



University of  
Sistan and Baluchestan



Uppsala University

## Exploring the Effects of AI-Assisted Translation on EFL University Students' Academic Writing Proficiency: A Longitudinal Study

Vahid Reza Mirzaeian<sup>1</sup>, Katayoun Oskoui<sup>2</sup>

<sup>1</sup>Associate Professor, Department of English, Faculty of Literature, Alzahra University, Tehran, Iran, Email: [mirzaeian@alzahra.ac.ir](mailto:mirzaeian@alzahra.ac.ir)

<sup>2</sup>PhD Candidate, Department of English, Faculty of Literature, Alzahra University, Tehran, Iran, Email: [oskoui@alzahra.ac.ir](mailto:oskoui@alzahra.ac.ir)

### Abstract

In spite of numerous studies on the effect of AIAT on foreign language learning, few have examined its impact on university students' academic writing proficiency over an extended period. This study, therefore, investigated the effect of Google Translate (GT) on the writing proficiency of English as a Foreign Language (EFL) university students. It examined fluency, lexical density, accuracy, and syntactic complexity across five versions of students' writings: the pretest, writing with the aid of GT, the posttest, and two retention tests conducted two to four months after the treatment. The findings indicated that syntactic complexity, accuracy, and fluency improved when students used GT for writing; however, these measures showed a significant decline in the posttest and retention test without GT. This decrease was more pronounced in lexical density and less so in syntactic complexity. Despite this decline, all factors assessed in the retention test still showed improvement compared to the pretest, indicating the positive effect of GT on students' foreign language writing performance.

**Keywords:** AI-assisted translation, google translate, academic writing, fluent language, lexical density, correct language use

Received: May 19, 2024

Revised: November 21, 2024

Accepted: November 29, 2024

Published: January 13, 2025

Article type: Research Article

DOI: 10.22111/ijals.2025.49435.2460

Publisher: University of Sistan and Baluchestan

© The Author(s).



How to cite: Mirzaeian, V. R., & Oskoui, K. (2025). Exploring the effects of AI-Assisted Translation on EFL university students' academic writing proficiency: A longitudinal study. *Iranian Journal of Applied Language Studies*, 17(2), 121-134. <https://doi.org/10.22111/ijals.2025.49435.2460>

## 1. Introduction

Recent studies on the application of AIAT to language learning indicate that most problems in AIAT output have been resolved, and that the technology is improving on a daily basis as it learns from user-reported errors and draws on a large language dataset (Ducar & Schocket, 2018; Lee, 2022a; Tsai, 2022). However, one persistent challenge in the field concerns less commonly taught languages, such as Persian (Mirzaeian, 2020, 2021, 2022, 2024; Mirzaeian & Maghsoudi, 2020, 2024; Mirzaeian & Oskoui, 2022, 2023). Another issue relates to translation irregularities, as discussed by Lee (2022a) and Shadiev et al. (2019). Thanks to improvements in AIAT output quality, the technology is gradually making its way into language classrooms and is being embraced by the language-learning community (Tsai, 2022). Research on the effects of AIAT on language learning has reported various outcomes, including positive impacts on content communication, lexical density, and syntactic complexity (Chang et al., 2022; Fredholm, 2019; Lee, 2022a; Niño, 2020; Tsai, 2019, 2022).

Overall, studies examining the impact of AIAT on language learning have been largely positive, indicating that this technological development can enhance language learning (Chon et al., 2021; Lee, 2020, 2022a; Tsai, 2019; Yang et al., 2023). However, almost all of these studies have focused on writing as a product rather than as a process. In other words, while students' written output showed significant improvement, it was unclear whether their writing proficiency itself had improved. One study that addressed this overlooked issue was Fredholm (2019), who demonstrated that when students were deprived of using AIAT, their lexical density dropped significantly. Building on this, the present study examined students' writing over a longitudinal period to determine whether the effects of AIAT were sustained over time.

