products. To this end, the purpose of this study was to examine how MT may have an effect on L2 learners' writing products relative to when writers wrote directly in L2, or translated a text to English from Korean. EFL university learners were asked to write for prompts that were counterbalanced for three writing modes and three writing topics. The learners' writing products were analyzed with Coh-Metrix to provide information on text characteristics at the multilevel. The results indicate that MT could facilitate the learners to improve fluency and cohesion, produce syntactically complex sentences, and write concrete words to express their target messages. Pedagogical implications are provided for how MT can be used to improve the quality of the L2 learners' writing products.

Key words: L2 writing, machine translation, direct writing, translated writing, machine-translated writing, Coh-Metrix

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A2A01024598).

* First Author: Yuah V. Chon, Professor, Department of English Education, Hanyang University
Corresponding Author: Dongkwang Shin, Professor, Department of English Education, Gwangju National University of Education; Pilmun-daero 55, Buk-gu, Gwangju 61240, Korea; Email: sdhera@gmail.com

Received 4 December 2019; Reviewed 29 December 2019; Accepted 1 March 2020

1. INTRODUCTION

The skill that is necessary for the 21st century is 'technology literacy', which requires learners to understand the machines that make information accessible (Jewitt, 2012). That is, technology literacy demands learners to become familiar with the machines (e.g., computers, cloud programming, and mobile devices) involved for accessing the information they need to solve their problems within the Information Age. In a similar vein, current second language (L2) writing may also require learners to extend their abilities to search and analyze information as they write online. In addition to dictionaries and concordancers, the machine translator can be considered the posterior reference tool for L2 writing in the current era. Although the facility cannot yet substitute the role of a fluent human translator, web-based machine translation (MT), most noticeably Google Translate, can be expected to be the referencing technology of the next generation (Kirchhoff, Turner, Axelrod, & Saavedra, 2011; van Rensburg, Snyman, & Lotz, 2012). Traditionally, MT had been based on the idea of parsing and generations—that is, the computer understanding the underlying meaning of a text in the source language, and then applying a new set of rules belonging to the target language. However, by the 1980s, this emphasis had shifted to statistics-based translation practice, and this is still the dominant paradigm. This means that the system calculates probabilities of various translations of a phrase being correct, rather than aiming for word-for-word translation (Groves & Mundt, 2015).

In contrast to the way the MT may be used for L2 writing, there have been two major approaches to L2 writing. The first method is to write texts as if a native speaker would (i.e., Direct Writing), and the second method would be to write a text in the learners' native language and then translate it into target language (i.e., Translated Writing) as if a novice second writer would. The comparison between these two approaches has already been analyzed through a number of studies (Ali, 1996; Cohen & Brooks-Carson, 2001; Kobayashi & Rinnert, 1992; Uzawa, 1996) for their writing products and writing processes. However, empirical research that subjects Google Translate to analysis of specific linguistic features, to date, is hard to find, and there is scant positive evaluation of MT as a reference tool for L2 writing (Groves & Mundt, 2015; Niño, 2008; Somers, 2003).

To this end, this study aimed to compare the characteristics of writing products obtained through different writing modes with a focus on the use of MT as a potential skill of L2 writing. That is, realizing that the use of MT is on the rise, it was felt necessary to examine the renditions of MT via measures of linguistic complexity (i.e., lexical diversity, syntactic complexity, cohesion) in order to seek affordances and caveats of MT for possible implementation in L2 writing. For this purpose, the researchers examined texts composed by three different modes of writing, that is *direct writing* (i.e., writing directly in L2), *translated writing* (i.e., learner writes in L1 and translates to L2 without the help of any reference source), and *machine-translated writing* (i.e., learner writes in L1 to submit to a machine translation for an L2 rendition) in order to conduct analysis with Coh-Metrix, a computational tool for analyzing linguistic and discourse representations of a text. The analysis sought to examine if Google Translate has the ability to produce stretches of communicative English by calculating indices of lexical diversity, syntactic complexity, and cohesion. The present study is expected to highlight possible changes that MT may bring to practices of L2 writing and to hopefully trigger more specific discussion on the benefits MT.

2. BACKGROUND

2.1. Direct Writing and Translated Writing

Although research on machine translated writing with a focus on analyzing the quality of the essays or written products are lacking within the field of ELT, previous research on direct writing versus translated writing may provide some preliminary evidence on the effects of MT on L2 writing. Uzawa (1996) compared second language learners' L1 writing, L2 writing, and translation from L1 into L2, focusing on writing and translating processes, attention patterns, and quality of language use. Thinking aloud, 22 Japanese ESL students studying at a Canadian college performed 3 tasks individually. When the think-aloud protocols were analyzed, it was found that (a) most students used a "what-next" approach both in the L1 and L2 writing tasks and a "sentence-by-sentence" approach in the translation task, (b) attention patterns in the L1 and L2 writing tasks were very similar, but quite different in the translation task. Attention to language use in the translation task was significantly higher than in the L1 and L2 writing tasks and, (c) scores on language use in the L1 and L2 writing tasks were similar, but scores on language use in the translation task were significantly better than in the L2 writing task.

Kobayashi and Rinnert (1992) examined English compositions written by 48 Japanese university students on: (1) differences between the texts resulting from two writing processes, one writing first in Japanese and then translating into English and the other composing directly in English. In terms of quality of content, organization, and style, lower-level writers tended to benefit from translation, whereas higher-level writers did not benefit much. Overall, syntactic complexity was greater in translations than in direct writings.

Cohen and Brooks-Carson (2001) asked thirty-nine intermediate learners of French to perform two essay writing tasks: writing directly in French as well as writing in the first language and then translating into French. Two-thirds of the students did better on the direct writing task across all rating scales; one-third, better on the translated task. While raters found no significant differences in the grammatical scales across the two types of writing, differences did emerge in the scales for expression, transitions, and clauses. In contrast to previous two studies, Cohen and Brooks-Carson (2001) found that a majority of the learners (2/3) did better on the direct writing task than on the translated writing tasks. However, the difference was found to be due to expression, transitions, and clauses.

