

ARTICLE



## Different effects of machine translation on L2 revisions across students' L2 writing proficiency levels

*Sangmin-Michelle Lee, Kyung Hee University*

### Abstract

*In recent years, machine translation (MT) has been gaining popularity, both in academic settings and in everyday life among foreign language students. However, insufficient research has been conducted in this field. Moreover, the findings of extant literature are often contradictory, and there are few empirical studies based on students' actual outcomes. Therefore, the present study investigates the effectiveness of using MT in English-as-a-Foreign-Language (EFL) writing classes. It particularly examines whether students' L2 writing proficiency levels influence their revisions when using MT. According to the results, using MT helped all levels of students improve their revisions, but to a different extent depending on their L2 writing proficiency levels. Compared to the higher-level students, the lower-level students made fewer changes per error, resulting in less improvement in the revised versions. Furthermore, this study found that the lowest-level students benefited the least from MT, mainly due to their limited L2 knowledge. Conversely, the higher-level students benefited more from MT by critically selecting better options between their own translations and those produced by MT. Overall, this study includes several pedagogical implications for using MT in L2 writing classrooms.*

**Keywords:** Machine Translation, L2 Revision, L2 Writing Proficiency

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

**APA Citation:** Lee, S.-M. (2022). Different effects of machine translation on L2 revisions across students' L2 writing proficiency levels. *Language Learning & Technology*, 26(1), 1–21.  
<https://hdl.handle.net/10125/73490>

### Introduction

During L2 writing, learners generally struggle with aspects such as vocabulary, grammar, content, organization, and stylistic conventions (Groves & Mundt, 2015; Tsai, 2019). Particularly, linguistic difficulties are often the most urgent issue for L2 learners, since linguistic accuracy is directly related to the overall quality of L2 writing (Min, 2006; Van Waes & Leijten, 2015). Hence, in order to compensate for their limited language proficiency, L2 learners often utilize various online resources, such as online dictionaries, grammar checkers, websites, and Google. In recent years, an increasing number of college students have been using machine translation (MT) for various academic purposes, including learning vocabulary, reading, and writing (Alhaisoni & Alhaysony, 2017; Garcia & Pena, 2011), due to its convenience, immediacy, and cost-effectiveness (Briggs, 2018). Moreover, since Google shifted from a phrase-based MT system to a neural MT (NMT) one in 2016, its accuracy has significantly improved (Le & Schuster, 2016; Tsai, 2019). Particularly, deep learning in NMT allows “the system [to] infer by itself the best representation from the data” (Poibeau, 2017, p. 184) rather than perform a word-by-word translation, and as a result, NMT received a remarkable BLEU score, a metric that determines MT accuracy (Bowker & Ciro, 2019; Koehn, 2020). It is also expected to become more accurate in the near future (Stapleton & Kin, 2019).

Previous studies have listed the benefits of using MT in L2 classrooms from various perspectives. Most

importantly, from the linguistic and cognitive perspectives, MT is a useful means of facilitating L2 learning because it highlights students' errors and helps them to notice and correct these errors, thus directing them to focus on the process of language learning (Lee, 2020; Niño, 2008). More specifically, these studies have revealed that MT raises students' linguistic awareness when conducting a post-editing task and comparing their own L2 writing with the MT outputs, which ultimately leads to reductions in the number of lexical and grammatical errors.

The process of detecting and correcting errors in their own writing or MT outputs, in turn, not only cultivates autonomous and self-directed learning (Kliffer, 2008; Lee, 2020), but it also encourages problem-solving and critical-thinking skills, since it requires students to reflect on their writing and MT outputs (Clifford et al., 2013; Niño, 2009). From an affective perspective, it reduces L2 writing anxiety, and it creates a more comfortable learning environment (Bahri & Mahadi, 2016; Niño, 2009). Consequently, the students in previous studies showed a positive attitude and greater satisfaction with regard to using MT in language classrooms (Bahri & Mahadi, 2016; Garcia & Pena, 2011; Niño, 2009).

However, the effectiveness of MT on L2 writing may vary according to learner variables. Particularly, considering that students' L2 proficiency level affects their revising strategies and writing outcomes (Cardelle & Corno, 1981; Chen et al., 2015; Lee, 2009; Qi & Lapkin, 2001), it also influences students' behaviors, attitudes, and writing outcomes while using MT. Although prior MT studies have assumed that MT may benefit students to different degrees depending on their proficiency levels (Bahri & Mahadi, 2016; Briggs, 2018; Lee, 2020; Tsai, 2019), the findings of these studies are still tentative and inconsistent. This is particularly because such studies have explored the effectiveness of using MT among single proficiency-level students without comparing different language proficiency levels. The issue of which groups of learners would benefit more from MT entails several important pedagogical concerns and will influence L2 teachers' pedagogical decisions and practices regarding the use of MT. Therefore, the present study investigates how English writing proficiency level may influence the effectiveness of using MT during English-as-a-Foreign-Language (EFL) students' revision processes.

## Literature Review

L2 writing is a difficult and complex skill in which students simultaneously face two particular tasks: a) organizing their ideas in terms of content and b) articulating their ideas in the L2. Generally, they struggle more often with the latter (Roca De Larios et al., 2002). Although adequate feedback and regular revisions can lead to substantial improvements in the L2 writing process, the opportunities are often limited in L2 writing classrooms for various reasons such as classroom constraints (Tsai, 2019). Under this situation, using MT can be an alternative to providing learners with individualized feedback with regard to their writing.

Even earlier MT studies have shown that the use of MT can facilitate L2 skills and help students enhance their L2 writing by assisting them in correcting their lexicogrammatical errors (Niño, 2008, 2009). MT can also help students deliver their intended meaning more efficiently and fluently (Garcia & Pena, 2011) and focus more on the discourse and content levels beyond the linguistic ones (Groves & Mundt, 2015). Moreover, with the introduction of NMT, the accuracy of MT has increased both in vocabulary and grammar, hence solving some of the lexical and grammatical problems of its earlier iteration (White & Heidrich, 2013). For example, according to Stapleton and Kin (2019) and Ducar and Schocket (2018), the current version of MT is not only capable of translating vocabulary more accurately and appropriately based on the context of the words used, but it also uses higher level vocabulary with greater lexical density. Furthermore, by using a large database, MT can translate colloquial language, less frequently used words, misspelled words, and less-common idioms (Ducar & Schocket, 2018). Also, regarding grammar, MT produces outputs with fewer linguistic errors than EFL students (Lee, 2020; Tsai, 2019). Accordingly, MT can help improve students' L2 writing, and L2 students consider MT outputs superior in terms of both vocabulary and grammar (Briggs, 2018; Stapleton & Kin, 2019). Tsai (2019) and Lee (2020) used MT as a CALL (computer-assisted language learning) model and found that students' L2 writing significantly

improved. In both studies, the students benefited from MT in terms of vocabulary expansion, content improvement, and reduction in the number of lexicogrammatical and mechanical errors.

