Training students to use online translators and dictionaries: The impact on second language writing scores

O'Neill, Errol M. ✉
The University of Memphis, USA (emoneill@memphis.edu)

Abstract

Online translators have become an ever-present component of language learning. Recent research has found that online translators are used for graded work at least sometimes by 87.7% of intermediate students learning Spanish and French (O'Neill, 2019). The same percentage of students self-reports using online dictionaries. Given how widespread these tools are, the impact of their usage needs to be explored. Although some research has looked into the effects of online translation on writing, no previous study has compared the composition scores of students using online translators with those using online dictionaries. This article reports the results of a study with over 1,000 compositions written by students in one of five groups: online translation use after training, online translation use with no training, online dictionary use after training, online dictionary use without training, and writing with the use of neither of these tools. The results ($p < 0.05$) show that students who used Google Translate after training scored the highest of all five groups on both experimental writing tasks, followed by those who used an online dictionary with prior training. The group that fared the least well was that which used neither tool. A posttest and delayed posttest however showed mixed results for online translators. Based on these findings, and given the difficulty of prohibiting the use of these tools on graded work outside the classroom, the author argues for training students in the responsible use of online dictionaries and translators for second language writing.

Keywords: online translation; online dictionaries; CALL; L2 writing; educational technology

Training students to use online translators and dictionaries: The impact on second language writing scores

1. Introduction

Technology is changing the way students approach learning a second language (L2). Learners no longer need to carry paper dictionaries with them for assistance with vocabulary words they would like to look up in the language they're learning; free online dictionaries (ODs) are readily accessible from any smartphone or tablet. With the click of a button or tap of a finger, language learners can have a word, phrase, sentence, or entire paragraphs translated for them in a free online translator (OT). Given the fact that online language resources can be accessed essentially at all times and from anywhere, instructors are faced with the tasks of knowing what resources are available and what effect their use might have on the process of learning a language.

Two common free tools that are used by language learners are online dictionaries (e.g. WordReference), which students can use to translate individual words or sometimes short phrases, and online translators (e.g. Google Translate), which can be used to convert text as short as a word and as long as multiple paragraphs. Previous research has shown that these tools are used widely by L2 learners. Jolley and Maimone (2015) found that 97.66% of students in their study had used online translation at least once. In a survey of 905 respondents, Clifford, Merschel, and Munné (2013), 71% of university Spanish and French students said they use OTs sometimes or often. More recent research (O'Neill, 2019) suggests that this number has increased, with 87.7% of students self-reporting the use of OTs for graded work sometimes, often, or always, even when this tool is prohibited by instructors. Online dictionaries are also used by a large percentage of students. Jin and Deifell's (2013) study of 265 language students found that 85% of respondents reported using an OD. In a study by Author (2019), 87.7% of students reported using ODs for graded coursework, in line with Jin and Deifell's finding and the same number that reported OT usage in Author's study.

Although online translators and dictionaries are widely used, opinions are varied on their usefulness and appropriateness for L2 learners. Initial reaction from instructors and researchers after the release of the first OT, Babel Fish, was largely negative. A number of authors (e.g. Luton, 2003, Burton, 2003, Stapleton, 2005, Williams, 2006) considered online translation use to be "an unethical practice" (Burton, 2003, p. 320), with online translations described as being often readily recognizable due to "red flags" (Luton, 2003, p. 768): errors such as idioms that are translated word-for-word, proper nouns treated as common nouns, and a general lack of contextual awareness. The general approach advocated in the literature was one of detection and prevention of OT use (e.g. Steding, 2009).

As time has progressed, and OT technology has improved, opinions have begun to shift. While some negative articles have still appeared (e.g. García & Peña, 2011, Correa, 2014), others (e.g. Fountain & Fountain, 2009, Niño, 2009) take a less polarized approach, discussing both positive and negative aspects of online translators depending on the level of the student or the goals of the assignment. In 2013, Clifford et al. (2013) found that only 42% of instructors surveyed considered OT use to be cheating, while Knowles (2016) finds this to be even lower at only 35% expressing such a view. Niño (2009) discussed positive aspects of OTs, including using online translators to make comparisons between students' L1 and L2 and to provide answers to their immediate linguistic needs, Some authors (e.g. White & Heidrich, 2013, Knowles, 2016, Ducar & Schocket, 2018) recognize a possible role for OTs in the classroom, for example to help students explore the target language, to learn about errors through post-editing of OT output, and to make comparisons between their L1 and L2.

Less work has been published related to online dictionaries. A few authors (e.g. Correa, 2014, Knospe, 2017) have pointed to potential issues, such as a lack of awareness among students that ODs do not provide complete

translations like OTs and the suggestion that ODs may interrupt the thought process of students who consult them frequently. Most discussion, however, has generally been positive towards online dictionaries. Elola, Rodríguez-García, and Winfrey (2008) mention the ability to check spelling and word meanings, as well as the presence of sample sentences to provide models of correct structures in the target language. Xu (2011) points to the advantages of ODs offering fast access to definitions, as well as interactivity (such as the ability to click on a word within a definition to look it up). Amirian and Hashmatifar (2013) report that instructors often prefer ODs to paper dictionaries and online translation, stating that ODs provide access to a greater amount of information about the entries, while Fredholm (2015) reports that student find ODs to be more reliable than OTs. Jin and Deifell (2013) view online dictionary use positively, stating that ODs are practical and easy for learners to use.

With nearly 9 in 10 students using OTs and ODs, and given the ubiquity of technology inside and outside the classroom, the question no longer seems to be if such technologies should be used, but which ones, when and how. In their seminal article, Ducar and Schocket (2018) state that it is not the role of instructors to prevent the use of online translation, but to encourage responsible use of OTs and related technologies "to help learners understand that positive progress toward greater proficiency and ethical use of technologies are critical 21st-century skills" (p. 15).

Despite frequent use of online translators and dictionaries, students may not be proficient in their effective use. In a survey of 126 English as a Foreign Language (EFL) learners in Iran, Dashtestani (2014) found that only 14.28% of respondents considered themselves proficient in using electronic dictionaries (which the authors defined as online, pocket, or CD-ROM based dictionaries), with 33.6% of students saying they were fairly proficient, 47.3% reporting they were a little proficient, and 4.76% saying they were not proficient at all. Their instructors were more categorical in their assessment; 80.3% of the 53 teachers surveyed said that EFL students lack adequate training in electronic dictionary usage. 78.6% of students agreed or strongly agreed that they needed training on how to use such tools. In their description of online dictionary use, Niitemaa and Pietilä (2018) report that students' skills in using online dictionaries are "often inadequate" and, consequently, students can find using a dictionary to be "slow and tedious" (p. 454). The authors state that "dictionary skills should be included in foreign language curricula as an essential part of literacy" (p. 461).