## 2. Literature Review

Recent studies show a significant increase in the use of AIAT in language classes (Hwang et al., 2022; Jolley & Maimone, 2022; Tourmen & Hoffmann, 2022), a trend that can be attributed to its improved quality in recent years. Most research prior to 2018 focused on the errors generated by AIAT and on techniques to help students diagnose and learn from these errors. However, since 2018, such errors have become almost negligible, and studies have shifted to exploring the positive aspects of using correct AIAT output in language classrooms (Chon et al., 2021; Kol et al., 2018; Stapleton & Kin, 2019). AIAT has also demonstrated that when the input contains minor syntactic inaccuracies, it can intelligently infer the correct form and provide an appropriate translation. For example, in Persian, a clitic attaches to the verb to indicate the subject, and in most cases, the subject is omitted, creating ambiguity. AIAT can identify the implied subject and provide the correct translation (Mirzaeian, 2024). Such capabilities have led some researchers to conclude that AIAT can produce better output than human translation (Stapleton & Kin, 2019; Tsai, 2019, 2022; Yang

et al., 2023). This improved quality has also contributed to AIAT's growing popularity among language teachers (Hwang et al., 2022; Tsai, 2022).

Despite considerable research in the field, some gaps remain. One such area is the long-term effect of AIAT on language learning, typically explored through longitudinal studies. Fredholm (2019), for example, investigated the impact of AIAT on vocabulary and found that although AIAT use greatly enhanced lexical density, it had no effect on vocabulary acquisition; students deprived of AIAT could not employ varied lexical items in their writing. Additionally, while some researchers have reported that AIAT improves both accuracy and fluency in language use (Ahn, 2021; Lee, 2020), others have found that it enhances accuracy but not fluency (Yang et al., 2023) or vice versa (Fredholm, 2019). Furthermore, no such study has examined Persian EFL university students. Therefore, the present study investigates the effect of AIAT on both the writing quality and writing retention of EFL university students.

The following research questions were formulated for the current study:

1. What is the effect of AIAT on the quality of FL writings produced by Persian EFL university students in terms of fluent language use, lexical density, correct language use and syntactic complexity?
2. What is the longitudinal effect of AIAT on the quality of FL writings produced by Persian EFL university students in terms of fluent language use, lexical density, correct language use and syntactic complexity two and four months after exposure to AIAT?

### 3. Method

The study adopted a quantitative approach, with the data consisting of compositions produced by the participants. Each composition was scored, and the number of errors in each was also recorded.

#### 3.1. Instruments

First, the Oxford Placement Test was administered to ensure the homogeneity of the participants. Next, they were asked to write five compositions on a single topic, four in English and one in Persian.

#### 3.2. Participants

The participants were initially 82 Persian EFL freshman university students enrolled in a female-only state university in Tehran. All were native Persian speakers with prior exposure to English in high school. Before the study began, they completed the Oxford Placement Test. Those who scored within one point above or below the mean score ( $M = 59$ ) were selected. Based on the

pretest results, their English proficiency level was classified as beginner (A2 according to the CEFR).

### **3.3. Procedure**

After the placement test, the 59 selected students were asked to write four compositions in English (one pretest, one while consulting AIAT, one two months after the study, and one four months after the study) and one in Persian, their native language. They were instructed to include at least 120 words in each composition. In the pretest, participants were not allowed to use any resources other than pen and paper and were asked to write a composition about the person they respected most and why. Next, they were given access to GT and instructed to write the same composition in Persian, translate it into English using GT, compare and contrast their own writing with the AIAT output, make the necessary corrections and adjustments, and submit the final draft. One week later, they were asked to write a composition on the same topic in class without access to AIAT (posttest).

To examine the longitudinal effect, 31 of the students were required to write the same composition again two months after the treatment, and the remaining 28 were asked to do so four months after the study. Participants were divided into two groups for the longitudinal stage to minimize the learning effect from the first retention test. Since the focus of the study was on form rather than content, during the retention stage participants were given their Persian compositions to recall what they had previously included.

### **3.4. Data Analysis**

As mentioned earlier, participants' compositions were evaluated based on four main criteria: fluent language use, lexical density, correct language use, and syntactic complexity. Fluent language use was measured using the total word count (Chon et al., 2021; Yang et al., 2023). Correct language use was assessed by identifying errors in syntax, lexis, spelling, and punctuation. To determine lexical density, both the textual lexical density and the frequency of words per composition were calculated (Yang et al., 2023). Syntactic complexity was evaluated by calculating the mean number of words per sentence (Lee, 2020) and weighting the distance between sentences (Chon et al., 2021; Yang et al., 2023).