Despite the potential benefits of MT in L2 writing classrooms, language teachers are still reluctant to utilize MT in their classrooms for a few reasons. The first reason is the historical inaccuracies of MT, which have been mostly caused by statistical machine translation, the predecessor of NMT. As mentioned earlier, recent studies have shown that MT accuracy has significantly improved. Hence, as the lexicogrammatical accuracy of MT exceeds that of intermediate EFL students, it can respond to students' queries and aid their L2 writing (Briggs, 2018; Lee, 2020; Stapleton & Kin, 2019). The second reason stems from L2 teachers' negative views toward translation, such as that translation is unnatural and ineffective, produces interference, or prevents students from communicating and thinking in the foreign language (Cohen & Brooks-Carson, 2001; Howatt, 2004; Malmkjær, 1998). However, according to Kim and Yoon (2014), translation tasks can bring about benefits in L2 writing classrooms depending on the involved language proficiencies, task types, and teacher interventions. Cook (2010) also viewed translation as pedagogically effective and educationally desirable because it can develop L2 awareness and use. Moreover, studies found that translation is the most frequently used method among Asian EFL students during L2 writing (Kim, 2011; Kim & Yoon, 2014; Wang & Wen, 2002) and that translation, including mental translation, naturally occurs during L2 writing (Liu, 2009). Kim (2011) further revealed that Korean EFL students often cannot utilize the L2 knowledge that they learned during high school in their L2 writing, and thus, translation tasks can help them activate their L2 knowledge and use it in their writing because they learned English mostly through the grammar-translation method during high school. In addition, L1 use helps students generate content, plan and monitor their writing process, search for better vocabulary, use compensatory strategies, and reduce their cognitive load during L2 writing, and thus enhances the quality of their L2 writing (Briggs, 2018; Choi & Lee, 2006; Kim & Yoon, 2014; Wang & Wen, 2002). While translation is indispensable in L2 learning (Butzkamm & Caldwell, 2009), as Cook (2010) emphasized, it should function as a means instead of an end in L2 classrooms. With the introduction of MT, translation has been reinstated in L2 writing classrooms (Briggs, 2018; Lee, 2020; O'Brien et al., 2018). For instance, Goulet et al. (2017) and O'Brien et al. (2018) recommended the maximized use of translation while using MT, that is, writing in the L1 first, machine-translating the text, and self-post-editing to maximize efficiency and reduce cognitive load.

Although the majority of previous studies have confirmed that MT can be helpful for L2 writing, they seldom considered learner variables, such as language proficiencies, strategies, or attitudes, when investigating MT. Moreover, concerning students' language proficiencies in the effectiveness of MT in L2 writing, previous studies have presented inconsistent results. On the one hand, a group of MT scholars have claimed that MT is more helpful for beginners and low- and intermediate-level students (Garcia & Pena, 2011; Kliffer, 2008) because "the cognitive demands associated with English production can be overwhelming to students of lower proficiency levels" (Briggs, 2018, p. 17). MT can reduce the cognitive load for them, and as a result, students write more fluently with less effort. On the other hand, Bahri and Mahadi (2016) and Tsai (2019) argued that it might not be appropriate to incorporate MT into beginner or low-level EFL classrooms. Lee (2020) speculated that lower-level students benefit less from MT due to their lack of L2 knowledge, confidence, and motivation. According to Valijärvi and Tarsoly (2019), beginner-level students rely more frequently on MT, but advanced-level students benefit more from it. In addition, several studies have maintained that MT grammatical accuracy exceeds that of intermediate EFL students and students view MT as superior to them (Briggs, 2018; Lee, 2020; Stapleton & Kin, 2019; Tsai, 2019). However, this view can be subjective and relative depending on their L2 proficiencies. In other words, students with different L2 proficiencies "may have different understanding of the errors produced by the translation program, or may interact differently" with MT (White & Heidrich, 2013, p. 243), and as a result, the effectiveness of MT can vary depending on students' L2 proficiency levels. This issue will, in turn, affect L2 teachers' decisions regarding whether and how to use MT in their classrooms, as well as students' perceptions and behaviors toward using MT. Despite its significance, none of these studies have actually compared different language proficiency levels and evaluated the effectiveness of MT according to such levels.

Another issue with previous MT studies lies in their scarcity of student outcomes, since they generally have focused on students' perceptions of using MT. Therefore, it is necessary to investigate the relationship between the effectiveness of using MT in L2 revisions and students' L2 writing proficiency based on their revision outcomes. In order to achieve this, the present study addresses the following research questions:

1. Did the number of changes in size, type, and function in the students' revisions with MT differ according to their L2 writing proficiency levels?
2. Did the students have different degrees of benefits from MT during revision, depending on their L2 writing proficiency levels?

## Method

### Research Procedure

In this study, the participants consisted of 80 students (male = 46; female = 34) enrolled in a general English class, which focused on reading and writing, at a four-year university in Korea. The task involved multiple stages of writing, and each student produced four versions: a) the L1 writing, b) their own L2 translation, c) the MT translation, and d) a revised version of their L2 translation. First, the students wrote essays (approximately half a page) about technology and human society in Korean. Second, they translated their essays into English on their own. Third, they used Google Translate to translate the essays. Finally, they compared their own L2 versions with the MT version and revised their L2 versions. This task was presented as a homework assignment and was not included in the final grade. The students were given one week to submit all four versions. The use of dictionaries or other resources was also allowed during the writing and revision process.

The current task was included in the curriculum for the following reasons. First, as prior studies have claimed that translation tasks facilitate L2 writing (Kim & Yoon, 2014; Liu, 2009) and particularly help Korean EFL students improve their L2 writing by activating their L2 knowledge (Kim, 2011), the task was expected to help the students connect their L2 knowledge with their L2 writing. Second, an increasing number of students are using MT for academic purposes, with more students translating larger passages of text than ever before (Murtisari et al., 2019), and the students in this study also showed a similar pattern. A brief survey conducted prior to the task revealed that approximately one-third of them (31.7%) wrote in their L1 and translated an entire paragraph or text into the L2 using MT and then post-edited the text. Therefore, this method allowed the author to investigate one of the popular ways in which L2 students used MT. In addition, this method allowed the author to compare the L1 writings, the students' L2 translations, the MT translations, and the revised versions, with specific focus on tracing the revision process and detecting the sources of the errors.

### Data Analysis

The present study employed mixed methods for a more robust and in-depth analysis of the data. The primary data sources included the students' original (their own L2 translation) and revised versions. For the analysis of these versions, multiple steps of quantitative analysis were employed. During the first step, two trained raters assessed the students' L2 translations and the revised versions based on a holistic six-point scale adapted from the TOEFL writing rubric (see [Appendix A](#)). Then, according to the scores of their original L2 texts, the students were categorized into five groups, ranging from Group 2 (the lowest level) to Group 6 (the highest level).

During the next step of analyzing errors and changes in the samples, the present study, which adopted Lee's (2020) framework, categorized errors into a) mechanics, b) vocabulary, or c) grammar. This study further classified grammatical errors into global and local errors because their significances can widely differ in L2 writing. Global errors refer to those that seriously impede intended meaning, while local errors refer to those that "although potentially irritating to an audience unfamiliar with multilingual writing," do not impede meaning (Kurzer, 2018, p. 15). Based on Kurzer's (2018) classification, fragments, run-ons, verb

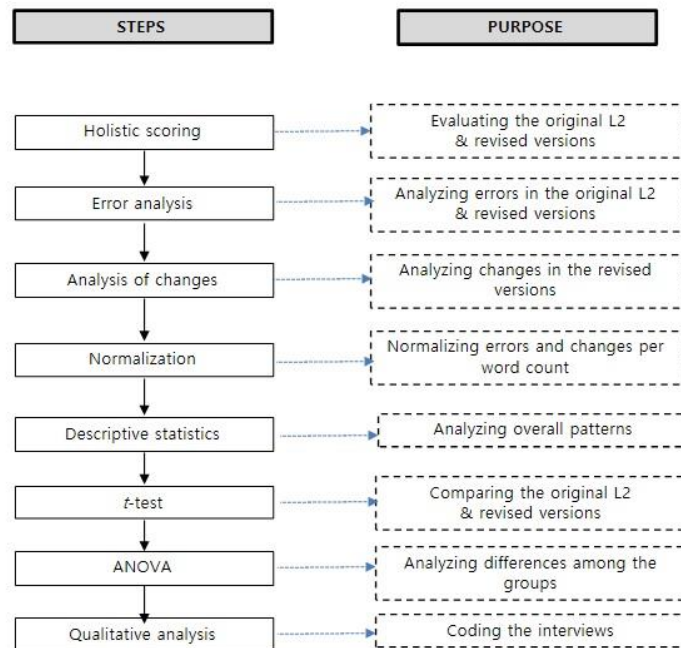
tenses, and sentence structures were classified as global grammatical errors, while the rest of the errors, involving, for instance, articles, prepositions, and word forms, were categorized as local errors. All of the errors and changes were normalized by dividing by the number of words for a clearer comparison among the groups. Finally, the descriptive statistics regarding the students' own L2 translations and their revised versions were calculated, after which *t*-tests were conducted to determine whether there were significant differences between the two versions.