Bahri and Mahadi (2016) called on instructors to help students "optimally benefit" from new technologies such as online translators (p. 165). One student in their study of 17 Malay L1 students learning English specifically mentions the need to be taught about Google Translate to learn more effectively, while other students cited OTs as encouraging learning, reducing stress, and increasing their enjoyment in learning the language. In Knowles' (2016) study of university instructors, several of the 20 participants interviewed mentioned a desire for, or a possible role for, training students in OT use. For example, one participant said in the future she might "show them [students] how it works," while another called training students for OT use "a really good thing" (p. 69). Knowles states that there is a need for "professional development for faculty and training for students in regards to pedagogically sound practices for using OMT (online machine translation) in the L2 classroom" (p. 79).

The second half of Ducar and Schocket (2018)'s recommendation, using technology ethically, has also been discussed in the literature related to online translation. The definition of ethical use of online tools can vary from instructor to instructor: Clifford et al. (2013) found that 42% of instructors considered any use of OT to be cheating, while 21% said it was not and 37% chose "other," some of whom indicating that it depends on the context or extent to which online translation is used. Conversely, Jolley and Maimone (2015) found that 87.18% of instructors judged online translation use for individual words to be "completely ethical," while the same percentage found that translating paragraphs or entire texts was "completely unethical." 59.98% considered translating short phrases to be "somewhat ethical" or "completely ethical." The authors suggest instructors consider the advantages and disadvantages of OT use and how OTs would fit into their teaching philosophy. Although Larson-Guenette (2013) similarly recommends creating a dialogue with learners about online tools, she states that many students in the study reported that teachers did not allow OTs to be used due to concerns

over academic honesty. No study was found investigating the opinions of instructors on the ethicality of online dictionary use, although the researcher has found general support anecdotally among instructors for its use on writing assignments.

Faced with the current state of affairs, in which a large number of students are using OTs but students and teachers feel students are not adequately prepared to use them, and a number of instructors question the ethicality of OTs, some researchers have attempted to analyze online translation and dictionary usage for student written production. García and Peña (2011) studied 16 students who first wrote using an online translator with the opportunity to post-edit their work, then wrote a second composition directly in the L2. Based on the number of pauses during writing and the number of errors produced, two raters found that 10-11 students (based on the rater) wrote better overall using online translation, with the remaining performing better writing directly in the L2. In their view, however, better performance did not necessarily indicate more learning had occurred, although they believe OTs could help motivate learners to communicate more in the language they are studying.

White and Heidrich (2013) asked students to write a composition in English and translate it into German using Google Translate. Students were allowed to post-edit their texts as they saw fit. 27.8% of students left the translation that the OT had created as-is or changed only one word. The remaining students did engage in more extensive post-editing and were able to reduce the number of errors, as compared to the raw output of Google Translate. At times, however, the authors note that the increased accuracy was due in part to simplifying the language used. White and Heidrich judged that online translation did not allow students to write at a level that was notably better than what they could produce on their own, but that it could be used as a support for learning as opposed to a replacement. Fredholm's (2014) study of Swedish students learning Spanish similarly found that the group who had access to an online translator committed fewer errors than the control group who wrote without the help of a translator, although the author judged that neither group wrote compositions that were judged to be well-written overall. Wuttikrikunlaya, Singhasiri, and Keyuravong (2018) ran a study with 14 first-year English learners in Thailand that allowed participants to use whichever online tool they preferred on a composition task. All of the participants used both an online dictionary and an online translator at least one time each, although they preferred OTs (74.91% of queries) to ODs (16.25%).

O'Neill (2016) looked not at frequency of use or number of errors, but rather the effects of online translation on writing scores. Thirty-two participants were asked to write compositions in one of three conditions: use of an OT for writing with prior training in its strengths and weakness, use of an OT without prior training, and no use of online translation allowed. The two groups using online translation did not perform worse than the control group on any measure overall, and the group using online translation after prior training actually received a significantly higher mean score (25.36 out of 30) on one of the two compositions participants wrote, when compared to the group not using an OT (20.65 out of 30). Although not outperforming the control group on all measures, the OT group without training did score significantly higher on four out of six categories (content, grammar, comprehensibility, and spelling/accents) on one or both compositions. The results suggest that the use of an online translator did not hurt, and may have helped, students in their writing.

A follow-up study to O'Neill (2016) was designed to confirm these results, which suggest online translation does not negatively affect and can improve student writing scores, and also to expand the analysis to an additional electronic tool: online dictionaries. The research questions for the current study are as follows:

➢ RQ1: Does the use of electronic tools (online translators or dictionaries) affect the writing scores of intermediate second language learners?

➢ RQ2: Does prior training in the use of electronic tools (online translators or dictionaries) affect the writing scores of intermediate second language learners?

➢ RQ3: Do learners who use online translation outscore those who use online dictionaries?

> ➤ RQ4: Are there any effects on learners' scores once they stop using the electronic tool (online translators or dictionaries) as compared to their peers who did not write using either tool?

2. Materials and methods

The current discussion focuses on one aspect of a larger study; as such, only the materials pertaining to this portion of the research will be presented. The design described below closely follows that of O'Neill (2016), with the addition of two new elements: the use of online dictionaries by two of the groups (O'Neill focused only on online translation) and a Delayed Posttest (which was absent from O'Neill, 2016). The study also differs from previous work such as Steding (2009), which focused on ethical issues and preventing online translation use; Tight (2017), which discussed instructing students about strengths and weaknesses of OTs but did not involve subsequent comparison of OT use with students using ODs or no tool; White and Heidrich 2013, which was an experimental study, but looked only at post-editing of online translation output, not at the effects of training or differences between OTs and ODs; and Wuttikrikunlaya et al. (2018), another experimental study in which students could use any tool they chose, but which did not seek to tease out the impact of each tool individually on student writing. The current study's design was approved by the author's Institutional Review Board (IRB), ensuring the appropriateness of the protocol used and the informed consent of participants in human subjects research.

2.1 Treatment groups

In order to determine the possible effects both of writing with one of the two tools (the OT Google Translate or the OD WordReference) and training in the use of each tool, participants were randomly placed into one of five groups: writing with the permitted use of Google Translate with prior training (GT+T), Google Translate without prior training (GT-T), WordReference with prior training (WR+T), and WordReference without prior training (WR-T), with an independent group that used neither tool (Indy). This design allows the analysis of the effects of just the use of the tool, as well as any positive or negative effects of being trained in how to use the tool effectively. Students were placed into treatment groups once they had completed the training or cultural session (see Section 2.3) to ensure that all students had participated in the appropriate session prior to the experimental tasks.

A total of 310 American students taking intermediate Spanish or French participated in the study. All participants provided informed consent to participated, as stipulated by the IRB at the author's university. Participants wrote 1,113 compositions in all. All students did not complete all writing tasks due to attrition and absences, but since student scores are being considered in the aggregate (not tracking the individual performance of a student, but rather comparing the means of the entire group) all compositions were retained for analysis. This study replicates O'Neill (2016), but with the addition of the two WordReference groups to compare the effects of OT use with those of ODs.