All in all, previous studies on direct writing versus translated writing and results on the quality of compositions are not straightforward. With our main interest on text-level analysis of the writing products, the studies do not allow for comparisons. In a similar vein, results for examining the quality of direct writing versus machine translated writing is left to be investigated since MT, the recent complementary writing tool, has not been empirically tested for its effect on improving the quality of English compositions.

2.2. The Use of Machine Translation for Writing

There have recently been an increasing number of Web Sites that offer the service of MT of individual sentences or even whole texts. Google Translate is one of the most popular services of this kind. At the moment, Google Translate is a free multilingual machine translation service developed by Google, to translate text, speech, images, sites, or real-time video from one language into another. In this study, the interest is confined to the Google Translate Service which is available for all Internet users.

MT research started after World War II with the advent of computers (Slocum, 1988), but it was not until the 1980s that processing speeds were high enough to allow for commercial applications. Nonetheless, translation has had a bad reputation in foreign/L2 learning and teaching (Garcia & Pena, 2011). It is often associated with the Grammar Translation Method that for centuries guided the discipline, and mostly involved translating source texts from the language being learned (L2) into the mother tongue (L1).

The earliest studies with MT focused on translation in the L1 to L2 direction and saw MT output typically as a “bad model”, which involved the risk of providing the learner with ill-formed examples of L2 (Niño, 2008; Somers, 2003). However, the recent growth of the Internet has brought MT (e.g., Google Translate) to the center of our attention for language learning and teaching. Nonetheless, the use of MT for L2 writing has not received sufficient attention from researchers and practitioners on the efficacy or the potential problems of this reference tool.

The MT studies that exist were initially conducted primarily within the context of translation studies, which saw MT often as sources of errors (Kliffer, 2005; La Torre, 1999; Niño, 2008). Recent studies have centered more on asking students to spot mistakes in the machine-translated writing products and correcting errors (Groves & Mundt, 2015; Ismail &

Hartono, 2016; Wallwork, 2016). Groves and Mundt (2015) asked students to submit an essay in their first language and this was then translated into English through a web-based translation engine. The resulting English text was analyzed for grammatical errors. While pointing out that studies on MT as a pedagogical tool are lacking, Lee (2019) asked Korean EFL university learners to translate their L1 writing into L2 writing without the help of MT and then correcting their L2 writing using the MT translation for comparison. Text analysis of students' writing outcomes revealed that the MT had helped the learners decrease the number of lexico-grammatical errors and improve revisions.

Taken together, the state of literature on machine translated writing indicates that there are few studies found for the language pair between Korean and English, which may render different results from previous studies due to complex interlingual relationships between L1 and L2. Second, with the recent developments of MT, there is need for an empirical study to examine how the use of MT fares in comparison to the L2 learners' usual approaches to writing, that is, direct writing and translated writing, by analyzing the quality of the written products in a systematic manner. That is, the core question would be to seek if the MT can be a reference source for scaffolding learners to produce writing of higher quality in comparison to when the learners have to write without access to the MT. In the end, the comparative analysis on the different modes of L2 writing is expected to provide pedagogical implications on how second language learners can benefit from the use of MT. With the overall aim of researching *linguistic complexity* measures in the writing products, the following research questions guided our study:

1. How do the writing products differ on *lexical measures* (i.e., low frequency words, word information, lexical diversity) in the three modes of writing—Direct Writing (DW), Translated Writing (TW), and Machine-Translated Writing (MW)?
2. How do the writing products differ on *syntactic complexity* (i.e., left embeddedness, number of modifiers per noun phrase) in the three modes of writing?
3. How do the writing products differ on measures of *cohesion* (i.e., text easability, referential cohesion, connectives, LSA) in the three modes of writing?

3. METHODS

3.1. Participants

A total of 65 college students (Male = 28, Female = 37) learning English as a Foreign Language (EFL) in the context of South Korea participated in the present study. The participants were from two municipal areas, both majoring in English Language Teaching.

All learners were participating in courses where L2 writing was being instructed. The age of the participants ranged from 20 to 26. A majority of the learners had been learning English as a foreign language since the 3rd year of elementary school. Their general language proficiency could be regarded as being high-intermediate to advanced according to a diagnostic writing test conducted at the beginning of the semester. Information on the learners' proficiency was important to ensure that the learners would be able to conduct post-editing for the machine translated texts. In order to control for the type of learners to be included in the study, a background questionnaire was further used to collect information on the learners' age, gender, English proficiency level, and experience of staying or studying English abroad as summarized in Table 1.

TABLE 1

Learner-Participant Information

Gender	Male	28 (42.4%)
	Female	37 (56.1%)
Age	20 years old	21 (31.8%)
	22 years old	25 (37.9%)
	23 years old	7 (10.6%)
	24 years old	6 (9.19%)
	25 years old	3 (4.6%)
	26 years old	4 (6.1%)
Experience of studying in English-speaking country	Yes	11 (16.7%)
	No	55 (83.3%)
Self-perceived L2 Writing Ability	Not very good	3 (4.5%)
	Not good	3 (4.5%)
	Normal	17 (25.8%)
	Good	34 (51.5%)
	Very Good	9 (13.6%)

For proficiency, learners were asked to rate their perceived level of writing proficiency on a 5-point Likert scale, which resulted in learners reporting on a modest level of writing proficiency (M = 3.65, SD = .94). For abroad experience, the learners were asked to check 'Yes/No' and write the specific duration of having lived in an English-speaking country. This was used to rule out any outliers (i.e., +3 years of residence) that could not be considered a L2 learner. Information was also collected on their usual writing activities and experience with different modes of writing (i.e., DW, TW, MW) by asking the learners to rate a 5-point Likert scale. While Mauchly's test on the assumption of sphericity was violated ($p < .05$), one-way repeated measures ANOVA indicated that their experience with the three writing modes was different, $F(1.616, 64.646) = 47.626$, $p < .001$ with Greenhouse-Geisser estimates of sphericity. The pattern of results indicated that DW (M = 3.78, SD = 0.88) had been significantly more adopted respectively over TW (M = 2.17, SD = 0.80) ($p < .001$) and MW (M = 1.80, SD = 1.17) ($p < .001$). However, there was no significant difference in the way the learners adopted TW over MT (p > .05). The information on their background indicates that the learners found MT to be a relatively new kind of reference tool for L2 writing.