In the third step, the raters tracked all of the changes made in the revised versions and categorized them in terms of the size, type, and function of the revisions, based on the framework adapted from Min's (2006) study (see [Appendix B](#) for a sample). In this study, size refers to linguistic units of change such as symbols, words, phrases, or clauses/sentences. A paragraph unit was not included in this study because replacing an entire paragraph with the MT outputs was not allowed in the current task. Type of revision refers to either surface (no change in the overall meaning of the original sentence), micro-text-based (changes affecting the meaning of a sentence but not altering the summary of the text), or macro-text-based (changes altering the overall summary of the text) changes. Each category includes the subcategories of Addition, Deletion, Substitution, Permutation (rephrasing information), Distribution (rewriting the same information in larger chunks), Consolidation (putting separate information together), and Reordering. Function refers to the reasons for making each change, including Mechanics, Word Choice, Grammar, Cosmetic (meaning that it preserves meaning and has the sole function of making the text look better), Texture (making the text more cohesive and coherent), Unnecessary Expression, and Explicature (making the text more explicit) (See [Appendix B](#)).

In the fourth and final step, an analysis of variance (ANOVA) was employed to determine whether there were any differences in the numbers of errors and changes, depending on the students' L2 writing proficiency levels. Overall, the interrater reliability values for scoring the two versions were 0.90 for the students' L2 translations and 0.91 for the revised versions.

**Figure 1**

*Data Analysis Procedure*



The quantitative analysis of the students' writings showed the overall patterns of their revisions as well as the patterns idiosyncratic to the different L2 writing proficiency groups. However, the analysis did not

explain the underlying reasons for the students' revising behaviors. Thus, in order to complement the quantitative data sources, retrospective interviews were conducted. Based on the analysis of both writing versions, 13 students were selected. Among them, six were selected because their revision patterns appeared unusual (e.g., they either made a significant number of changes or rarely made a change, or they either greatly improved or did not improve in the revised versions). The rest of the students were randomly selected from among the volunteers.

Overall, the interviews were free-structured in form and took approximately 20 min per student. During the interviews, the students freely discussed their experiences and perceptions about using MT during writing as well as the benefits and difficulties of using MT during the task (see [Appendix C](#) for the interview questions). Each interview was recorded, transcribed, and thematically coded by the same raters who evaluated the students' writings. In addition, the students' remarks were coded as positive or negative responses, after which the results were compared across the groups. Since the categories were clear, the intercoder reliability was high (96%). The data analysis procedure is summarized in [Figure 1](#).

## Results

### The Students' Writing Outcomes

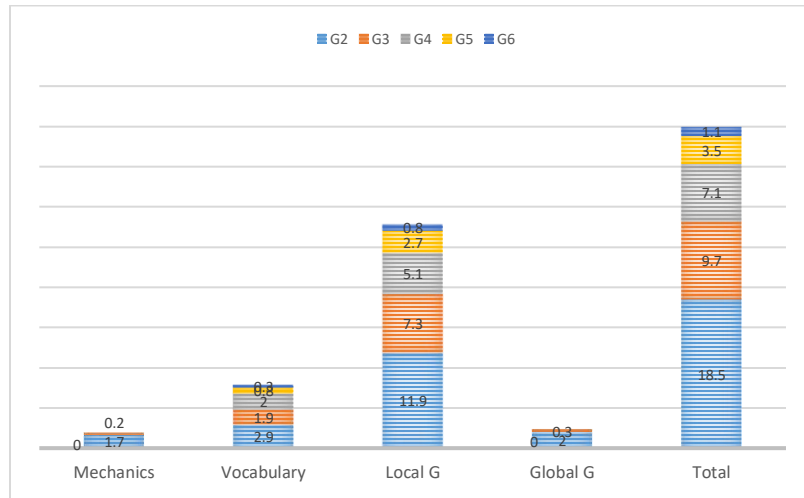
With regard to the students' own L2 translations (original versions), 14 students scored 2 (Group 2), 18 students scored 3 (Group 3), 22 students scored 4 (Group 4), 18 students scored 5 (Group 5), and eight students scored 6 (Group 6). The average total word count was 177 words, ranging from 89 to 385. Group 6 produced the longest essays (mean = 241), whereas Group 2 produced the shortest (mean = 161). The average number of total errors was 8.4. The students made local grammatical errors most frequently (mean = 5.8), followed by lexical errors (mean = 1.7), global grammatical errors (mean = 0.6), and mechanical errors (mean = 0.3) relative to the total word count. Unsurprisingly, as shown in [Figure 2](#), Group 2 scored the highest in the means of total errors as well as in the means of mechanical, lexical, and grammatical errors, whereas Group 6 scored the lowest in all of these aspects, with none in the global grammatical error category. Particularly, the global grammatical errors frequently appeared in the writings of Group 2 (39 cases among 11 students), less frequently in Group 3 (nine cases among four students), and none in the writings of the higher-level groups. On average, Group 2 made 18.5 errors, Group 3 made 9.7, Group 4 made 7.1, Group 5 made 3.5, and Group 6 made 1.1 relative to the total word count. As shown in [Table 1](#), the results of the ANOVA confirmed that the differences among the groups were statistically significant in all of the categories ( $p < .001$ ).

**Table 1**

*Analysis of the Differences in Errors Among the Groups (ANOVA)*

	Sum of Squares	Mean Square	<i>F</i>	<i>p</i>
Mechanics	35.368	8.842	8.935	< .001
Vocabulary	50.276	12.569	8.139	< .001
Local grammar	936.755	234.189	27.953	< .001
Global grammar	45.060	11.265	15.313	< .001
Total errors	2345.070	586.267	38.375	< .001

*Note.*  $N = 80$

**Figure 2***Group Means of Errors in the Original Versions*

Note. Local G = Local Grammar; Global G = Global Grammar

The results indicated that the majority of the students actively and frequently made changes. On average, the students made 6.7 changes in the revised versions. For size, more than half of the changes were made to words, followed by phrases and clauses/sentences. For type, Substitution and Addition were the most frequently used subcategories both in surface and micro-text-based changes. Surface changes occurred approximately twice as frequently as micro-text-based changes. There were no macro-text-based changes found. The students made changes the most frequently for the functions of Grammar and Cosmetic and the least frequently for Mechanics and Unnecessary Expression (Table 2).