2.2 Pretest

All participants were asked to complete a Pretest in their target language, Spanish or French, in a computer lab on Microsoft Word running on a Windows 8 PC. This task involved writing a composition on a preset topic. The goal of this first task was to establish a baseline for participants by having them write without using either tool, online translators or dictionaries, to allow for comparisons between the experimental groups who would later use one of the tools, and the control group that was not permitted to use an OD or OT throughout the study. In order to ensure the validity of the study's results composition tasks were designed so that the Pretest and posttests (described in Section 2.5) simulate how students might write in a typical classroom situation where the teacher instructs students to complete a composition without access to online translators or dictionaries. This follows the design of O'Neill (2016). The Pretest prompt was as follows; it should be noted that the versions of all writing task prompts for French students used francized names (e.g. Corine) and French place names (e.g.

Lyon):

> *You are writing a letter to Corina, a student from Madrid, Spain who is writing an article about summer at American universities. (3 paragraphs)*

> *1) Greet Corina, then give 2 reasons why summer can be interesting/fun for American students and 2 reasons why it can be boring/annoying.*

> *2) Talk about some interesting activities you did last summer, either for school/work, or with your family and friends.*

> *3) Convince Corina why it is important for students to have a long summer vacation or why it is better for there to be a shorter (or no) summer vacation.*

2.3 Training

A week after the Pretest, participants attended one of two 20-minute sessions. Students placed into the GT+T and WR+T groups participated in a training session exploring the strengths and weakness of the tool students had been assigned to: Google Translate[1] or WordReference[2]. The remaining groups participated in a culture session about the profession of translation, which was attended by the Indy group as well as those participants assigned to the groups GT-T and WR-T, who would be using their respective tool without being trained in its use. The training session included the five following steps:

- ➢ Introduction to the tool

- ➢ Demonstration of how the tool works

- ➢ Testing of the tool through examples

- ➢ Review of the tool's strengths and weaknesses

- ➢ Explanation of how the tool should be used during the study

The first step defined what the tool was and gave some potential advantages and disadvantages of using it. The second part introduced students, through screenshots and explanations, to some key functions available for that tool. The third step provided sample words, expressions, and sentences to translate into the target language (Spanish or French) using the tool. The next step provided an overview of some of the strengths and weaknesses of the tool, including some that participants encountered while performing the third step. For the last step, students were instructed that they could "decide how much or how little to use [the tool]," "how much or how little editing you want to do," and that they could "not use any other website, including Google search while you're doing the composition." For each step, there were True/False, fill-in-the-blank, and short response comprehension-check questions to help ensure students were engaged in the lesson. Several authors (e.g. Ducar & Schocket, 2018, Knowles, 2016, McCarthy, 2004) have mentioned the possibility of instructing students on how best to use or not use online translators, and O'Neill (2016) described a similar study where such training was performed. This portion of the study was designed to represent how an instructor might carry out such training with a real-world class.

2.4 Experimental writing tasks (Compositions 1 and 2)

One week after the training or culture session they attended, participants completed Composition 1 in one of the five conditions mentioned above: GT+T, GT-T, WR+T, WR-TR, or Indy. The composition followed the same format as the Pretest, but in response to a different topic. The following week, participants completed

---

[1] The training module for Google Translate for students of Spanish can be accessed at: https://goo.gl/C8GhYF.
[2] The training module for WordReference for students of Spanish can be accessed at: https://goo.gl/deo624.

Composition 2 in the same condition as before but on a new topic. The purpose of these tasks was to determine how students who used online translators or dictionaries, with or without prior training, performed in comparison to the control group and to each other. In a classroom setting, some instructors might allow students to use an online translator or dictionary, while others might prohibit their use. In addition, some, but not all, instructors might give explicit instructions on how best to use or not use each tool. The design of this study, which replicates that of O'Neill (2016) but with the addition of ODs, is meant to account for each of these eventualities. If the mean scores obtained on one or both of the compositions were different from other groups to statistically significant level, this would suggest that the treatment had an effect on student writing. The prompts for Compositions 1 and 2 are reproduced below:

Composition 1 prompt

You are writing a letter to María Clara, a student from Toledo, Spain who is considering studying at your university in the United States. (3 paragraphs)

1) Greet María Clara, then give 2 reasons why your university is better/more desirable than other schools and 2 reasons why it is worse/less desirable than other schools.

2) Talk about some interesting activities you did last semester while attending your university, either at school for class or in town with friends.

3) Convince María Clara either why it is important for her to come to your university or why she might consider another school.

Composition 2 prompt

You are writing a letter to Juan Pedro, a student from Leon, Spain who is considering applying for a job in your field in the United States. (3 paragraphs)

1) Greet Juan Pedro, then describe 2 things that are potentially good about jobs in your field and 2 things that are potentially negative about jobs in your field.

2) Talk about some important things you did, either at school or elsewhere, which helped you to prepare for having a job in this field.

3) Convince Juan Pedro either why it is important for him to apply for your field in the United States or why he might consider another field (either related or completely different) instead.

2.5 Posttest and Delayed Posttest

A week after writing Composition 2, participants completed the Posttest, which again followed the same format as the other writing tasks. As with the Pretest, participants were not allowed to use either tool in order to gauge any effect the training or use of the tool would affect student writing once the tool was no longer authorized. The Posttest was meant to evaluate students' performance right after the experimental tasks to determine any possible immediate effect of the use of OTs and ODs, or of the training some participants had received, on student writing scores. Finally, four to five weeks after Composition 2, participants completed the Delayed Posttest. This composition, which followed the same format as the other writing tasks, was administered to detect any longer-term change in student performance when writing without the tools. This addresses a limitation of O'Neill (2016), which only gauged student performance in a posttest immediately after the experimental tasks without addressing possible effects after more time had passed. The prompts for the Posttest

and Delayed Posttest are as follows:

Posttest prompt

You are writing a letter to Bernardo, a student from Barcelona, Spain who is writing an article about winter at American universities. (3 paragraphs)

1) Greet Bernardo, then give 2 reasons why winter can be interesting/fun for American students and 2 reasons why it can be boring/annoying.

2) Talk about some interesting activities you did last winter, either for school/work or with your family and friends.

3) Convince Bernardo why it is important for students to have a long winter vacation or why it would be better for there to be a shorter (or no) winter vacation.

Delayed Posttest prompt

You are writing a letter to Jorge, a student from Salamanca, Spain who is writing an article about spring at American universities. (3 paragraphs)

1) Greet Jorge, then give 2 reasons why spring can be interesting/fun for American students and 2 reasons why it can be boring/annoying.

2) Talk about some interesting activities you did last spring, either for school/work or with your family and friends.

3) Convince Jorge why it is important for students to have a long spring vacation or why it would be better for there to be a shorter (or no) spring vacation.