3.2. Instruments

3.2.1. Writing tasks

In this study, three writing items were developed for three writing modes, that is DW, TW, and MW. Picture description tasks (Park, Min, & Kim, 2014) were adopted as prompts for which the learners had to write narratives based on pictures with an ending to create (See Appendix A for an example). However, in the process of designing the writing tasks, it was necessary to ensure that the learners would not be subject to a practice effect. As such, three writing topics were assigned to different writing modes for the three groups of students. The configuration of three writing topics and three writing modes as illustrated in Table 2 was employed to eventually winnow out the effects of writing mode.

TABLE 2

Writing Mode and Writing Topics

	N	Writing Task 1	Writing Task 2	Writing Task 3
Group 1	20	DW	TW	MW
		Writing Topic A	Writing Topic B	Writing Topic C
Group 2	23	MW	DW	TW
		Writing Topic B	Writing Topic C	Writing Topic A
Group 3	22	TW	MW	DW
		Writing Topic C	Writing Topic A	Writing Topic B

Note: DW = Direct Writing, TW = Translated Writing, MW = Machine translated writing; Writing Topic A = Football Game, Writing Topic B = Subway Station, Writing Topic C = Theme Park

3.2.2. BNC-COCA 25000 RANGE

In order to conduct a preliminary analysis on the single word items for low frequency words in the writing products, the *Range* program with BNC/COCA lists of 25,000 words (https://www.wgtn.ac.nz/lals/about/staff/paul-nation) was used to analyze the vocabulary load of texts. While being based on a combined word list from British and American English, the program BNC-COCA 25,000 RANGE is expected to indicate how much and what vocabulary occurs in a particular text or group of texts. The development of BNC-COCA 25,000 RANGE is based on the recent compilation of Nation and Webb's (2011) vocabulary list of 25 word bands (each consisting of 1,000 word families) and an additional list of 3,000 compound nouns, proper nouns and interjections. In our utilization of the program, the focus was to analyze how the learners had been able to incorporate low frequency vocabulary (i.e., 3rd 1,000–4th 1,000 word families) versus high frequency vocabulary (i.e., 1st 1,000–2nd 1,000 word families) as an indication of sophisticated vocabulary use.

3.2.3. Coh-Metrix

Coh-Metrix 3.0 is a computational tool that produces indices for linguistic and discourse representation of a text (Graesser, McNamara, Louwerse, & Cai, 2004). The analysis tool being available at http://tool.cohmetrix.com/ was developed at the Institute for Intelligent Systems at The University of Memphis. Coh-Metrix 3.0 analyzes texts on many levels of language and discourse, such as, word concreteness, syntax, and cohesion. Coh-Metrix allows readers, writers, educators, and researchers to instantly gauge the difficulty of written text for the target audience. It is arguably the broadest and most sophisticated automated textual assessment tool currently available on the Web (McNamara, Graesser, McCarthy, & Cai, 2014). Coh-Metrix automatically provides numerous measures of evaluation at the levels of the text, the paragraph, the sentence, and the word. Coh-Metrix uses lexicons, part-of-speech classifiers, syntactic parsers, semantic analyzers, and several other components that are widely used in computational linguistics.

There have been studies published to validate the use of Coh-Metrix, for instance, to assess the cohesion of text (McNamara, Louwerse, McCarthy, & Graesser, 2010). Collectively, these studies have used Coh-Metrix to distinguish a wide range of texts. For example, Crossley, Louwerse, McCarthy, and McNamara's (2007) investigation of L2 learner texts revealed a wide variety of structural and lexical differences between texts that were adopted (or authentic) versus adapted (or simplified) for second language learning purposes.

The indices in Coh-Metrix are categorized into eleven categories: (1) Descriptive, (2) Text Easability Principal Component Scores, (3) Referential Cohesion, (4) LSA (Latent Semantic Analysis), (5) Lexical Diversity, (6) Connectives, (7) Situation Model, (8) Syntactic Complexity, (9) Syntactic Pattern Density, (10) Word Information, and (11) Readability (Total: 106 measures).

However, not all elements of language complexity were relevant within and across the specific genre (i.e., narrative) we were investigating (McNamara et al., 2014). As such, a tripartite approach was adopted for the analysis of lexical, syntactic and discourse features of learners' writing. On the one hand, there was adaptation of 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), which has been operationalized as syntactic complexity. At the level of lexis, "systemic lexical complexity" (Bulté & Housen, 2012, p. 28) was modified to incorporate Jarvis' (2013) work on lexical diversity, which Jarvis sees as consisting of rarity, volume, variability, disparity, and dispersion. The third approach was conducted to examine the discourse features of texts by analyzing multiple levels of cohesion (McNamara, Louwerse, McCarthy, & Graesser, 2010). Initially, there was calculation for descriptive indices (No. of words and sentences) for an overview of the writing products. Thereafter, the texts were analyzed for *lexical diversity* (e.g., low frequency words, type-token ratio, measure of textual lexical diversity), *syntactic complexity* (e.g., left embeddedness, number of modifiers per noun phrase), and *cohesion* (e.g., text easability, referential cohesion, connectives) in line with our frame of analysis. The indices of Coh-Metrix that were deemed most relevant to the analysis of the genre and those results that suggested statistical significance were given priority in the analysis for focused analysis. The measures utilized from Coh-Metrix are presented in Table 3.

TABLE 3

Indices of Coh-Metrix (Adapted from McNamara et al., 2014)

Indices of Coh-Metrix	Measures	Explanation of Index
1. Descriptive Indices	Number of Words, Number of Sentences	Provides basic information about the text
2. Text Easability	Word Concreteness, Verb Cohesion	Provides scores on the ease or difficulty of the linguistic characteristics of text
3. Referential cohesion	Noun Overlap, Content Word Overlap	Provides information about the overlap in content words between local sentences or co-references within the text
4. Lexical Diversity	Type-Token Ratio (TTR), Measure of Textual Lexical Diversity (MTLD), Vocabulary Diversity (VOCD)	Provides information concerning the variety of unique word types that appear in the text in relation to the overall number of words in the text.
5. Connectives	Number of Connectives, Diversity of Connectives	Provides information concerning the cohesive links between ideas and clauses within your text.
6. Syntactic Complexity	Left Embeddedness (words before main verb), Number of Modifiers (per noun phrase)	Provides information regarding the part-of-speech categories, groups of word phrases, and syntactic structures for sentences
7. Word Information	Noun Incidence, Pronoun Incidence, Familiarity for Content Words, Hypernymy for nouns and verbs	Provides information on the syntactic categories (e.g., nouns, verbs, adjectives, and adverbs) of the words used and the function (e.g., preposition, pronouns) of the words.
8. Latent Semantic Analysis (LSA)	LSA similarity between all possible pairs of sentences in a paragraph	Provides information about the semantic overlap between the sentences or paragraphs in your text.