**Table 2***Changes in Size, Type, and Function*

Size	Type		Surface		Micro		Function	
	N(%)		N(%)	N(%)	N(%)		N(%)	
Symbol	9(1.0)	Addition	192(20.3)	31(3.3)	Mechanics	9(1.0)		
Word	499(52.7)	Deletion	71(7.5)	4(0.4)	Word Choice	112(11.8)		
Phrase	288(7.3)	Substitution	291(30.4)	83(8.8)	Grammar	311(32.8)		
Clause / Sentence	151(15.9)	Permutation	186(19.6)	19(2.0)	Cosmetic	340(35.9)		
		Distribution	8(0.8)	7(0.7)	Texture	40(4.2)		
		Consolidation	10(1.1)	3(0.3)	Unnecessary Expression	22(2.3)		
		Reordering	28(3.0)	14(1.5)	Explicature	105(11.1)		

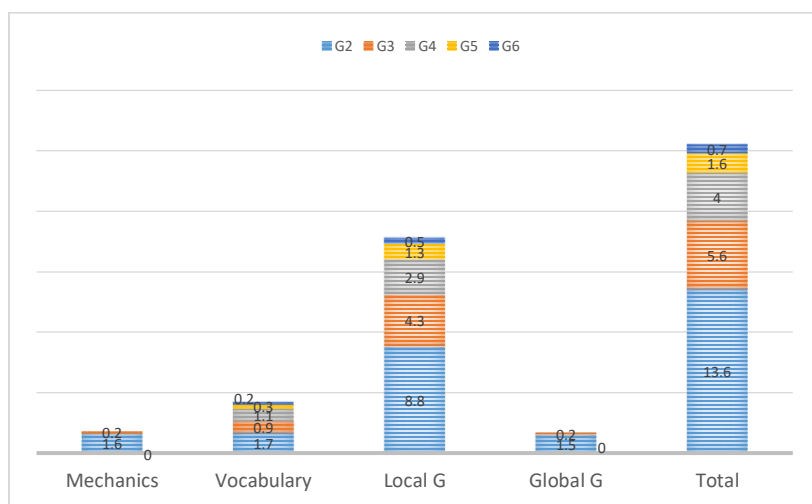
Note.  $N = 80$

The students made a significant number of changes during revision, which enhanced scores in the revised versions. In this regard, five students scored 2, 13 students scored 3, 21 students scored 4, 21 students scored 5, and 20 students scored 6 in the revised versions. Compared with the scores of the first versions, the scores in the revised versions improved by approximately 22% on average (Group 6 excluded). More specifically, both lexical (mean = 0.9) and local grammatical (mean = 3.3) errors were reduced in the revised versions

(Figure 3). This indicates that, in the final versions, vocabulary and grammar improved by 46.0% and 44.1%, respectively. Furthermore, the results of the *t*-test showed that the differences between the two versions were statistically meaningful, except for the mechanical errors (Table 3).

**Figure 3**

*Group Means of Errors in the Revised Versions*



Note. Local G = Local Grammar; Global G = Global Grammar

**Table 3**

*The Results of the Paired t-Test*

Measured Area	Original Version		Revised Version		<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Mechanics	0.3	1.177	0.29	1.105	1.000	.320
Vocabulary	1.74	1.450	0.94	1.036	6.374	<.001
Local grammar	5.86	4.450	3.27	3.511	8.867	<.001
Global grammar	0.60	1.420	0.45	1.101	2.101	.039
Total errors	8.40	6.647	5.36	5.312	8.877	<.001
Average score	3.85	1.244	4.47	1.211	-9.267	<.001

Note. *N* = 80

The results further showed that the lower-level students (Group 2 & 3) made the fewest corrections per error, whereas Group 6 made the most changes in the revised versions. The ratio of changes to errors for Group 6 was 1.519, followed by 1.410 for Group 5, 1.001 for Group 4, 0.773 for Group 3, and 0.426 for Group 2. These results imply that while the lower-level groups made errors more frequently, they did not make changes as often. Although there were differences in the *t*-test results, many of the global grammatical errors committed by Groups 2 and 3 remained unresolved in the revised versions. For example, out of the 48 global grammatical errors, only 12 errors were corrected. In some cases, the students did not replace their outputs even when the MT provided better options, as shown in the following example:

...it is certain that using Korea language perfectly\* as Korean but if adolescent use abbreviatton\* too long time, they will be poor at mother language. Finaily\* they will be adult,they\* get accustomed to abbreviation so that they are poor at standard language and the spelling system of Hangul. (WH from Group 2, original version)



Koreans should be good at Korean, but if they use abbreviations for a long time, they will become accustomed to abbreviations. So, when they become adults, they may not be able to use the standard language and spelling well. (MT version)

...it is certain that using Korea language perfectly\* as Korean but if adolescent use abbreviatton\* for a long time, they will be poor at mother language. Finaily\* they will become adult, they get accustomed to abbreviation so that they may not be able to use standard language and the spelling system of Hangul. (Revised version)

Moreover, new errors were found in the revised versions of Groups 2 and 3. Overall, six students in Group 2 made 12 new grammatical errors and two lexical errors (i.e., 14 errors in total), three students in Group 3 made four grammatical errors in the revised versions, while no additional errors were found in the revisions of the higher-level groups. As a result, Group 2 improved the least (the mean average score of the revised versions improved 28.8% compared to that of the original versions), while the other groups showed greater improvement in the revised versions: Group 3 with 44.0%, Group 4 with 46.9%, Group 5 with 47.3%, and Group 6 with 41.7%. [Table 4](#) presents examples of the changes.

**Table 4**

*Examples of the Changes*

Category	Original Version	Revised Version
Mechanics	Although we use letters* original idea	Although we use letters, (corrected) <u>unipue</u> idea (newly added/incorrect)
	communication disorder*	communication breakdown (corrected)
Vocabulary	Slang and abbreviations	Slang and <u>profanity</u> (newly added/incorrect)
Global grammar	...it can be confusing for foreigners who are studying children or Korean*	...it can be confusing for foreigners or children who are studying Korean (corrected)
Local grammar	...used by young generation*	...used by the young generation (corrected)
	...this rule...	...this rules... (newly added/incorrect)

The results of the ANOVA confirmed that there were no significant differences among the groups except in Substitution and Reordering in micro-text-based changes and Word Choice and Texture in relation to function ([Table 5](#), see [Appendices D, E, and F](#)). In addition, the *t*-test results showed that the lower-level groups rarely corrected mechanical errors ([Table 2](#)).

**Table 5**

*Analysis of the Differences in Changes Among the Groups (ANOVA)*

Size	Type		(Surface)		(Micro)		Function			
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>		
Symbol	2.148	.083	Addition	2.054	.095	3.480	.12	Mechanics	2.147	.830
Word	0.785	.538	Deletion	2.118	.087	1.404	.241	Word Choice	5.468	.010
Phrase	0.212	.931	Substitution	0.749	.562	10.625	<.001	Grammar	2.109	.080
Clause / Sentence	1.251	.297	Permutation	1.332	.266	1.252	.296	Cosmetic	1.478	.217
			Distribution	0.934	.449	2.365	.060	Texture	5.260	<.001
			Consolidation	2.139	.084	.880	.480	Unnecessary Expression	0.293	.882
			Reordering	1.883	.122	21.656	<.001	Explicature	2.485	.051

*Note.*  $N = 80$

### Analysis of the Retrospective Interviews

Three students each from Groups 2, 3, and 6, and two each from Groups 4 and 5 (13 in total) participated in the retrospective interviews. Regardless of the group, the students stated that MT was helpful in selecting vocabulary, correcting grammar in some instances, and making replacements to convey better and more authentic expressions. They also listed instant feedback and convenience as some of the advantages of using MT. Moreover, they remarked that using MT was efficient because they “could receive individualized feedback like peer feedback” and it “enabled/helped self-editing.” While they generally agreed that MT was useful in facilitating their revisions, they also stated that they “could not and should not entirely trust it,” mainly because of the grammatical errors in the MT outputs.