2.6 Scoring

After compositions were collected, twelve raters were recruited to score a subset of each of the five writing tasks (Pretest, Compositions 1 and 2, Posttest, and Delayed Posttest). Given the fact that there were over 1,000 compositions collected, it was not feasible to have two raters score all compositions. Instead, each individual composition was scored blindly by two of the raters; neither of the raters knew what scores the other had given. Raters were paired off, meaning for example that Rater 1 and Rater 2 scored the exact same portion of compositions without consulting each other; Rater 3 and Rater 4 scored the same compositions as each other; and so on. Raters scored the compositions out of 30 possible points, using a rubric (Appendix) taking into account the following features: Content, Grammar, Spelling/Accents, Syntax, Vocabulary, and Understandability. These were the same categories used in O'Neill (2016), except that Comprehensibility and Remaining Grammar were renamed Understandability and Grammar for the sake of clarity and simplicity. Raters read the compositions and submitted their scores to the researcher for analysis.

3. Results

After the scores were received from the raters, two sets of analyses were run. The first concerns interrater reliability, in order to ensure that the two scores obtained by the raters for each composition were significantly comparable to represent a valid measure. The second set of analyses involves determining and comparing the mean composition scores for each of the five treatment groups on the five writing tasks, in order to determine any differences in the scores based on the use of online translators, dictionaries, or neither tool; as well as any

differences related to whether or not those participants using the tools had received training prior to completing Compositions 1 and 2.

3.1 Interrater Reliability

It is important to establish that the scores given by raters are reliable in order to permit further analysis comparing the performance of the different groups. In order to do so, the Reliability Analysis procedure was run in SPSS. A strong level of interrater reliability was found on all five writing tasks, with all intraclass correlations for average measures exceeding .700 ($p < .001$). Specifically, the correlations were as follows: Pretest, .737; Composition 1, .759; Composition 2, .809; Posttest, .812; Delayed Posttest, .703. These results mean that the two scores from raters across compositions on each writing task were statistically similar for all five tasks, indicating that the composition scores do not vary significantly depending on which rater was scoring them. Given the level of interrater reliability, the results can be considered reliable and further analysis is possible.

3.2 Writing Tasks

The mean scores of the treatment groups (GT+T, GT-T, WR+T, WR-T, Indy) were obtained for each of the five writing tasks (Pretest, Composition 1, Composition 2, Posttest, Delayed Posttest). One-way ANOVA tests were conducted to compare means from each of the treatment groups for the five writing tasks to determine if there was any statistically significant difference between the means. The results of each writing task will be described below.

Pretest - Since the pretest was administered before training was conducted and before the electronic tools (online translators and dictionaries) were used by participants, the expectation was that means between the groups would be statistically similar. For eight out of the ten pairwise comparisons, there was no significant difference ($p < 0.05$) between the means of the five groups in the initial pretest, which would be expected if each group started out with the same level in written Spanish or French prior to the start of the study. However, the Indy group had a significantly lower mean ($\bar{x} = 18.7857$ out of 30 possible points) on the Pretest as compared to the GT-T ($\bar{x} = 20.57$), the Google Translate group that had no prior training ($p = 0.040$) and as compared to WR+T ($\bar{x} = 20.88$), the WordReference group that had prior training ($p = 0.005$). This means that any results between, Indy and GT-T, and between Indy and WR+T, groups should be interpreted with caution since their initial scores were not statistically similar.

All of the remaining eight comparisons for the Pretest (Indy versus GT+T, Indy versus WR-T, and all pairwise comparisons among the GT+T, GT-T, WR+T, and WR-T groups) did not find a difference at a significant level ($p < 0.05$), meaning that their written performance in Spanish or French was found to be statistically similar at the onset of the study. Given the statistical similarity among eight out of ten of the groups in the Pretest, while keeping in mind the caveat for Indy vs. GT-T and WR+R, further analysis can be made for the subsequent writing tasks.

Composition 1 - On the first experimental task, Composition 1, students were allowed to use Google Translate, WordReference, or neither, depending on which treatment group they were placed into. In addition, groups GT+T and WR+T had received training in the use of their respective tool, while other groups attended a control session about the profession of translation. On Composition 1, there was a significant difference in all 10 comparisons between groups, meaning that each difference seen in score was statistically significant. The highest scoring group was Google Translate with prior training ($\bar{x} = 25.53$), earning greater than two points more than the next closest group, Google Translate without prior training ($\bar{x} = 23.69$). The group that had the next highest score was WordReference with training ($\bar{x} = 21.78$), followed by WR without training ($\bar{x} = 20.69$). All of these groups scored significantly higher than the control group: Indy's mean was 19.6731. For the Composition 1, the results are clear: both groups using Google Translate, and especially those who had prior training in the use of

this tool, outperformed the groups using WordReference and the Indy group, which used no tool at all. In turn, the two groups using WordReference, and the students with training more so than those without it, did better than the group that did not use either tool.

Composition 2 - For this second experimental task, students were again allowed to use GT, WR, or neither, depending on which group they were in. The same pattern seen for the first experimental composition is confirmed for the second composition. GT+T ($\bar{x} = 25.05$) achieved the highest score, followed closely by GT-T ($\bar{x} = 24.19$). The next two highest means were that of WR+T ($\bar{x} = 21.32$), in this case scoring over two-and-a-half points lower on average than GT-T, and that of WR-T ($\bar{x} = 20.11$). Lastly, the independent group once again had the lowest mean score of the five groups ($\bar{x} = 19.67$). As with Composition 1, the Google Translate groups received higher mean scores than the WordReference groups; in both cases, the group with training in each tool significantly outperformed the group that had no such training.

Posttest - Approximately a week after the experimental tasks were completed, participants wrote a composition as a posttest measure to see if there were any changes in student performance after having written with or been trained in using one of the tools. A difference in means could indicate that using a tool with or without training might have an effect on students' subsequent performance when writing on their own. For the Posttest, as with the Pretest, no participants in any of the groups were allowed to use any tools.

Similar to the Pretest, most of the comparisons between groups were not statistically different. The three exceptions to this, however, were significant differences involving the Google Translate group that had received training. The means between the Indy group and GT+T were statistically different (barely reaching significance at $p = 0.050$), as were those of GT+T and GT-T ($p = 0.031$), and GT+T and WR-T ($p = 0.036$). In all three cases, the group who had used Google Translate after training did significantly worse than the other group mentioned. The GT+T scored lower than any other group ($\bar{x} = 18.20$); the WR+T group's mean ($\bar{x} = 19.04$) was also higher than GT+T's mean, but not to a statistically significant level. All comparisons among the Indy group, GT-T, WR+T, and WR-T were not statistically significant, indicating no meaningful difference among these groups' scores in the Posttest. Overall, seven of the ten pairwise comparisons showed no difference, with the only exceptions being three of the comparisons involving the GT+T group. These results merit further analysis in the Discussions section (Section 4).