For correct language use, two independent raters analyzed the frequency and distribution of errors. The interrater reliability coefficient was 0.91, indicating high agreement. For fluent language use, lexical density, and syntactic complexity, Coh-Metrix 3.0 was used, following previous research (Chon et al., 2021; Graesser et al., 2004; McNamara et al., 2014; Yang et al., 2023). After the initial analysis, inferential statistics were applied to compare the compositions across several conditions: two-month versus four-month retention tests; pretest versus retention test; posttest versus retention

test; and AIAT-assisted writing versus posttest performance. IBM SPSS (version 26) was used for all statistical analyses. Descriptive statistics were computed to examine overall patterns in the data. To assess changes across compositions, repeated-measures ANOVA was conducted to test for statistically significant differences among the four compositions. Paired-sample t-tests were then employed for pairwise comparisons. In the repeated-measures ANOVA, Mauchly's test of sphericity was performed, and the homogeneity of variances for the between-subjects ANOVA was confirmed ( $p=.850$ ).

## 4. Results

The results clearly indicated statistically significant differences among the compositions. Students produced higher-quality compositions when consulting AIAT, as well as in the posttest and retention tests, compared to the pretest, in terms of both fluent language use and correct language use. Table 1 presents these results ( $p < .001$ ). In the AIAT-assisted compositions and subsequent tests, word counts increased substantially, while the number of errors decreased significantly. Regarding overall scores, the AIAT-assisted composition ranked highest, followed by the posttest, the retention test, and the pretest, respectively. Analysis of lexical density and syntactic complexity revealed similar patterns, further confirming the positive impact of AIAT.

**Table 1**

*Descriptive Statistics and Repeated Measure ANOVA: (Fluent Language Use, Lexical Density, Correct Language Use, Syntactic Complexity)*

|                        | Item                    | Pretest |       | AIAT   |       | Posttest |       | Retention |       | F      |
|------------------------|-------------------------|---------|-------|--------|-------|----------|-------|-----------|-------|--------|
|                        |                         | Mean    | SD    | Mean   | SD    | Mean     | SD    | Mean      | SD    |        |
| Score                  |                         | 1.6     | 0.96  | 4.82   | 0.69  | 3.65     | 1.65  | 2.5       | 1.49  | 114.54 |
| Fluent Use of Language | Word Count              | 43.01   | 32.59 | 123.58 | 19.34 | 103.19   | 40.43 | 73.01     | 45.65 | 118.88 |
| Lexical Density        | Different Words         | 18.25   | 22.16 | 73.42  | 11.90 | 59.85    | 24.49 | 47.19     | 25.98 | 154.96 |
|                        | Textual lexical density | 32.87   | 22.88 | 57.80  | 18.94 | 53.40    | 20.76 | 48.57     | 27.06 | 24.84  |
| Correct language use   | Total No of errors      | 8.40    | 4.59  | 1.94   | 1.83  | 3.72     | 4.12  | 5.44      | 6.27  | 10.98  |
| Syntactic Complexity   | Words per sentence      | 4.93    | 2.51  | 11.65  | 3.06  | 8.82     | 3.21  | 7.17      | 3.93  | 57.45  |
|                        | Weighted distance       | 0.56    | 0.39  | 0.85   | 0.10  | 0.64     | 0.41  | 0.62      | 0.36  | 6.47   |