3.3. Procedure

Three writing sessions were run consecutively, with each writing session lasting approximately 40 minutes in a computer lab. As presented in Table 2, the three groups were asked to write for different configurations of writing modes and writing topics. That is, a counterbalanced design was used to avoid the pitfalls of a repeated measure design, where none of the learners had to write for identical writing topics or writing mode.

Between the writing sessions, a 10-minute break was provided. For DW, the learners were asked to compose their texts directly on their word processors and submit an electronic version of their essays. In the case of TW, the learners were asked to write in Korean and translate into English on their word processors, and edit their essays as needed to express their target message. When submitting their writing, the learners were asked to provide their Korean essays together with their English translations. In the MW tasks, the learners were asked to write in Korean, enter their essays to the MT, and post-edit renditions of MT as needed. When submitting, the learners were asked to provide their Korean essays, the renditions of Google Translate, and post-edited versions of their machine translated essays. Dictionaries were not allowed in any of the writing tasks.

3.4. Data Analysis

The first stage of data analysis involved conducting computational analyses for the writing products collected via the three modes of writing (i.e., DW, TW, MW). With regard to examining low frequency words in research question one, BNC-COCA 25,000 RANGE was used to analyze the lexical sophistication level of the writing products for each writing mode. It was considered most relevant to analyze the vocabulary at the 3rd 1,000 and 4th 1,000 word levels since vocabulary use was not exhibited beyond the 4th 1,000 word level in high proportions. Coh-Metrix was used for analysis of other lexical measures (i.e., word information, lexical diversity) in research question one. Coh-Metrix was utilized for the second and third research questions to analyze the corpus on measures of syntactic complexity and cohesion. In the second stage of the analysis, one-way ANOVA with repeated measures was conducted in order to examine any significant effect of writing mode (i.e., DW, TW, MW) on the writing texts. When Mauchly's test on the assumption of sphericity was violated ($p < .05$) in the statistical tests, degrees of freedom were corrected using Greenhouse-Geisser estimate of sphericity.

4. RESULTS

4.1. Comparison of Three Writing Modes by Lexical Measures

4.1.1. Low frequency words

The descriptive values for 3rd 1,000 and 4th 1,000 word levels were subject to repeated measures one-way ANOVA to seek if writing mode would affect the products of writing. The results of inferential statistics with pairwise comparisons are presented in Table 4.

TABLE 4

Analysis of Single Words by Three Writing Modes

	DW		TW		MW		*F*	Post hoc
	M	*SD*	*M*	*SD*	*M*	*SD*		
No. 3rd1000 Words	1.44	1.84	1.71	1.80	2.35	1.93	5.836**	DW = TW DW < MW** TW = MW
% of 3rd 1000 Words	0.84	0.97	0.91	0.93	1.21	1.01	3.402*	DW = TW DW < MW* TW = MW
No. 4th1000 Words	0.88	1.13	2.32	2.57	1.88	1.69	9.512***	DW < TW*** DW < MW*** TW = MW
% of 4th 1000 Words	0.54	0.68	1.19	1.28	1.04	0.98	6.736**	DW < TW** DW < MW** TW = MW

Note: DW = Direct Writing, TW = Translated Writing, MW = Machine Translated Writing
**p* < .05, ***p* < .01, ****p* < .001

The number and percentage of words at the 3rd 1,000 level and 4th 1,000 level indicated that the MT as a whole had helped the learners to utilize more low frequency vocabulary in comparison to those written for DW ($p < .05$). The results as a whole indicated that the MT had facilitated the learners to retrieve target words that are relatively more lexically sophisticated (i.e., lower in frequency) in comparison to those retrieved for DW. The number and percentage of words at the 4th 1,000 word level also indicated that TW had facilitated the learners to produce words at a higher sophistication level. This may indicate that the pre-constructed text in L1 had become a platform through which the learners could more carefully attend to selecting the target vocabulary to express their messages (Cohen & Brooks-Carson, 2001).

4.1.2. Word information

Word information refers to the idea that each word is assigned a syntactic part-of-speech category. Thus, syntactic categories are segregated into content words (e.g., nouns, verbs, adjectives, adverbs) and function words (e.g., prepositions, determiners, pronouns). Many words can be assigned to multiple syntactic categories. However, Coh-Metrix assigns only one part-of-speech category to each word on the basis of its syntactic context, and computes the relative frequency of each word category by counting the number of instances of the category per 1,000 words of text, called incidence scores (McNamara et al., 2014). For this category, measures such as, *Noun Incidence*, *Pronoun Incidence*, *Familiarity* for content words, and *Hypernymy* for nouns and verbs were examined since they were most viable for analyzing the texts for writing mode difference. Results on the indices of Coh-Metrix and statistical analyses are summarized in Table 5.

Regarding *Noun Incidence*, the MW (M = 249.29) had helped the learners retrieve more content words, particularly over those retrieved via DW (p < .01). Through the use of content words, it is likely that more important information may have been projected in the MW mode. In comparison, the written texts produced by MW were found to have the least incidence of pronouns compared to DW (p < .001) and TW (p < .01). This indicates that the higher density of pronouns in the DW and TW written products can create referential cohesion problems if the reader cannot make direct connections to the referent. The results indicate that the more use of pronouns in DW and TW may cause relatively more reading difficulties for comprehension. For instance, the referent for 'we' can be considered ambiguous in the following: "We decided to go cafe near the park for changing our plan and communicate with friends about how we change plan eating food and beverage" (Essay No.13, DW).