Delayed Posttest - In order to judge any longer-term effects of the use of online translators and dictionaries, and training in each tool, a final writing task was administered three to four weeks after the Posttest. Any statistically significant differences could indicate longer-term effects of using these tools with or without training. Unlike with the Posttest, there were no differences ($p < 0.05$) found in the comparisons between any of the groups. The mean scores ranged from 19.19 (GT+T) to 20.08 (WR+T). At the end of the study, once online translators and dictionaries were no longer permitted, all groups had statistically similar scores. The means for all writing tasks are summarized in Table 1 below, while Table 2 shows the p values for each pairwise comparison. Further discussion of these and other results will be undertaken in the Section 4 below.

Table 1

Means on Compositions 1 through 5 by group

Group	Pretest	Composition 1	Composition 2	Posttest	Delayed Posttest
Indy	18.78	18.92	19.67	19.58	19.61
GT +training	19.99	**25.53**	25.05	**18.20**	19.19
GT –training	20.57	23.69	24.19	20.17	20.07
WR +training	20.88	21.78	21.32	19.04	20.08
WR –training	20.12	20.69	20.11	20.07	19.60

Note. Scores are out of 30 possible points. The highest and lowest mean across all tasks are bolded.

Table 2

Pairwise significance values for the five writing tasks

Pairwise Comparison	Pretest	Composition 1	Composition 2	Posttest	Delayed Posttest
Indy vs. GT+T	0.075	**4.413E-29**	**1.709E-17**	**0.050**	0.511
Indy vs. GT-T	**0.040**	**8.869E-09**	**4.699E-09**	0.448	0.577
Indy vs. WR+T	**0.005**	**6.285E-06**	**9.812E-02**	0.407	0.481
Indy vs. WR-T	0.138	**1.365E-02**	6.984E-01	0.527	0.994
GT+T vs. GT-T	0.530	**5.530E-03**	1.979E-01	**0.031**	0.311
GT+T vs. WR+T	0.248	**3.449E-11**	**2.436E-13**	0.235	0.193
GT+T vs. WR-T	0.904	**7.274E-14**	**3.952E-11**	**0.035**	0.616
GT-T vs. WR+T	0.754	**2.039E-02**	**1.125E-06**	0.163	0.987
GT-T vs. WR-T	0.688	**7.698E-04**	**2.745E-06**	0.913	0.632
WR+T vs. WR-T	0.461	**1.439E-01**	**3.456E-01**	0.199	0.565

Note. Significant results are bolded.

4. Discussions

This section will present an analysis of the results in relation to the research questions, a discussion of the appropriateness of these tools for second language learning, and the implications of the study for second language learning and instruction.

4.1 Use of online tools without training

The first research question was the following: "Does the use of electronic tools (online translators or dictionaries) affect the writing scores of intermediate second language learners?" The results of this study indicate a positive response to this question. Both pairwise comparisons on the two experimental tasks found significant differences between groups. As presented in Table 1, statistically significant differences were found on Compositions 1 and 2, and between the following pairs who did not have any training: Indy vs. GT-T, and Indy vs. WR-T. The Indy group scored the lowest of the three groups (18.92 on Composition 1; 19.67 on Composition 2), with WR-T achieving the next highest means (20.69 on Composition 1; 20.12 on Composition 2) and GT-T doing the best (23.69 on Composition 1; 24.19 on Composition), all results $p < 0.05$.

These results suggest that students who use an online translator may receive higher scores on their writing assignments as compared to their classmates who use an online dictionary or neither tool. These results confirm the tentative findings of O'Neill (2016), which reported that students using an online translator did not score worse than, and sometimes performed better than, students who did not use the OT. This finding is particularly important since many instructors prohibit OT use on writing assignments. Depending on the instructor's perspective, this result could be positive since it allows students to have better compositions overall. On the other hand, instructors who do not want their students to use online translation cannot use the argument that it will make compositions worse overall. Prohibiting OT use may not be feasible, however. If instructors could always determine whether or not an online translator had been used, one strategy might be to punish students who are attempting to gain an unfair advantage over their peers either by assigning a lower grade or pursuing sanctions related to academic dishonesty. However, previous research (O'Neill, 2013) found that instructors were only able to identify OT use correctly 70.7% of the time, meaning that nearly three out of ten times, instructors either wrongly suspected OT use or did not judge a composition to be helped by online translation even though it had been.

Given the stakes for students' grades and academic record, other strategies for dealing with online translation use are discussed in Sections 4.4 and 5. It is important to consider the fact that students using online dictionaries also outperformed the control group that used no tool, which is a new finding. For instructors who accept the fact that most students are going to use online tools for their language learning, but are less sure about

the role OTs should play, the results point towards ODs as an option that will help students significantly improve their performance on writing assignments. Instructors might encourage the use of online dictionaries to assist students with their writing or create in-class assignments where online tool use can be more closely monitored.

4.2 Use of online tools with training

The next research question, "Does prior training in the use of electronic tools (online translators or dictionaries) affect the writing scores of intermediate second language learners?", can also be responded to affirmatively. Two pairwise comparisons from Table 1 address this question. The group that used Google Translate after training outscored the group that used this tool without training, on Composition 1 ($\bar{x}$ = 25.53 versus $\bar{x}$ = 23.69) and Composition 2 ($\bar{x}$ = 25.05 versus $\bar{x}$ = 24,19). Training in how to use Google Translate, as well as information about potential strengths and weaknesses, appears to have allowed participants to use this tool more effectively. This finding again confirms the analysis of O'Neill (2016), which found that the group using online translation outperformed the control group on one of two writing tasks and did not do worse that the control group on the other. Since GT+T in the current study scored significantly higher than Indy on both experimental tasks, and also did better than the OT group without training, this finding strengthens the previous analysis and can more clearly state that training in online translation did have a positive effect on overall writing scores.

The same can be said for WordReference, with WR+T receiving a higher score on both compositions as compared to WR-T (Composition 1: $\bar{x}$ = 21.78 versus $\bar{x}$ = 20.69; Composition 2: $\bar{x}$ = 21.32 versus $\bar{x}$ = 20.11). As with the groups without training, however, WR+T performed less well than GT+T, suggesting that use of an online dictionary after training was beneficial to writing scores, but not as beneficial as using the online translator after training. The decision over which tool(s) to allow may once again come down to instructor preference. If the teacher is comfortable with students using online translation, training students in its use should be beneficial. Otherwise, ODs can also be somewhat beneficial, especially with prior training. In view of the fact that 87.7% of students already use OTs and ODs on graded work (O'Neill, 2019), these results suggest that those instructors who do not currently address these tools openly with their students might do so to improve how well students can use OTs or ODs when they write compositions on their own or in class.

4.3 Differences between scores for ODs and OTs

Another consideration that the results address concerns the performance of students who used an online dictionary as compared to those who used online translation. This analysis responds to the third research question: "Do learners who use online translation outscore those who use online dictionaries?"