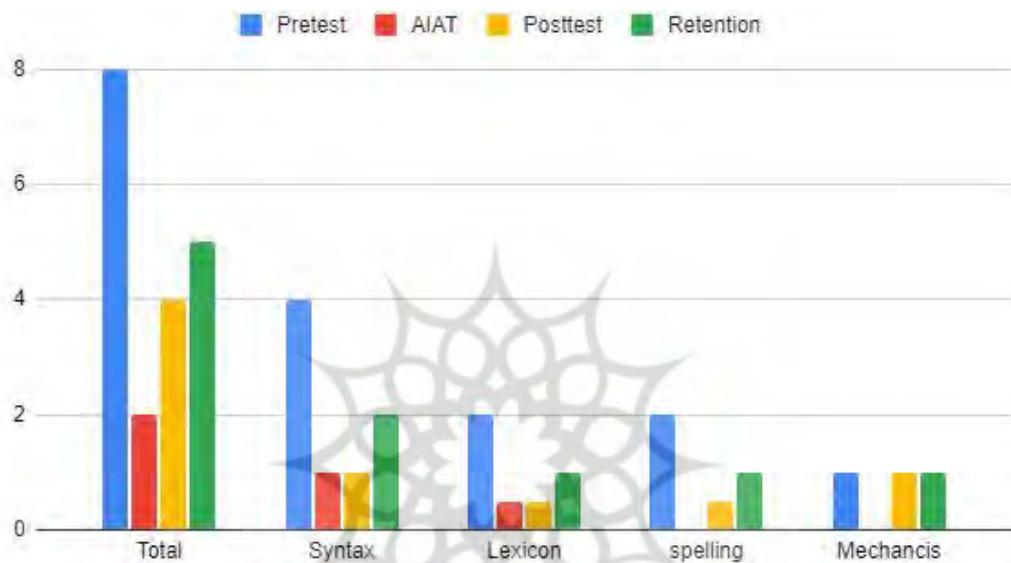
Level of significance at  $<0.001$

Errors related to syntax, lexicon, spelling, and mechanics were also analyzed. The results indicated that syntactic errors were the most frequent ( $M=2.17$ ), followed by lexical errors ( $M=0.98$ ), spelling errors ( $M=0.89$ ), and mechanical errors ( $M=0.83$ ), respectively. The pretest recorded the highest number of errors ( $M=8.41$ ), followed by the posttest ( $M=3.73$ ) and the AIAT-assisted compositions ( $M=1.95$ ). Retention tests showed a higher error rate ( $M=5.45$ ), as depicted

in Figure 1. Notably, the number of syntactic errors decreased sharply from the pretest to the other composition types. To determine whether these differences were statistically significant, an ANOVA was conducted. As shown in Table 2, the results indicated significant differences in all error categories except for lexical errors. Mechanical errors decreased in the AIAT-assisted compositions but increased in subsequent tests.

**Figure 1**

*Error Analysis Result*



**Table 2**

*Descriptive Statistics and Repeated Measure ANOVA: (Correct Language Use)*

| Unit      | Pretest |      | AIAT |      | Posttest |      | Retention |      | F    | p      |
|-----------|---------|------|------|------|----------|------|-----------|------|------|--------|
|           | Mean    | SD   | Mean | SD   | Mean     | SD   | Mean      | SD   |      |        |
| Syntax    | 4.23    | 2.61 | 1.05 | 1.21 | 1.27     | 1.48 | 2.18      | 2.77 | 9.05 | <0.001 |
| Lexicon   | 1.78    | 1.38 | 0.57 | 1.17 | 0.65     | 0.85 | 0.98      | 1.39 | 2.26 | 0.119  |
| Spelling  | 1.36    | 0.53 | 0.24 | 0.66 | 0.65     | 1.48 | 1.11      | 1.65 | 4.41 | 0.026  |
| Mechanics | 1.07    | 1.74 | 0.12 | 0.49 | 1.19     | 2.15 | 1.21      | 2.91 | 6.47 | <0.005 |

To further investigate the differences among compositions and assess the effect of AIAT on subsequent writing after the treatment, paired t-tests were conducted. As shown in Table 3, the comparison between the AIAT-assisted composition and the posttest revealed significant differences in mechanics, overall score, and word count, whereas no significant differences were found for syntax, lexicon, or spelling. The paired t-test between the AIAT-assisted composition and the retention test indicated statistically significant differences in all features except lexicon. These results suggest that correct language use in terms of lexicon remained consistent between the posttest and retention test, while correct language use in other areas, including fluent language use,

declined substantially. In contrast, the paired t-test between the pretest and retention test showed significant differences in all areas except mechanics.

**Table 3***Paired t-test between Compositions in Correct Language Use*

|           | AIAT vs Posttest |        | AIAT vs Retention |        | Pretest vs Retention |        |
|-----------|------------------|--------|-------------------|--------|----------------------|--------|
|           | t                | p      | t                 | p      | t                    | P      |
| Syntax    | -0.732           | 0.466  | -3.241            | 0.001  | -3.945               | <0.001 |
| Lexicon   | 0.138            | 0.891  | -1.511            | 0.135  | -3.321               | 0.002  |
| Spelling  | -1.668           | 0.100  | -2.661            | 0.011  | -4.008               | <0.001 |
| Mechanics | -3.833           | <0.001 | -3.201            | 0.001  | -1.867               | 0.66   |
| Score     | 5.816            | <0.001 | 11.872            | <0.001 | 12.813               | <0.001 |

Using paired t-tests, the study also examined whether statistically significant differences existed between the two-month and four-month retention tests. As shown in Table 4, a significant difference was observed only in syntax. While minor differences were noted in other aspects between the two retention compositions, these did not reach statistical significance. This finding indicates that syntactic errors increased significantly from the two-month to the four-month retention test, whereas other features of writing quality remained relatively stable over time.