Familiarity for content words is a rating of how familiar a word seems to an adult. For example, the words 'milk' (588), 'smile' (594), and 'pony' (524) have an average Familiarity of 569 compared to the words 'cornet' (364), 'dogma' (328), and 'manus' (113), which have an average Familiarity of 268 (McNamara et al., 2014). In the written text, the content words produced respectively in DW (M = 586.98) and TW (M = 585.60) could be regarded as being more familiar than those retrieved in MW (M = 583.50) (p < .05). This may mean that words offered by MT are those that the learners do not usually use. This accords with the previous results presented in Table 3 for 'Analysis of Single Words by Three Writing Modes' where there were more words at the 3rd 1,000 level and 4th 1,000 word level in the machine translated texts in comparison to those that were written without the reference source.

Regarding hypernymy for nouns and verbs, Coh-Metrix reports word hypernymy (i.e., word specificity) by locating each word on a hierarchical scale by allowing for the measurement of the number of subordinate words below and superordinate words above the target word. As a result, a lower value reflects an overall use of less specific words, while a higher value reflects an overall use of more specific words (McNamara et al., 2014).

As presented in Table 5, the indices indicated that MW (M = 1.74) had situated the learners to use the most specific words in comparison to when the texts had been composed via DW (M = 1.61) (p < .001). TW (M = 1.73) also exhibited the use of more specific words in comparison to DW (p < .05). This is plausible considering that the L1 texts had guided the learners to more directly tap into the meaning of the words that the

TABLE 5

Summary of Coh-Metrix Results

	DW		TW		MW		*F*	Post hoc
	M	*SD*	*M*	*SD*	*M*	*SD*		
Descriptive Indices								
Number of words	167.95	51.28	181.23	46.35	194.35	54.76	11.41***	DW < TW** DW < MW*** TW < MW*
Sentence length (No. of words)	11.47	2.99	12.56	3.32	13.87	3.34	18.727***	DW < TW** DW < MW*** TW < MW**
Text Easability								
Word Concreteness (z score)	0.49	0.86	0.88	0.91	0.67	0.96	3.81*	DW < TW* DW = MW TW = MW
Verb Cohesion (z score)	0.22	0.84	-0.18	0.79	0.24	0.80	5.78**	DW > TW** DW = MW TW < MW**
Referential Cohesion								
Noun overlap, adjacent sentences (mean)	0.25	0.15	0.31	0.19	0.32	0.19	3.75*	DW < TW* DW < MW* TW = MW
Noun overlap, all sentences	0.19	0.11	0.23	0.15	0.27	0.16	11.65***	DW < TW* DW < MW*** TW < MW*
Stem overlap, all sentences (mean)	0.21	0.11	0.26	0.15	0.29	0.16	10.51***	DW < TW** DW < MW*** TW = MW
Lexical Diversity								
TTR, content word (lemmas)	0.74	0.07	0.74	0.08	0.72	0.08	3.15*	DW = TW DW > MW* TW = MW
MTLD, all words	56.26	16.27	62.43	17.34	59.70	17.16	3.28*	DW < TW** DW = MW* TW = MW
VOCD, all words	55.08	17.75	61.77	16.63	60.27	16.86	3.64*	DW < TW** DW = MW TW = MW
Connectives								
Adversative and contrastive connectives incidence	11.38	8.52	14.02	8.69	17.72	9.33	8.99***	DW = TW DW < MW*** TW < MW*
Syntactic Complexity								
Left embeddedness	2.73	1.02	2.82	0.98	3.56	1.26	15.00***	DW = TW DW < MW*** TW < MW***
Number of modifiers per noun phrase	0.65	0.16	0.69	0.13	0.79	0.14	15.90***	DW = TW DW < MW*** TW < MW***

Word Information								
Noun incidence	233.55	41.64	247.22	49.58	249.29	38.00	3.34*	DW = TW DW < MW** TW = MW
Pronoun incidence	129.96	38.27	123.15	37.81	100.49	32.07	11.92***	DW = TW DW > MW*** TW > MW**
Familiarity for content words	586.98	5.92	585.60	4.51	583.50	5.62	6.89**	DW = TW DW > MW** TW > MW*
Hypernymy for nouns and verbs	1.61	0.26	1.73	0.32	1.74	0.27	6.13**	DW < TW* DW < MW*** TW = MW
Latent Semantic Analysis (LSA)								
LSA overlap, all sentences in paragraph (mean)	0.16	0.07	0.17	0.06	0.17	0.08	.801	N/A

Note: DW = Direct Writing, TW = Translated Writing, MW = Machine Translated Writing; TTR = Type-Token Ratio, MTLD = Measure of Textual Lexical Diversity, VOCD = Vocabulary Diversity
*p < .05, **p < .01, ***p < .001

learners would have wanted to express. All in all, the DW essays were composed of the least specific words.

4.1.3. Lexical diversity

Lexical diversity refers to the variety of unique words (tokens) that occur in a text in relation to the total number of words (tokens) (McNamara et al., 2014). Lexical diversity was calculated by Type-Token Ratio (TTR), Measure of Textual Lexical Diversity (MTLD) and Vocabulary Diversity (VOCD), as presented in Table 5. MTLD is calculated as the mean length of sequential word strings in a text that maintain a given TTR value. The index produced by VOCD is calculated through a computational procedure that fits TTR random samples with ideal TTR curves. In particular, MTLD and VOCD can overcome the potential confound of text length by using sampling and estimation methods (McCarthy & Jarvis, 2010).

Contrary to the expectation that MW would produce lexical diversity at the highest level, lower TTR evidenced for MW (M = 0.72) in comparison to that of DW (M = 0.74) (p < .05). However, caution is needed in interpreting the results of TTR since there is a lower likelihood of the different words being unique when the number of word token increases in an essay (McCarthy & Jarvis, 2010).

When MTLD and VOCD were calculated, index for MTLD was highest for TW (M = 62.43) and significantly different from that of DW (M = 56.26) (p < .01). A speculation can be made for how lexical diversity was not the highest for MW. Learners may have been incomplete in writing the pre-edited version of their L1 essays (when the learners were hard pressed for time). Also, students who were not experienced in the use of MT may have had higher expectations for what the MT could do to facilitate their writing process so that their L1 texts may have been prematurely submitted to the MT.