Once again, the results are clear on this question. Based on the data obtained, and as suggested above, it is possible to respond in the affirmative. This finding applies regardless of whether or not prior training was given to students. The highest mean scores on Composition 1 were by GT+T ($\bar{x}$ = 25.53) and GT-T ($\bar{x}$ = 23.69), followed by WR+T ($\bar{x}$ = 21.78) and WR-T ($\bar{x}$ = 20.69). The two pairwise comparisons between GT groups and WR groups were significantly different, indicating that students who used the OT did indeed receive higher scores than those who used the OD. Similar results were found in the second composition, confirming this finding. No previous study was found in the literature that addressed the specific issue of comparing the effects of online dictionary use on L2 student writing to that of online translator use. If the results of this study are confirmed by future research, it appears that OTs help students to perform better when writing as compared to ODs.

Online dictionaries are sometimes touted by instructors as a way to improve student writing while representing an alternative to OTs. The results from Sections 4.1 and 4.2 would support the former motivation for recommending OD use, since participants who use WordReference did better on both compositions; however, recommending ODs over OTs is not supported by the data since the two groups using online translation scored

higher than the two groups using the OD. This observation puts aside for the moment the important consideration of the appropriateness of OT use, which will be discussed further in Sections 4.4 and 5.

4.4 Effect of ODs and OTs on future performance

One concern some instructors have is whether the use of online tools, in particular OTs, will negatively affect the future performance of students. Learners may become dependent on using online resources and rely on them rather than their own knowledge of the language. The fourth research question addresses this issue. Unlike the previous research questions, the results are mixed for the RQ4: "Are there any effects on learners' scores once they stop using the electronic tool (online translators or dictionaries) as compared to their peers who did not write using either tool?"

The immediate posttest, administered one week after Composition 2, shows that students who used WordReference did not do significantly worse than students who did not use the tool at all during the study. This finding holds true for both the WordReference group that had training and that which received no training. An analysis of the Delayed Posttest, given four to five weeks after Composition 2, also confirms that there was no significant difference between the Indy group and the two WordReference groups. These results suggest that instructors need not worry about detrimental effects on overall performance immediately after using the tool or approximately a month later.

The same cannot be said for online translation, for which results were mixed based on the group and task considered. The Posttest revealed that the Google Translate group that received training ($\bar{x}$ = 18.20) did significantly worse than three of the other groups: the Indy group ($\bar{x}$ = 19.58), Google Translate without training ($\bar{x}$ = 20.17), and WordReference without training ($\bar{x}$ = 20.07). Conversely, GT-T did not perform statistically worse than the other groups. The results are somewhat perplexing, although there is one possible explanation that can be offered. One reason that GT+T may have done worse than the Indy, GT-T, and WR-R groups might be related to the efficacy of the training. Participants in GT+T learned how to use Google Translate well, as compared to the three groups that received no training and lower scores on Compositions 1 and 2. It is possible that once there was no longer the ability to use Google Translate, students in GT+T could no longer put their training into use and had become more reliant on the tool they had begun to master. Participants in GT-T and WR-R had not learned to use the tool as effectively, and the Indy group never used either tool, so there was less of a negative effect on their writing in the Posttest, where no online tool was allowed. More research would be needed to confirm this result and the conjecture about it.

The results to the Delayed Posttest suggest that there was no longer-term effect on student performance after using online translation, for the GT-T group or GT+T. Once the tool was no longer allowed after some time had passed, students performed similarly to their peers. One possible explanation for the change in GT+T's results between the Posttest and the Delayed Posttest is that students once again became accustomed to writing without the OT and their scores were comparable to the other groups.

There are at least three possible interpretations of these results. First, the fact that none of the experimental groups performed significantly worse than the control group on the Delayed Posttest could suggest that instructors do not need to be overly concerned about the effects of online translator or dictionary use on future writing in the long term. Since students performed best using Google Translate after training, an argument could be made that instructors should embrace OTs and train students in their use. On the other hand, the situation may not be as clear cut. Since nearly nine out of ten students use OTs and ODs at least sometimes, it could be argued that the Delayed Posttest, while important for experimental design, might not represent a likely scenario on the ground: many students might not go a month without using these tools. Since GT+T was the only group that did worse than the others on the Posttest, one might argue that ODs or no tool at all should be promoted over OTs since students using OT, at least after training, may not do as well as their peers on assessments where the OT is not allowed, such as exams. Another interpretation would be that since most students use online translators and

dictionaries anyway, the most relevant results are those on the two experimental compositions, where students trained in OT and OD use outperformed those with no training or not using either tool. Instructors can openly discuss these tools, showing learners how they are used, what their strengths and weaknesses are, and encourage informed use of online tools to assist in students' writing.

A final perspective on the results for the Delayed Posttest can be considered. While students using online translators or dictionaries did not perform worse than the Indy group on this measure, they also did not perform better. The higher scores on the compositions may not be leading to increased learning. If the objective is to increase student learning, the results of this study do not support using OTs or ODs to do so. If, on the other hand, the goal is to allow students to express themselves better using their current knowledge and the assistance of OTs or ODs, then their use might be beneficial for encouraging effective communication.

4.5 Implications for instructors and learners

This study provides the first concrete data on two important topics: how training students can improve their use of online translators and dictionaries in L2 writing; and how OTs and ODs, two tools that are used by an overwhelming majority of language students, compare when used for writing compositions. The results of this study suggest several possible paths moving forward for instructors and students. Learners using modern technology, both online translators and online dictionaries, generally wrote better compositions than those not using either tool. If the goal is encouraging students to communicate effectively in writing in the target language, teachers may consider allowing the use of OTs and ODs for student composition assignments. The only case where participants in one of the groups using an online tool did significantly worse than participants who used neither tool was in the case of the immediate posttest for students using an OT without training once they were no longer allowed to use online translation. An initial reaction might be to reinforce the notion that OTs can be harmful for students; one path forward would then be for an instructor to continue to ban or at least discourage the use of OTs.

This take-away, however, is not supported by the data as a whole: students who received training in the use of OTs actually outperformed those not using this tool on both experimental tasks to a statistically significant degree. This group of students did not perform significantly worse on either the Posttest or Delayed Posttest once they were no longer allowed to use the tool, suggesting that their subsequent writing ability was not negatively affected. This result was also the case for students using ODs. The issue may not be whether or not the tools are used, but rather if students are using them effectively. With nearly nine out of ten language students using OTs and ODs, whether or not they are permitted, the most practical question that suggests itself is not if these tools should be allowed, but how best to assist students in using these tools in the most effective ways possible. The data support this path forward: the groups using online translators and online dictionaries after training significantly outscored both their counterparts who did not have such training, as well as the group that did not use either tool.