However, there were a few disparities among the groups regarding their perceptions of MT. One student from Group 2 simply stated that MT was “useless because it had errors and [they] did not know which one [they] should take.” In contrast, another student from Group 2 mentioned that MT was immensely helpful to him and he did not have any complaints. Meanwhile, four students from Groups 2 and 3 stated that the MT translations, despite having some noticeable errors, were superior to their versions. They explained that the MT translations “were easier to understand” than their translations and that they “sounded more natural and authentic, like real English.” They also talked about their struggles with making revisions using MT. Because they were aware that neither MT nor their work was perfectly accurate, they were not certain as to which would be better. Furthermore, since many parts of the MT translations were considerably different from the students’ own translations, “it was very challenging to compare two versions sentence by sentence, correct the errors gradually, and incorporate the changes.” They particularly faced challenges in making changes at the clause/sentence level, stating “I was not sure how much I was allowed to change following the MT version” and “I was worried about whether the writing would still sound like mine after revision.”

Conversely, the students from Group 4 and above were aware that MT could be incorrect, but they were more capable of detecting and correcting their own errors, as shown in the following quote:

The machine translation helped me notice some things. It helped me when I did not really know how to write certain things or I had wrong sentences. However, you must be careful with it, since it does not always translate correctly. It sometimes makes awkward literal translations and cannot translate specific expressions. When you’re not sure, you need to double-check with other resources. (MJ, Group 5)

Although the students from Group 4 and above did not feel that MT generally produced better quality versions than they did, they still found some expressions that were “more articulate and precise than [theirs].” When they found differences, they could evaluate “which options were better for the context” and “sounded more authentic” and then decide whether to replace them. In addition, it is worth noting that the students

from Group 6 did not simply limit MT to fixing errors, but they related the usage of MT to language learning. They said that using MT “helped improve [their] English ability in general,” “helped [them] learn new expressions and vocabulary,” and “this individual learning was a good way to gain language skills.” In sum, during the interviews, the lower-level groups exhibited more struggles with detecting errors and revising them due to their limited language proficiency than the higher-level groups.

## Discussion

The first research question examined whether there was a difference in the number of changes depending on the students’ proficiency levels in writing. In this study, the lower-level students produced more errors in all of the mechanical, lexical, and grammatical categories. However, the number of changes did not correlate with the number of errors. The results also revealed that the higher-level groups made changes more actively than the lower-level groups in terms of the ratio of changes to errors. In this study, the students’ errors were not only quantitatively different, but they were also qualitatively different according to their L2 writing proficiency levels. More specifically, while the lower-level groups corrected their lexical and local grammatical errors, many of the global grammatical errors remained unresolved, even when better options were provided by MT.

A few explanations for this behavior are possible. First, although some students stated during the interviews that the MT outputs were more accurate and more authentic than theirs, the interviews also indicated that the lower-level groups had difficulty evaluating their own erroneous sentences and those of MT. Thus, they were unable to determine which one better represented their intended meaning in the L2. Briggs (2018) argued that the ability to notice errors depends on the students’ degree of self-perception and confidence in L2 writing. Bitchener and Knoch (2010) also contended that the extent to which students decide to act on feedback can differ according to their language proficiency levels. They explained that advanced writers have a greater capacity for attention and that they correct their errors more effectively than lower-level students. Similarly, in the present study, the students’ L2 proficiency and their confidence in writing seemed to influence their ability to notice and correct errors. When there was a large discrepancy between the MT interpretation and their version, it became more difficult to compare individual sentences and correct errors. In addition, the global grammatical errors mostly required changes at the clause/sentence level, and the lower-level students’ reluctance to make changes at this level may have affected their decisions not to replace them. In fact, options for replacing entire sentences seemed to place the lower-level students in a dilemma: since they perceived the MT outputs to be superior to their own versions, frequent changes for a sentence would have resulted in them replacing most of their writing. Thus, ironically, this situation forced them to be somewhat reluctant to make changes at the clause/sentence level, as stated by the students during the interviews. On the other hand, since the higher-level students were more capable of making judgments about which alternative was better in a given context, they seemed to be more confident about replacing larger chunks in their revisions.

Another possible explanation can be found in the students’ L1 writings. As Lee (2020) ascertained, and as voiced by a student in the present study, the students’ L1 writings (source texts) strongly affected the MT outputs and, subsequently, their revised versions. After tracing all of the global grammatical errors that remained unresolved in the revised versions back to their L1 writings and MT versions, this study found that approximately 46% of them were incorrectly translated in the MT versions and the sources of errors in the MT versions were often discovered in their L1 writings. For example, the equivalent Korean sentences in the L1 writings contained errors such as missing periods, missing subjects/objects (common in the Korean language), and/or run-on sentences. This definitely caused inaccurate translations by MT, after which the erroneous sentences in the MT versions further confused the students. These errors in the L1 versions frequently caused syntactic errors in the MT outputs, and as Niño (2008) argued, correcting syntactic or sentence structure errors seemed to be more challenging for the lower-level students. While the higher-level students were able to discard erroneous sentences, the lower-level students were not.

The second research question examined whether the students benefited from using MT in the revision

process and whether the degree of benefits varied depending on the students' L2 writing proficiency levels. The results of this study indicated that using MT helped the students, regardless of their proficiency levels, and improved their revisions in all of the investigated areas, except for reducing mechanical errors. The students, in fact, often made changes to grammar (32.8%) and vocabulary (11.8%). As a result, the total number of errors and the number of lexical and grammatical errors were greatly reduced in the revised versions, and the revised versions were 42.1% better in terms of reductions in the total number of errors. The *t*-test results confirmed that the differences in the number of lexical and grammatical errors between the original and revised versions were statistically meaningful. Previous studies have indicated that lexicogrammatical errors often diminish the quality of L2 writing (Lee, 2009, 2020; Min, 2006), and likewise, the overall quality of the students' revised versions in the present study improved through the decrease in the number of lexicogrammatical errors. However, many of the global grammatical errors in Group 2 remained unresolved, and the students in this group made the least number of changes per error. Moreover, new errors were introduced in the revisions of Groups 2 and 3. Consequently, the rate of improvement in their writing was lowest in Group 2, in contrast to the other groups which showed greater improvement.

Since MT did not provide one-to-one direct feedback regarding errors in the students' original writings, MT-assisted revision required a complex process of noticing, comparing, checking, and making decisions that was too overwhelming for the lower-level students. The results of this study are consistent with previous L2 studies on feedback and revision strategies that found that students with high proficiency levels produced better-quality revisions (Cardelle & Corno, 1981; Lee, 2009; Qi & Lapkin, 2001). Since previous studies have pointed out that knowledge about L2 comes into play during revisions (Ferris & Roberts, 2001; Lee, 2009), the higher-level groups had more advantages from using MT in the present study. As Barkaoui (2007) claimed, the more proficient students in this study responded to the MT outputs more confidently and critically, and they produced better-quality revisions "by changing what needs to be changed and preserving what can be preserved" (Barkaoui, 2007, p. 82).

In addition, the students in this study exhibited different patterns in a few subcategories depending on their proficiencies. For type, the lower-level students made changes more frequently than the higher-level students in Substitution of micro-text-based changes (see [Appendix E](#)). These students made more lexical errors in the original text, and with the help of MT, they corrected those errors. Lexical errors are more local and are thus easier to compare with MT outputs and to correct than syntactic or sentence structure errors (Niño, 2008); thus, the lower-level students were able to substitute them with other options. This result was congruent with the findings that these students also made changes more frequently for Word Choice. On the other hand, while Group 6 frequently made changes for Texture, the lower-level groups rarely made such changes. This indicated that changes related to Texture were challenging for the lower-level students, even when MT provided options. Lastly, the difference found in relation to Reordering was ascribed to the fact that the higher-level students seldom made sentence structure errors in the original text and thus rarely made changes related to this subcategory.