4.2. Comparison of Three Writing Modes by Syntactic Complexity

Before analysis of syntactic complexity in the three modes of writing, descriptive statistics were examined for an overall view of the writing products. As indicated in Table 5, the descriptive statistics and inferential statistics indicated that the learners had produced the most number of words in the order of MW, TW, and DW. For TW, it seems that the essays initially composed in L1, which became the basis for producing essays in L2, had encouraged the learners to express themselves at higher levels of fluency. Also, the MT had facilitated learners to produce longer essays when the MT took on the role to help learners express themselves. Similarly, sentence length, which may be an indicator of more fluent writing (Abdel Latif, 2012; Chenoweth & Hayes, 2001) was also the longest for the essays composed via MW.

To assess syntactic complexity, measures of *Left embeddedness* (words before main verb) and *Number of modifiers* (per noun phrase) were utilized. Left embeddedness refers to the mean number of words before the main verb of the main clause in sentences. As presented in Table 5, the indices indicated that the most number of words had been retrieved before the verb when writing via MW (M = 3.56), and the differences were significant from DW (p < .001) and TW (p < .001). This indicates that the MT may have helped the learners to retrieve more complex syntactic structures that otherwise may have not been possible with DW or TW. In a similar vein, indices also indicated that the MT had enabled the learners to produce more *Number of modifiers* per noun (M = 0.79) at a significant level (p < .001) over both DW (M = 0.65) and TW (M = .69). This again may be an indication that the MT can help learners to free their cognitive load in utilizing complex sentence structures while writing. Although the MT was not used in their study, Kobayashi and Rinnert's (1992) study demonstrates in a similar way that syntactic complexity was greater in translated writing than when composing directly in English. It seems that when learners' writing in L1 is composed as a platform, translations will be produced for improved syntactic complexity.

4.3. Comparison of Three Writing Modes by Measures of Cohesion

4.3.1. Text easability

In line with text easability being a component of overall cohesion scores (McNamara et al., 2014), indices of text easability was examined. To analyze text easability, which provides scores on the ease or difficulty of the linguistic characteristics of text, measures of *word concreteness* and *verb cohesion* were utilized. Higher values on word concreteness would indicate that the type of text contains more content words that are concrete, meaningful, and evoke mental images, which as a result are easier to process and understand (McNamara et al., 2014). Word concreteness indicated that essays written via TW included more content words relative to the other two writing modes. This makes sense since L1 texts may have helped learners focus more on retrieving L2 words that are semantically closest in meaning to the L1 words that they were using. On the other hand, in spite of the use of MT, the writing products did not indicate the level of concreteness expected from using the reference tool.

According to McNamara, Graesser, and Louwerse (2012), *verb cohesion* reflects the degree to which there are overlapping verbs in the text. This component score is likely to be more relevant for texts intended for younger readers and for narrative texts. As presented in Table 5, indices for *verb cohesion* illustrated that it was the MT that had helped the learners achieve verb cohesion, that is, by the repeated use of verbs. The recurrent use of verbs is likely to have facilitated and enhanced the comprehension level of the text. When there are repeated verbs, the text is likely to include a more coherent event structure that will facilitate and enhance understanding of the subject matter that the text is describing (McNamara et al., 2014).

4.3.2. Referential cohesion

Coh-Metrix tracks different types of word co-reference, for instance, *noun overlap* (adjacent sentences, all sentences), and *stem overlap*. Referential cohesion occurs when a noun, pronoun, or noun phrase refers to another constituent in the text. For example, in the sentence "When the intestines absorb the nutrients, the absorption is facilitated by some forms of bacteria," the word "absorption" in the second clause refers to the event associated with the verb "absorb" in the first clause. A referential cohesion gap occurs when the words in a sentence or clause do not connect to other sentences in the text. Such cohesion gaps at the textbase level increase reading time and may disrupt comprehension (Graesser, McNamara, & Kulikowich, 2011).

In particular, noun overlap (adjacent sentences) represents the average number of sentences in the text that have noun overlap from one sentence back to the previous sentence. Among the co-reference measures, it is the most strict, in the sense that the noun must match exactly, in form and plurality (McNamara et al., 2014). As presented in Table 5, the statistical results for *noun overlap* (adjacent sentences) indicated that writing with MT had produced the highest mean (M = 0.32). While noun overlap is the proportion of sentence pairs that share one or more common nouns, this indicated that MT had helped the learners achieve higher levels of cohesion between sentences over DW ($p < .05$). Index for TW indicated higher levels of cohesion over DW ($p < .05$). The lowest level of cohesion appeared for DW, which may be due to the way the writers were concentrating on retrieving word items to express themselves at the sentence level rather than at the discourse level.

In a similar vein, cohesion measures for noun overlap (all sentences) was highest for MW, which indicated significant differences from both DW ($p < .001$) and TW ($p < .05$). While a stem overlap refers to "A noun in one sentence ..... being matched with a content word (i.e., nouns, verbs, adjectives, adverbs) in a previous sentence that shares a common lemma (e.g., tree/treed; mouse/mousey; price/priced)" (McNamara et al., 2014; p. 65), indices for stem overlap confirmed that MW (M = .029) and TW (M = 0.26) had respectively improved cohesion in comparison to those texts written via DW ($p < .001$). Overall, the findings indicate that the MT is likely to facilitate learners to achieve higher levels of cohesion, which is likely to increase ease of readability.

4.3.3. Connectives

Connectives play an important role in the creation of cohesive links between ideas and clauses and provide clues about text organization (Crossley & McNamara, 2012). Coh-Metrix provides an incidence score (occurrence per 1,000 words) for all connectives as well as different types of connectives. Indices are provided on five general classes of connectives: causal (e.g., because, so), logical (e.g., and, or), adversative/contrastive (e.g., although, whereas), temporal (e.g., first, until), and additive (e.g., and, moreover) (McNamara et al., 2014). As presented in Table 5, it was only *Adversative and Contrastive Connectives Incidence* that was significantly different between writing modes ($p < .001$) where connectives appeared significantly more in MW (M = 17.72) than in DW (M = 11.38) or TW (M = 14.02) ($p < .05$). While the increased number of connectives would reflect the number of logical relations in the text that are explicitly conveyed, it seems that the MT had helped the learners to improve their instances of using adversative and contrastive connectives (e.g., *though*, *even though*, *even if*, *while*). It can possibly be predicted that Korean learners of English in the present study were not accustomed to using the types of connectives within their usual practices of writing, but the MT had assisted the learners to more readily make use of those connectives in line with their target message. According to McNamara et al. (2014), the use of connectives has often been assumed to be crucial components of higher-quality writing. That is, higher-quality writing has the quality of being more coherent and better organized.