If one allows for the possibility of these tools being used by learners for foreign language writing, the results of the current study indicate that it would be incumbent upon teachers to instruct students in the responsible, effective use of online tools. This research included a relatively brief treatment (only 20 minutes in length) introducing the tool and involving students in the process of actively learning its features, its strengths and its weaknesses. A hands-on introduction at the beginning of the semester or school year could set the tone for responsible technology use among learners; interventions with additional practice, based on errors the instructor notes in student writing, may be called for to follow up and provide additional guidance. Undertaking such an approach could assist students in writing better than they would if left to their own devices, when they might not be aware of the limitations of the output provided by the tool and could overuse or ill use it.

Learners stand to benefit from such an approach. Instead of the current situation in many classrooms, where online translation use is either forbidden or tolerated in spite of the fact that most students are using it, students

could openly come to their instructor with questions related to the output of OTs instead of hiding their use. Even in the case of online dictionaries, for which instructors may initially be more receptive of student use, the findings of this study suggest that explicit instruction can assist students in learning how to use this resource well. Instead of technology being banned or begrudgingly tolerated, instructors and students alike can embrace technology to foster better writing among language learners.

4.6 Limitations and future research

As with any research, some limitations should be noted for the current study. The subject pool comes from one institution at the university level; results may not necessarily apply to students at other institutions or at the primary or secondary school levels. Since students chose whether or not to participate in the study, a selection bias may be present; students who were interested in completing the study may not be entirely representative of the general population of students. In spite of efforts to ensure validity, the fact that students were writing compositions in a computer lab as part of a proctored study may have affected their use of the online translator or online dictionary. In addition, given the rate of usage of these tools among foreign language students overall, it is likely that in a real-world situation some students may be using both tools on a given assignment; a follow-up study could investigate how students who are allowed to use both tools fare as compared to those using only one of the tools or none at all. As stated in Section 3.2, some caution should be noted when comparing the Indy group with GT-T, as well as the Indy group and WR+R, since the first group scored worse on the Pretest than the latter two groups. Since those groups were not statistically the same at the onset, one might expect them to not be the same in the Posttest and Delayed Posttest as well, which was not the case. Further research will be needed to confirm these two comparisons both on the Posttests as well as overall.

More research is called for to explore the effects of online translation and dictionary usage on students of different levels and different language pairs. An investigation into the impact of OTs and ODs on specific linguistic features (grammatical accuracy, vocabulary choice, etc.) could also prove to be fruitful. An analysis of the extent each tool is used by participants might provide insight into how much use of each tool on a given writing task is helpful or harmful on writing scores and student learning. Other tools, such as search engines (e.g. Yahoo search) or free online bilingual corpora (e.g. Linguee), could be investigated to see how the use of these tools by learners compares to that of OTs and ODs. Longer-term follow-up of students a semester, a year, or further into the future could track any residual effects of online translator or dictionary usage. The scope of investigation could be broadened beyond any effects on students' written expression, to reading comprehension in the L2 and the relatively new usage cases of automatic translation of spoken word and printed word provided by online translation apps on smartphones and tablets. There is much that can be and, given the prevalence of online tools, arguably needs to be investigated.

5. Conclusions

No one study or article can provide a definitive answer concerning the use of online translation or dictionaries by L2 learners. As the first experimental study to compare the scores of students using OTs and ODs, after prior training or no training, it is hoped that this research will advance the discussion of the place of online tools in the 21st-century classroom.

Participants in the study who used an OT with training performed significantly better than all other groups on the experimental writing tasks. One can conclude, at least tentatively, that training students in how to use online translation can be beneficial as compared to the current situation, where many instructors ban or do not discuss online translation with their students despite this tool's widespread use. Since there is no way to prevent students from using OTs outside of the classroom, training students in their use should allow students to utilize them more effectively. If the five-step approach from Section 2.3, or a similar training method, is undertaken, instructors can help guide students in the responsible use of the tool by showing strengths and weaknesses of OTs and giving some guidelines for their use, overall or for specific assignments. For example, students could be

advised to use online translation for short phrases but not entire sentences or paragraphs. Instructions could call for specific structures or vocabulary to be used in compositions, reducing the likelihood of students using an OT or at least encouraging them to do so "mindfully" (Knowles, 2016) instead of taking the results of the online translators as-is without further consideration or editing. Alternately, grading rubrics could also deemphasize grammatical accuracy so that students feel less pressured to turn to online translation for a better grade and be more likely to rely on their own knowledge of the language to accomplish the more important goal of communication (Ducar & Schocket, 2018).

Based on the findings of this study, online dictionaries are also a good option for writing, particularly when students are trained in their use. While participants using the OD did not perform quite as well as those using online translation, they still scored significantly higher on their compositions than students who did not use either tool. Since some questions remain about the ethicality of using online translation on graded work, and given the mixed results for students using online translation on the immediate Posttest, some teachers may be more comfortable with allowing students to use an online dictionary. This approach could work especially for in-class assignments, where the instructor's presence might encourage students to use only authorized tools. Training as well could emphasize effective ways of using online dictionaries and encourage responsible use, for example stressing that online dictionaries might be best for looking up occasional words or expressions but should not be used to translate a composition written first in the students' L1 word-by-word into the target language. Teachers can use or adapt the five-step strategy employed in the current study or create their own lessons introducing students to, and training students in, using online tools based on learner, instructor, or institutional needs.

There may never be a one-size-fits-all solution to the issue of online tools for second language learning. With an overwhelming majority of students now using online translation and dictionaries, the time has come for instructors to learn more about these tools, develop clear policies they are comfortable with concerning their use, and train students on if, how, and how much they should be using online resources to assist them in their coursework. In turn, by becoming informed on the advantages and disadvantages of OTs and ODs, and the potential effects of these tools on writing, learners can be empowered to make responsible choices in using modern technologies on their language learning journey.

6. References

Amirian, S., & Heshmatifar, Z. (2013). The impact of using electronic dictionary on vocabulary learning and retention of Iranian EFL learners. *International Journal of Research Studies in Educational Technology*, *2*(1), 1-10. https://doi.org/10.5861/ijrset.2013.384

Bahri, H., & Mahadi, T. S. 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. https://doi.org/10.7575/aiac.alls.v.7n.3p.161

Burton, C. (2003). The online translator: Implementing national standard 4.1. *Hispania, 86*(2), 320-21. https://doi.org/10.2307/20062865

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *Research in Education and Learning Innovation Archives, 10*, 108-121. https://doi.org/10.7203/attic.10.2228

Correa, M. (2014) Leaving the "peer" out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content & Language Integrated Learning*, *7*(1), 1-20. https://doi.org/10.5294/laclil.2014.7.1.1

Dashtestani, R. (2013). EFL teachers' and students' perspectives on the use of electronic dictionaries for learning English. *CALL-EJ, 14*(2), 51-65.

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

Elola, I., Rodríguez García, V., & Winfrey, K. (2008) Dictionary use and vocabulary choices in L2 writing. *ELIA, 8,* 63-89.