Based on the aforementioned results, this study includes several pedagogical implications for using MT in L2 writing assignments. First, it would be beneficial to teach writing strategies using MT. Chen et al. (2015) revealed that merely using technology cannot bring educational benefits to all students. In their study, while some students who were less proficient but more motivated and more cautious about revising with the paraphrase-supporting tool benefited from using technology in the paraphrasing task, the rest did not indicate sufficient awareness of grammar rules in technology-supported L2 writing. In the present study, while the students in Groups 5 and 6 were particularly aware of MT's inaccuracies, they still benefited from MT by selectively adopting options. They also mentioned in the interviews that they could further increase their L2 skills (beyond merely fixing the errors) by using MT. Conversely, the students in Group 2, due to their limited capacity in the L2, did not benefit as much as those in Group 6. These results indicate that the use of MT might further increase the gap between higher- and lower-level students. The results of the current study and Chen et al.'s (2015) study imply that teachers should consider students' differences when using a technological tool in writing. Thus, proper training and different MT-use strategies, based on the

students' L2 writing proficiency levels, can ensure the effectiveness of MT.

Second, as Jolley and Maimone (2015) noted, the quality of students' source texts (L1 writing) is indispensable to achieving high-quality results in the use of MT. Similarly, this study also revealed that the students' L1 writing proficiency levels held relevance when using MT because they significantly influenced the MT versions. Therefore, pre-editing or extra care in L1 writing may further benefit L2 learners when using MT in L2 writing.

Lee (2020) argued that MT is similar to peer feedback in that neither MT nor peer feedback is perfect, yet both are beneficial to a certain degree. However, unlike teacher or peer feedback in which criticism is usually more specific about individual errors, MT merely provides an alternative version of writing. Thus, feedback is often neither specific nor targeted to individual errors. In other words, in the MT versions, some errors are directly replaced, while other errors are hidden in a different formulation of a particular sentence. From this perspective, the usefulness of revisions with MT is somewhere between the usefulness of peer feedback and self-editing and between direct and indirect feedback. In this study, MT sometimes overwhelmed the lower-level students because correcting errors by comparing two very different versions was often beyond their capacity. Self-editing demands a significant amount of attentional, cognitive, and linguistic resources (Ferris & Roberts, 2001; Waller & Papi, 2017), which the lower-level students seemed to lack. Hence, along with using MT, teachers need to implement diverse types of feedback to complement MT. For instance, facilitating writing conferences with students after revising with MT can be helpful in reducing their uncertainties regarding linguistic items. As direct correction is more effective for lower-level students (Chandler, 2003; Cheng et al., 2015; Saito & Fujita, 2004), teachers should provide further direct explanations about the rules through writing conferences, which can reduce students' uncertainties regarding linguistic items and prevent them from making passive or superficial error corrections when using MT. By doing so, teachers can offer positive feedback to students and motivate them to write, which are also features lacking in MT.

## Conclusion

The present study showed that, although using MT greatly helped the students' L2 revisions by reducing the number of lexicogrammatical errors, the degrees of pedagogical benefits from using MT can vary depending on the students' L2 writing proficiency levels. Language proficiency, as well as other individual characteristics, shapes students' opinions regarding feedback, and as a result, it has extensive effects during the writing and revision process (Cardelle & Corno, 1981; Lee, 2009; Qi & Lapkin, 2001). Similarly, while the higher-level students in the present study were capable of evaluating the MT versions and critically selecting the appropriate options, the lower-level students were less capable and thus less confident. In particular, this study revealed the lack of changes made among the lower-level students, their lower rates of improvement, and the introduction of new errors in their revised versions. A possible reason for this is that the lower-level students may have struggled to determine what to do with the two versions. In the worst case, their corrections may have simply been a copy of the MT versions or based on random guesses, which does not cultivate language learning or the ability to make conscious error corrections. Consequently, they might make the same errors in the future. Truscott (2007) contended that the findings of L2 writing studies on corrective feedback have often overestimated students' abilities in error corrections, and furthermore, their corrections during revisions should not be considered the culmination of language learning. Hence, merely allowing students to use MT does not benefit L2 writing classrooms, and further teacher guidance and effective pedagogical design should be implemented alongside the consideration of students' needs.

The present study includes several limitations. Because the task was a home assignment, the process was not controlled; thus, unexpected variables may have intervened. Also, given that MT quality varies depending on the topic, the topic of the task may have influenced the result of the study. In this study, only a small number of participants were interviewed. Since each student had their own idiosyncratic characteristics and cognitive styles, further extensive interviews or other qualitative data resources can be helpful to better understand their writing or revision process with MT, thus drawing more teaching

implications. In addition, this was a short-term experimental study. However, since other studies have mentioned that students' error corrections do not guarantee language learning (Chandler, 2003; Truscott, 2007), an analysis of delayed subsequent writing tasks is required in order to obtain a more precise measure of their language learning after using MT. Similarly, more longitudinal studies should be conducted to examine the far-reaching effects of MT on students' L2 learning. Finally, this study employed a specific, controlled task procedure of using MT; however, future research should employ more diverse approaches.

Despite some deficiencies in the current MT iteration, many studies have found it to be a useful pedagogical tool that can benefit students' language learning, particularly with regard to L2 writing (Chandra & Yuyun, 2018; Lee, 2020; Tsai, 2019). Meanwhile, other studies have shown that MT is capable of producing texts that are of better quality than the writing of intermediate-level EFL students (Briggs, 2018; Stapleton & Kin, 2019) and demonstrating higher writing proficiency in various linguistic parameters (Ducar & Schocket, 2019; Tsai, 2019). Moreover, the accuracy of MT is likely to improve in the near future. As Belam (2003) asserted, students are highly likely to use it at some stage in their lives and careers, and they can benefit from it by gaining a new perspective on the target language. Therefore, simply prohibiting MT in L2 classrooms will be ineffective (Briggs, 2018; Lee, 2020). Since more students are likely to use MT for educational purposes in the future, L2 writing teachers should find effective strategies for integrating MT into language learning. They should also be adequately prepared for potential problems such as academic dishonesty, overdependence on MT, and the possible redefinition of their roles as teachers in L2 writing classrooms.

## References

- Alhaisoni, E., & Alhaysony, M. (2017). An investigation of Saudi EFL university students' attitudes towards the use of Google Translate. *International Journal of English Language Education*, 5(1), 72–82. <https://doi.org/10.5296/ijele.v5i1.10696>
- Bahri, H., & Mahadi, T. (2016). Google Translate as a supplementary tool for learning Malay: A case study at Universiti Sains Malaysia. *Advances in Language and Literary Studies*, 7(3), 161–167.
- Barkaoui, K. (2007). Revision in second language writing: What teachers need to know. *TESL Canada Journal*, 25(1), 81–92. <https://doi.org/10.18806/tesl.v25i1.109>
- Belam, J. (2003). *Buying up to falling down: A deductive approach to teaching post-editing* [Paper presentation]. MT Summit IX, Workshop on Teaching Translation Technologies and Tools, New Orleans, LA. <https://aclanthology.org/volumes/2003.mtsummit-tttt/>
- Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), 207–217. <https://doi.org/10.1016/j.jslw.2010.10.002>
- Bowker, L., & Ciro, J. B. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Publishing.
- Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses. *JALT CALL Journal*, 14(1), 3–24.
- Butzkamm, W., & Caldwell, J. (2009). *The bilingual reform: A paradigm shift in foreign language teaching*. Narr Studienbücher.
- Cardelle, M., & Corno, L. (1981). Effects on second language learning of variations in written corrective feedback on homework assignments. *TESOL Quarterly*, 15(3), 251–261. <https://doi.org/10.2307/3586751>

- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267–296. [https://doi.org/10.1016/S1060-3743\(03\)00038-9](https://doi.org/10.1016/S1060-3743(03)00038-9)
- Chandra, S. O., & Yuyun, I. (2018). The use of Google Translate in EFL essay writing. *LLT Journal: A Journal on Language and Language Teaching*, 21(2), 228–238. <https://doi.org/10.24071/lt.2018.210212>
- Chen, M.-H., Huang, S.-T., Chang, J. S., & Liou, H.-C. (2015). Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Computer Assisted Language Learning*, 28(1), 22–40. <https://doi.org/10.1080/09588221.2013.783873>
- Cheng, K.-H., Liang, J.-C., & Tsai, C.-C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education*, 25, 78–84. <https://doi.org/10.1016/j.iheduc.2015.02.001>
- Choi, Y. H., & Lee, J. (2006). L1 use in L2 writing process of Korean EFL students. *English Teaching*, 61(1), 205–225.
- Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *The Acquisition of Second Languages and Innovative Pedagogies*, 10, 108–121. <https://doi.org/10.7203/attic.10.2228>
- Cohen, A., & Brooks-Carson, A. (2001). Research on direct versus translated writing: Students' strategies and their results. *The Modern Language Journal*, 85(2), 169–188. <https://doi.org/10.1111/0026-7902.00103>
- Cook, G. (2010). *Translation in language teaching: An argument for reassessment*. Oxford University Press.
- Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google Translate. *Foreign Language Annals*, 51(4), 779–795. <https://doi.org/10.1111/flan.12366>
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10(3), 161–184. [https://doi.org/10.1016/S1060-3743\(01\)00039-X](https://doi.org/10.1016/S1060-3743(01)00039-X)
- Garcia, I., & Pena, M. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning*, 24(5), 471–487. <https://doi.org/10.1080/09588221.2011.582687>
- Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, 37, 112–121. <https://doi.org/10.1016/j.esp.2014.09.001>
- Goulet, M.-J., Simard, M., Parra Escartín, C., & O'Brien, S. (2017). Using machine translation for academic writing in English as a second language: Results of an exploratory study on linguistic quality. *ASp. La Revue du GERAS*, 72, 5–28. <https://doi.org/10.4000/asp.5045>
- Howatt, A. (2004). *A history of English language teaching*. Oxford University Press.
- Jolley, J. R., & Maimone, L. (2015). Free online machine translation: Use and perceptions by Spanish students and instructors. In A. Moeller (Ed.), *Learn languages, explore cultures, transform lives* (pp. 181–200). Robert M. Terry.
- Kim, E.-Y. (2011). Using translation exercises in the communicative EFL writing classroom. *ELT Journal*, 65(2), 154–160. <https://doi.org/10.1093/elt/ccq039>

- Kim, Y., & Yoon, H. (2014). The use of L1 as a writing strategy in L2 writing tasks. *GEMA Online Journal of Language Studies*, 14(3), 33–50. <https://doi.org/10.17576/GEMA-2014-1403-03>
- Kliffer, M. (2008). Post-editing machine translation as an FSL exercise. *Porta Linguarum*, 9, 53–67. <https://doi.org/10.30827/Digibug.31745>
- Koehn, P. (2020). *Neural machine translation*. Cambridge University Press.
- Kurzer, K. (2018). Dynamic written corrective feedback in developmental multilingual writing classes. *TESOL Quarterly*, 52(1), 5–33. <https://doi.org/10.1002/tesq.366>
- Le, Q. V., & Schuster, M. (2016, September 27). A neural network for machine translation, at production scale. *Google Research Blog*. <https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>
- Lee, S.-M. (2009). Exploring the writing and editing patterns of the different language proficiency groups on a wiki. *Foreign Languages Education*, 16(3), 95–122.
- Lee, S.-M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, 33(3), 157–175. <https://doi.org/10.1080/09588221.2018.1553186>
- Liu, Y. (2009). *Translation in second language writing: Exploration of cognitive process of translation*. VDM Verlag.
- Malmkjær, K. (Ed.). (1998). *Translation and language teaching*. St. Jerome.
- Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15(2), 118–141. <https://doi.org/10.1016/j.jslw.2006.01.003>
- Murtisari, E., Widiningrum, R., Branata, J., & Susanto, R. (2019). Google Translate in language learning: Indonesian EFL students' attitudes. *The Journal of Asia TEFL*, 16(3), 978–986. <https://doi.org/10.18823/asiatefl.2019.16.3.14.978>
- Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, 21(1), 29–49. <https://doi.org/10.1080/09588220701865482>
- Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, 21(2), 241–258. <https://doi.org/10.1017/S0958344009000172>
- O'Brien, S., Simard, M., & Goulet, M.-J. (2018). Machine translation and self-post-editing for academic writing support: Quality explorations. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 237–262). Springer.
- Poibeau, T. (2017). *Machine translation*. MIT Press.
- Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, 10(4), 277–303. [https://doi.org/10.1016/S1060-3743\(01\)00046-7](https://doi.org/10.1016/S1060-3743(01)00046-7)
- Roca De Larios, J., Murphy, L., & Marin, J. (2002). Critical examination of L2 writing process research. In S. Ransdell & M.-L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 11–47). Kluwer Academic Publishers.
- Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, 8(1), 31–54. <https://doi.org/10.1191/1362168804lr133oa>
- Stapleton, P., & Kin, B. (2019). Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes*, 56, 18–34. <https://doi.org/10.1016/j.esp.2019.07.001>



- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272. <https://doi.org/10.1016/j.jslw.2007.06.003>
- Tsai, S.-C. (2019). Using Google Translate in EFL drafts: A preliminary investigation. *Computer Assisted Language Learning*, 32(5–6), 510–526. <https://doi.org/10.1080/09588221.2018.1527361>
- Valijärvi, R. L., & Tarsoly, E. (2019). Language students as critical users of Google Translate: Pitfalls and possibilities. *Practitioner Research in Higher Education Journal*, 12(1), 61–74.
- Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers & Composition*, 38(A), 79–95. <https://doi.org/10.1016/j.compcom.2015.09.012>
- Waller, L., & Papi, M. (2017). Motivation and feedback: How implicit theories of intelligence predict L2 writers' motivation and feedback orientation. *Journal of Second Language Writing*, 35, 54–65. <https://doi.org/10.1016/j.jslw.2017.01.004>
- Wang, W., & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11(3), 225–246. [https://doi.org/10.1016/S1060-3743\(02\)00084-X](https://doi.org/10.1016/S1060-3743(02)00084-X)
- White, K. D., & Heidrich, E. (2013). Our policies, their text: German language students' strategies with and beliefs about web-based machine translation. *Die Unterrichtspraxis*, 46(2), 230–250. <https://doi.org/10.1111/tger.10143>

## Appendix A. Rubric for the Students' L2 Writing

Score	Description
6	<p>The writing shows a good command of the language (i.e., very few incorrect uses of language).</p> <p>The development of ideas is logical, and the intended meanings are very clear.</p> <p>The sentences and word choices are varied and adequate in the context.</p>
5	<p>The writing includes some language errors.</p> <p>The development of ideas is clear, and the meanings are easy to understand.</p> <p>The sentences and word choices are somewhat varied and adequate.</p>
4	<p>Language errors are often found, but they do not interfere with the understanding of the essay.</p> <p>The development of ideas is adequate, and the meanings are conveyed.</p> <p>Overall, the sentences and word choices are adequate.</p>
3	<p>Language errors are frequently found, and they sometimes interfere with the understanding of the essay.</p> <p>The development of ideas is simple, and it sometimes creates confusion.</p> <p>The sentences show little variety, and the word choices are sometimes inappropriate.</p>
2	<p>The language errors frequently distract from the understanding of the writing.</p> <p>The development of ideas is unclear and difficult to follow.</p> <p>The sentences and word choices are simple and often inappropriate.</p>
1	<p>The language errors frequently distract and significantly impede the understanding of the writing.</p> <p>The development of ideas is unclear and difficult to follow. The paragraphs are rarely developed.</p> <p>The sentences and word choices are simple and frequently inappropriate.</p>