4.3.4. Latent semantic analysis

LSA (Latent Semantic Analysis) computes the semantic similarities between words, sentences, and paragraphs (Landauer, McNamara, Dennis, & Kintsch, 2007) to provide measures of semantic overlap between sentences or between paragraphs. Coh-Metrix 3.0 provides eight LSA indices, each of these measures varies from 0 (low cohesion) to 1 (high cohesion). For instance, LSA considers semantic overlap between explicit words and words that are implicitly similar or related in meaning. For example, “home” in one sentence will have a relatively high degree of semantic overlap with words such as “house” and “table” in another sentence. However, the non-significant results for LSA as presented in Table 5 did not allow the researchers to conduct a valid analysis. We surmised that each writing product that the learners had produced had not been long enough in the number of words for the calculation of LSA to produce valid results since the learners had written a relatively short narrative based three picture strips.

5. DISCUSSION AND CONCLUSION

Despite the fact that the quality of machine translated texts are poor in comparison to human translations, the use of MT is now reaching a much wider audience than before (Kirchhoff et al., 2011; van Rensburg et al., 2012). Given the acceptance of other digital technology for teaching and learning, it seems likely that MT will become a tool students will rely on to complete their assignments in a second language (Groves & Mundt, 2015). With the potential benefits that they may offer L2 learners for second language acquisition, the purpose of the present study was to evaluate the effect of writing mode on L2 writing by conducting multilevel analyses with BNC-COCA 25,000 and Coh-Metrix. The results are expected to provide pedagogical implications on how the renditions of *Google Translate* may have a role in helping learners write in L2 possibly beyond their current level of writing proficiency.

Taken together, the results indicated that the machine translated texts, in comparison to those written by the learners, were significantly superior on a number of components. That is, the machine translated texts produced more words and longer sentences, which can be deemed indicators of improved writing fluency (Abdel Latif, 2012; Chenoweth & Hayes, 2001). The significantly increased use of lower frequency vocabulary in the MW products also adds to our findings that retrieval of productive low-frequency word families is an important repertoire of vocabulary knowledge for communicative needs. (Johnson, Acevedo & Mercado, 2016).

The MT was also found to be beneficial in improving cohesion (i.e., verb cohesion, noun overlap, stem overlap). As another means for improving cohesion, the use of connectives (i.e., adversative and contrastive connectives) was found to be superior in the MW texts relative to those produced by DW or TW. Syntactic complexity measures were also robust for MW over both TW and DW which demonstrates that the machine translated texts are characterized by complex, embedded syntax. Regarding word information, higher incidence of nouns and specific words (hypernymy) and lower incidence of pronouns and familiar words in the machine translated texts demonstrated higher writing quality particularly in relation to DW. To extrapolate, the higher index scores for noun incidence indicate that the machine translated texts are semantically more packed for meaning. The lower index for familiarity of content words illustrates that the machine translated texts consists of lexical items that are of higher sophistication levels. In a similar vein, the higher index for hypernymy also indicates that more instances of specific lexical items had been utilized in the machine translated texts.

Crossley, Weston, McClain-Sullivan, and McNamara's (2011) study demonstrates the validity of Coh-Metrix for assessing the quality of writing texts. In their study, the college students' essays were characterized as being more syntactically complex, greater in diversity of words, and more frequent in the use of rare, unfamiliar words in comparison to ninth grade essays that were characterized by higher word frequency (i.e., more familiar words) and lower syntactic complexity (i.e., simple sentences). As such, the study corroborates that Coh-Metrix is a valid tool for conducting a comparative analysis of texts produced under the three writing conditions.

In contrast, what is noteworthy in the present study is that none of the lexical diversity measures calculated by Coh-Metrix evidenced superiority of the machine translated texts. Even when measures such as MTLD and VOCD were adopted to overcome the potential confound of text length, it was the texts written via TW rather than MW that indicated higher degrees of lexical diversity. A possible explanation for this is that the learners may have been limited in their repertoire of post-editing strategies when their language proficiency was underdeveloped to notice problematic words or expressions in the machine translated texts. Learners may have met further problems when they were unable to edit the texts for authentic language use or contextually-appropriate words so that they may have reverted to the words they knew already. That is, while attending to post-editing, lexical diversity may have deteriorated. When the MT produced "Then after connecting to locate DMB broadcasting a football match", learners' post-editing led to "Then he founds the football match broadcasting after connecting to DMB" where the word 'locate' was deleted (See Appendix B for more machine translated writing).

The results of the study indicate that MT may expedite the learners' writing process to write in another language, especially for improving fluency and cohesion; and for producing syntactically complex sentences and concrete words. However, from a pedagogical perspective, there are several precautionary guidelines to be reminded of before L2 learners can be instructed to make use of MT.

First, writing continually only with the MT may deprive the learners' opportunities to retrieve second language items and acquire them. Problematic language items should be purposefully examined through focus-on form instruction (Long, 1991) either through explicit or implicit instruction (i.e., consciousness-raising activities). Second, access to MT should not be used by students to leisurely obtain a translation for the learners' L1 text. That is, before submitting a L1 text to MT, learners should be instructed to spend time to carefully construct their texts. Third, when discourse structures are different between L1 and L2, such as between Korean and English, the learners will need to be made conscious about the difference in discourse structures. That is, learners may want to use the MT more for solving their lexico-grammatical problems. For instance, the MT can be most effective for improving the use of connectives, and also in enhancing syntactic complexity. MT can also be informative in providing concrete or specific words (i.e., hypernymy). Fourth, teachers will need to train learners to be able to employ post-editing strategies since the use of MT necessitates the use of them for high-quality outcomes (Austermühl, 2011; Kirchhoff et al., 2011). Teachers will need to conduct strategy-based instruction where training can be devoted to noticing the errors or inappropriate uses of the language in the MT renditions. Austermühl (2011) discusses the issue of the quality of MT in some depth and from a number of perspectives. He described the idea that MT is not aiming for 'high quality' translation. Instead, it is aiming for 'usable' translation.