Fountain, C., & Fountain, A. (2009). A new look at translation: Teaching tools for language and literature. In C. Wilkerson (Ed.), *Empowerment through collaboration: Dimension 2009* (pp. 1–15). Valdosta, GA: Southern Conference on Language Teaching.

Fredholm, K. (2014). Effects of online translation on morphosyntactic and lexical-pragmatic accuracy in essay writing in Spanish as a foreign language. In *CALL Design: Principles and practice. Proceedings of the 2014 EUROCALL Conference* (pp. 96-101). Groningen, The Netherlands. https://doi.org/10.14705/rpnet.2014.000201

Fredholm, K. (2015). *El uso de traducción automática y otras estrategias de escritura* [The use of machine translation and other writing strategies]. *Estudios de Lingüística Aplicada*, *62*, 9-31. https://doi.org/10.22201/enallt.01852647p.2015.62.415

García, I., & Peña, M. I. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning, 24*(5), 471-487. https://doi.org10.1080/09588221.2011.582687

Jin, L., & Deifell, E. (2013). Foreign language learners' use and perception of online dictionaries: A survey study. *MERLOT Journal of Online Learning and Teaching*, 9, 515-532.

Jolley, J. R., & Maimone, L. (2015). Free online machine translation: Use and perceptions by Spanish students and instructors. In A. J. Moeller (Ed.), *Learn languages, explore cultures, transform lives* (pp. 181-200). Minneapolis: 2015 Central States Conference on the Teaching of Foreign Languages.

Knospe, Y. (2017). *Writing in a third language: A study of upper secondary students' texts, writing processes and metacognition* (Dissertation Thesis). Umeå University.

Knowles, C. L. (2016). *Investigating instructor perceptions of online machine translation and second language acquisition within most commonly taught language courses* (Unpublished doctoral thesis). The University of Memphis, Memphis, TN.

Larson-Guenette, J. (2013). It's just reflex now: German language learners' use of online resources. *Die Unterrichtspraxis, 46*(1), 62-74. https://doi.org/10.1111/tger.10129

Luton, L. (2003). If the computer did my homework, how come I didn't get an "A"? *French Review, 76*, 766-770.

McCarthy, B. (2004). Does online machine translation spell the end of take-home translation assignments? *CALL-EJ online, 6.1.* Retrieved from http://www.clec.ritsumei.ac.jp/english/callejonline/9-1/mccarthy.html

Niitemaa, M. L., & Pietilä, P. (2018). Vocabulary skills and online dictionaries: A study on EFL learners' receptive vocabulary knowledge and success in searching electronic sources for information. *Journal of Language Teaching and Research, 9*(3), 453-462. https://doi.org/10.17507/jltr.0903.02

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'Neill, E. M. (2013). Online translator usage in foreign language writing. *Dimension, 2013*, 74-88.

O'Neill, E. M. (2016). Measuring the impact of online translation on fl writing scores. *IALLT Journal of*

Language Learning Technologies, 46(2), 1-39. https://doi.org/10.17161/iallt.v46i2.8560

O'Neill, E. M. (2019). Online translator, dictionary, and search engine use among L2 students. *CALL-EJ, 20*(1), 154-177.

Stapleton, P. (2005). Using the web as a research source: Implications for L2 academic writing. *The Modern Language Journal*, *89*, 177–189. https://doi.org/0.1111/j.1540-4781.2005.00273.x

Steding, S. (2009). Machine translation in the German classroom: Detection, reaction, prevention. *Die Unterrichtspraxis / Teaching German*, *42*(2), 178–189. https://doi.org/10.1111/j.1756-1221.2009.00052.x

Tight, D. G. (2017). Tool usage and effectiveness among L2 Spanish computer writers. *ELIA: Estudios de Lingüística Inglesa Aplicada*, *17*, 157-182. https://doi.org/10.12795/elia.2017.i17.07

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

Williams, L. (2006). Web-based machine translation: A tool for promoting electronic literacy and language awareness. *Foreign Language Annals, 39*, 565-578. https://doi.org/10.1111/j.1944-9720.2006.tb02276.x

Wuttikrikunlaya, P., Singhasiri, W., & Keyuravong, S. (2018). The use of online tools in L2 writing: A study of Thai university students. *Journal of English Language Teaching and English Linguistics, 30*(1). Retrieved from http://culi.chula.ac.th/publicationsonline/files/article2/EOVAIjqqFhSat105231.pdf

Appendix: Scoring rubric

*Please **read** each composition and **evaluate it based on the following categories**. IMPORTANT — You should NOT let whether or not you think Google Translate or WordReference was used effect your score; just score it on its merits.*

Content — Are the ideas used sufficiently <u>creative</u> and <u>appropriate</u> to the subtopics outlined in the task?
5   Yes, the writer addressed the theme by covering all 3 subtopics
4   Yes, the writer addressed the theme by covering 2 of the 3 subtopics
3   The writer addressed the theme by covering 1 of the 3 subtopics
2   No, the writer did not address any of the subtopics, but addressed the overall theme
1   No, the writer did not address any of the subtopics, nor the theme

Grammar — Is <u>grammar</u> appropriate (tense, subject-verb agreement, number/gender agreements, articles, negation, etc.)
5   Yes, the grammar throughout is appropriate
4   Yes, the grammar is appropriate in the vast majority of cases
3   The grammar is appropriate about half of the time
2   No, the grammar is inappropriate in the vast majority of cases
1   No, the grammar throughout is inappropriate

Spelling / Accents — Are the <u>letters & accents</u> written as would be expected? (Note: This does <u>not</u> include conjugation.)
5   Yes, all or very nearly all of the words used are written perfectly
4   Yes, the vast majority of the words used are written correctly
3   About half of the words used are written correctly
2   No, the vast majority of the words used are written incorrectly
1   No, all or very nearly all of the words contain mistakes

Syntax — Is the <u>word order</u> used appropriate?
5   Yes, word order throughout is appropriate
4   Yes, word order is appropriate in the vast majority of cases
3   Word order is appropriate about half of the time
2   No, word order is inappropriate in the vast majority of cases
1   No, word order throughout is inappropriate

Vocabulary — Are the <u>words</u> <u>accurate</u> and <u>effective</u>? (This does <u>not</u> include spelling, accents, or grammatical usage.)
5   Yes, all or very nearly all of the words used are accurate and effective
4   Yes, the vast majority of the words used are accurate and effective
3   About half of the words used are accurate and effective
2   No, the vast majority of the words used are not accurate and effective
1   No, all or very nearly all of the words are not accurate and effective

Understandability— Could this writing be <u>understood</u> by a native Spanish speaker?
5   Yes, with no difficulty
4   Yes, with a little difficulty
3   Yes, with moderate difficulty
2   Yes, with great difficulty
1   No, it could not be understood

Note. The version of the scoring rubric for French instructors mentioned "French" instead of "Spanish."