## Appendix B. Samples of the Students' Changes

Category	Subcategory	Original	Revised
<b>Size</b>	Symbol	...easy, because it...	...easy because it...
	Word	It is hard to...	It is <u>difficult</u> to...
	Phrase	...ancestors and scholars have pointed out the flaw of common people's writing skills.	...ancestors and scholars <u>have continued to question people's writing skills.</u>
	Clause / Sentence	Also, using slangs and abbreviation allows people to express intimacy toward others even easily by SNS	Also, <u>we can express the intimacy of each other more easily by using the abbreviations and slang which are characteristics of the text message.</u>
<b>Type</b>	Addition	...use languages.	...use languages <u>perfectly.</u>
	Deletion	Almost everyone should follow...	Everyone should follow...
	Substitution	...creates an oxymoron that...	...creates <u>a contradiction</u> that...
	Permutation	These skills make it easy for others to communicate.	<u>Others can communicate easily with these skills.</u>
	Distribution	Korean education focuses on reading and listening.	Korean education focuses on reading and listening <u>when learning English.</u>
	Consolidation	Young generations use their native language poorly. They often learn the language from social networks.	<u>Young generations use their native language poorly because they often learn the language from social networks.</u>
	Reordering	They see the benefits hardly.	They <u>hardly</u> see the benefits.
<b>Function</b>	Mechanics	They changed their attitudes, So they...	They changed their attitudes, so they...
	Word Choice	views of a life	<u>value</u> of a life
	Grammatical	A new trend was appeared.	A new trend <u>has</u> appeared.
	Cosmetic	There were limitations with existing materials.	<u>Existing materials had limitations.</u>
	Texture	Young people use social networks everyday, and we create many new words there.	Young people use social networks everyday, and <u>they</u> create many new words there.
	Unnecessary Expression	Surely, AI can be effectively used.	AI can be effectively used.
	Explicature	Texting has been criticized for killing language.	Texting has been criticized for killing language <u>by overusing abbreviations, spellbinding, and etc.</u>

## Appendix C. Interview Questions

1. For what and how do you usually use MT, if you use it?
2. What is your general impression toward using MT?
3. How did you use MT during the task?
4. What were the advantages and disadvantages of using MT?
5. Are you going to use it again for L2 writing? Why and why not? How would you use MT for L2 writing, if you are going to use it again?

## Appendix D. Differences Among the Groups: Size of Changes

Size	Mean (SD)					F	p
	2 (n = 14)	3 (n = 18)	4 (n = 22)	5 (n = 18)	6 (n = 8)		
Symbol	0.2(0.611)	0.1(0.471)	0.0(0.000)	0.0(0.000)	0.0(0.000)	2.148	.083
Word	4.1(3.655)	4.3(1.968)	3.4(2.267)	2.3(2.163)	3.6(3.906)	0.785	.538
Phrase	2.6(2.618)	2.7(1.820)	1.8(1.541)	1.9(1.364)	1.9(2.055)	0.212	.931
Clause / Sentence	0.9(1.466)	0.9(1.110)	1.0(1.541)	0.7(1.118)	1.0(0.487)	1.251	.297
Total	7.8(5.680)	8.0(3.420)	6.2(4.475)	4.9(3.245)	6.5(5.594)	0.731	.574

## Appendix E. Differences Among the Groups: Type of Changes

Category Subcategory		Mean (SD)					F	p
		2 (n = 14)	3 (n = 18)	4 (n = 22)	5 (n = 18)	6 (n = 8)		
<b>Surface</b>	Addition	2.0(2.09)	3.5(2.50)	2.7(3.18)	1.4(1.46)	1.7(1.58)	2.054	.095
	Deletion	0.7(0.80)	1.6(1.57)	0.7(1.31)	0.5(0.61)	0.7(1.8)	2.118	.087
	Substitution	2.6(3.58)	3.7(2.67)	3.45(2.40)	4.0(4.14)	5.0(3.66)	0.749	.562
	Permutation	1.07(1.81)	1.7(1.01)	2.9(1.86)	2.2(1.59)	2.13(1.80)	1.332	.266
	Distribution	0.1(0.36)	0.0(0.00)	0.05(0.21)	0.2(0.51)	0.3(0.70)	0.934	.449
	Consolidation	0.07(0.26)	0.0(0.00)	0.2(0.39)	0.1(0.32)	0.4(0.51)	2.139	.084
	Reordering	0.8(1.12)	0.3(0.66)	0.2(0.50)	0.2(0.38)	0.5(1.40)	1.883	.122
Subtotal		7.6(6.28)	10.8(6.08)	9.3(7.46)	8.6(6.37)	11.13(9.37)	0.616	.652
<b>Micro-text-based</b>	Addition	0.0(0.00)	0.06(0.23)	0.5(1.01)	0.7(.097)	0.9(0.64)	3.480	.120
	Deletion	0.0(0.00)	0.06(0.23)	0.2(0.35)	0.0(0.00)	0.0(0.00)	1.404	.241
	Substitution	2.9(2.43)	1.6(1.68)	0.5(0.96)	0.2(0.38)	0.0(0.00)	10.625	<.001
	Permutation	0.3(0.61)	0.1(0.47)	0.09(0.29)	0.3(0.68)	0.6(1.40)	1.252	.296
	Distribution	0.3(0.46)	0.0(0.00)	0.1(0.46)	0.0(0.00)	0.0(0.00)	2.365	.060
	Consolidation	0.07(0.26)	0.0(0.00)	0.05(0.21)	0.0(0.00)	0.1(0.35)	0.880	.480
	Reordering	1.0(0.87)	0.0(0.00)	0.0(0.00)	0.0(0.00)	0.0(0.00)	21.656	<.001
Subtotal		4.5(1.98)	1.8(1.73)	2.3(3.98)	1.2(1.20)	3.0(4.20)	3.320	.150

## Appendix F. Differences Among the Groups: Function of Changes

Function	Mean (SD)					<i>F</i>	<i>p</i>
	2 ( <i>n</i> = 14)	3 ( <i>n</i> = 18)	4 ( <i>n</i> = 22)	5 ( <i>n</i> = 18)	6 ( <i>n</i> = 8)		
Mechanical	0.2(0.611)	0.1(0.471)	0.0(0.000)	0.0(0.000)	0.0(0.000)	2.147	.830
Word Choice	3.3(2.76)	1.56(1.68)	0.9(2.04)	0.4(0.91)	1.2(1.58)	5.468	.010
Grammar	3.7(2.68)	5.1(3.00)	4.5(4.41)	3.1(3.72)	1.2(1.28)	2.109	.080
Cosmetic	3.3(2.70)	5.0(3.48)	3.4(2.80)	4.3(2.51)	6.4(7.05)	1.478	.217
Texture	0.3(1.33)	0.1(0.32)	0.4(0.66)	0.4(0.97)	2.1(2.29)	5.260	.001
Unnecessary Expression	0.2(0.57)	0.4(1.29)	0.3(0.076)	0.2(0.54)	0.13(0.35)	0.293	.882
Explicature	0.8(1.67)	0.3(0.76)	2.2(3.36)	0.9(1.16)	1.6(1.06)	2.485	.051

### About the Author

Sangmin-Michelle Lee is a professor of global communication at Kyung Hee University in Korea. She earned her PhD from Pennsylvania State University in language education. She has published papers on language learning in a technology-enhanced learning environment, machine translation, L2 writing, game-based learning, and digital creativity.

**E-mail:** [sangminlee@khu.ac.kr](mailto:sangminlee@khu.ac.kr)