The study is not without its limitations. The writing products yielded from the three writing modes could also be analyzed through error analysis to ascertain if the instances of lexico-grammatical errors actually decrease with the use of MT. In other words, it could be empirically tested as to if MT can help L2 learners to remove the writing errors that they usually commit when having to write L2 compositions on their own. In tandem, the writing products could have been assessed by human raters based on a valid assessment rubric to examine how the writing mode may have an effect on the writing products. In the instance, the writing products may be able to highlight if human raters are in fact able to perceive the quality of MT-produced texts in any different ways from those essays that have been written without reference to a MT. Also, studies in the future that make use of different discourse structures may highlight findings that offer more discussion on the use of MT.

Another limitation and recommendation for research is that future studies should be able to focus on examining the post-editing process of attending to machine-translated texts not only to form a taxonomy of the post-editing strategies, but also to analyze how learners attend to the cross-linguistics features, that is between L1 and L2. There is also research to be conducted for scrutinizing the process of attending to these cross-linguistic differences after training has been conducted for post-editing strategies. Most of all, due to the characteristics of MT in being able to take statistics from the large language data entered by its users, future studies that analyze MT renditions may evidence improved features of writing products for Korean-English, that we would have wanted to see more of in our study.

Applicable levels: Secondary, tertiary

REFERENCES

Abdel Latif, M. M. M. (2012). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, *34*(1), 99-105.

Ali, N. H. (1996). *Translation versus direct composition: Effects of first language on second language Arab writers.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Illinois.

Austermühl, F. (2011). On clouds and crowds: Current developments in translation technology. *Translation in Transition*, *9*, 1-27.

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.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, *18*(1), 80-98.

Cohen, A. D., & Brooks-Carson, A. (2001). Research on direct versus translated writing: Students' strategies and their results. *The Modern Language Journal*, *85*(2), 169-188.

Crossley, S. A., Allen, D., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language, 23,* 84-102.

Crossley, S. A., Louwerse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, *91*(1), 15-30.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115-135.

Crossley, S. A., Weston, J., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28,* 282– 311.

Garcia, I., & Pena, M. I. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning*, *24*(5), 471-487.

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. *The Elementary School Journal*, *115*(2), 210-229.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223-234.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instrument & Computers*, *36*(2), 193-202.

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes, 37*, 112-121.

Ismail, A., & Hartono, R. (2016). Errors made in Google Translate in the Indonesian to English translations of news item texts. *ELT Forum: Journal of English Language Teaching, 5*(2), 1-6.

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.

Jewitt, C. (2012). *Technology, literacy, learning: A multimodal approach*. New York: Routledge.

Johnson, M. D., Acevedo, A., & Mercado, L. (2016). Vocabulary knowledge and vocabulary use in second language writing. *TESOL Journal*, *7*(3), 700-715.

Kirchhoff, K., Turner, A. M., Axelrod, A., & Saavedra, F. (2011). Application of statistical machine translation to public health information: A feasibility study. *Journal of the American Medical Informatics Association, 18*(4), 473-478.

Kliffer, M. D. (2005). An experiment in MT post-editing by a class of intermediate/advanced French majors. In *Proceedings of EAMT, 10th Annual Conference*. Budapest, Hungary.

Kobayashi, H., & Rinnert, C. (1992). Effects of first language on second language writing: Translation versus direct composition. *Language Learning*, *42*(2), 183-209.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

La Torre, M. D. (1999). A web-based resource to improve translation skills. *ReCALL, 11*, 41-49.

Lee, S. M. (2019). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2018.1553186

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective*. (pp. 39-52). Amsterdam: John Benjamins.

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*(2), 381-392.

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across genres and grades. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89-116). Plymouth, UK: Rowman & Littlefield Education.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292-330.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle Cengage Learning.

Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, *21*(1), 29-49.

Park, Y., Min, H., & Kim, J. (2014). *Development of writing assessment software for high school classroom in Korea* (Report No. RRE 2014-11). Seoul: Korea Institute for Curriculum and Evaluation.

Slocum, J. (1988). A survey of machine translation: Its history, current status, and future prospects. In J. Slocum (Ed.), *Machine translation systems* (pp. 1-47). Cambridge: Cambridge University Press.

Somers, H. (2003). Machine translation in the classroom. In H. Somers (Ed.), *Computers and translation: A translator's Guide* (pp. 319-340). Amsterdam: John Benjamins.

Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing*, *5*(3), 271-294.

Van Rensburg, A., Snyman, C., & Lotz, S. (2012). Applying Google Translate in a higher education environment: Translation products assessed. *Southern African Linguistics and Applied Language Studies*, *30*(4), 511-524.

Wallwork, A. (2016). *English for academic research: A guide for teachers*. Pisa, Italy: Springer.

APPENDIX A

An Example of the Writing Prompt for Writing Topic

(Adopted from Park et al., 2014)

다음 그림 1, 2, 3은 순서대로 일어난 일이다. 그림 1과 2에 나타난 상황을 각각 묘사하고, 이에 따른 그림 3의 내용을 추론하여 쓰시오. (250-300단어 내외)

1	2	3
		?

APPENDIX B

Sample Writing Produced by the Machine Translator

<u>Rendition of the Machine Translator:</u>

Saturday is the football match between South Korea and Greece at 8:00. A boy looks at a billboard advertising the fight. The boys are very fond of football. So the boy to think I enjoy this feeling to see a football match. The TV would I have failure. Already 8 seam could not even watch TV Boy is embarrassed. But soon the boy writes his head. The boy he had to eject your smart phone. Then after connecting to locate DMB broadcasting a football match. Although the screen is small, but boy is he who enjoys soccer match expectations.

<u>Post-edited version of the Machine Translator:</u>

There are a football match between South Korea and Greece at Saturday 8:00. A boy looks at a advertisement of that match at a billboard. The boy likes soccer very much. So the boy is amused at watching the match. But unexpected problem arises. The TV breaks down. The boy gets embarrassed because he cannot watch the game although it’s 8:00. But soon he uses his brain. The boy takes out his smart phone which he has. Then he founds the football match broadcasting after connecting to DMB. Although the screen is small, the boy enjoys the soccer match which he expects.