**Table 4***Descriptive Statistics and T-Test for Retention (Two and Four Weeks)*

| Item       | Two weeks |       | Four weeks |       | t     | P     |
|------------|-----------|-------|------------|-------|-------|-------|
|            | Mean      | SD    | Mean       | SD    |       |       |
| Syntax     | 1.48      | 2.02  | 3.01       | 3.29  | 2.18  | 0.034 |
| lexicon    | 0.93      | 1.51  | 1.01       | 1.22  | 0.171 | 0.865 |
| spelling   | 0.82      | 1.43  | 1.42       | 1.83  | 1.407 | 0.166 |
| Mechanics  | 1.18      | 3.17  | 1.21       | 2.61  | 0.046 | 0.965 |
| Score      | 2.82      | 1.48  | 2.26       | 1.51  | 1.452 | 0.151 |
| Word count | 82.53     | 41.64 | 65.01      | 47.98 | 1.482 | 0.142 |
| Total      | 4.42      | 5.96  | 6.64       | 6.58  | 1.365 | 0.179 |

## 5. Discussion

This study demonstrated that AIAT can significantly enhance various aspects of students' foreign language writing. Improvements were observed across multiple measures: fluent language use (word count), correct language use (total and category-specific errors in syntax, lexicon, spelling, and mechanics), syntactic complexity (weighted sentence distance and words per sentence), lexical density (variety of lexical items), and overall message conveyance. Error frequency also declined across all categories, and repeated-measures ANOVA indicated statistically significant differences in scores between compositions. As noted by previous scholars (Garcia & Peña, 2011; Chang et al., 2022; Tsai, 2022), elementary-level students in this study

benefited from AIAT, as it enabled them to express ideas they could not otherwise articulate in the target language.

The findings further revealed that AIAT-assisted compositions showed the greatest improvement in fluent language use, correct language use, and syntactic complexity. However, given the participants' low proficiency, such gains may reflect reliance on AIAT rather than genuine language acquisition. This concern aligns with studies suggesting that AIAT use does not necessarily lead to lasting language improvement (Chen et al., 2015; Tang & Liu, 2018; Zhang & Cheng, 2021). Because AIAT does not offer targeted feedback on students' errors, beginner learners may struggle to identify and correct mistakes solely by comparing their writing with AIAT-generated translations (Lee, 2022 b). Consequently, it is plausible that participants in this study reproduced AIAT's output directly in the posttest, raising the question of whether such performance gains represent authentic learning or mere imitation.

The second research question:

The present study found that students retained approximately half of the material they had been exposed to through AIAT, indicating a positive long-term impact. Retention test results showed higher scores and greater word counts compared to the pretest, although declines were observed in syntactic complexity, fluent language use, and correct language use. Notably, the retention compositions contained significantly more words per sentence and a greater variety of vocabulary items. Given that these learners had previously lacked opportunities to fully express themselves in writing, AIAT appeared to boost their confidence in producing written work. This finding aligns with Fredholm's (2021) assertion that AIAT can enhance learners' self-confidence and self-esteem in writing in the target language.

When compared to earlier research, this study produced both convergent and divergent results. Consistent with prior findings (Chang et al., 2022; Yang et al., 2023), AIAT in the present study improved fluent language use, correct language use, and syntactic complexity. However, unlike Chang et al.'s (2022) study, where AIAT did not increase word count—the current study found a significant increase. This discrepancy may be explained by differences in participants' proficiency levels; whereas Chang's participants were intermediate learners, the participants in this study were predominantly beginners, who could produce longer sentences only when supported by AIAT. The results also corroborate Fredholm's (2015, 2019, 2021) findings that AIAT can enhance lexical density, fluent language use, and syntactic complexity. However, the current study diverges from Fredholm's (2015) work in two key respects: first, correct language use improved significantly here, both with AIAT and in the posttest, whereas Fredholm found no such improvement; second, the positive effect of AIAT on lexical density persisted in this study even without AIAT, whereas Fredholm observed its disappearance in the absence of the tool. These differences may stem from variations in implementation strategies, baseline levels of correct language use, or the language pairs involved (Spanish-Swedish in Fredholm's study vs. Persian-English here). Additionally, in Fredholm's (2015) research, students wrote on different topics in the pretest and posttest, while in

in the current study the same topic was used for both, potentially contributing to greater lexical retention.

## 6. Conclusion

This study demonstrated that AIAT can help language learners reduce syntactic errors, use a more diverse range of lexical items, and employ grammatical structures beyond their current level of proficiency. Consequently, students were able to produce richer output in both the AIAT-assisted compositions and the posttest compared to the pretest. The study examined the effect of AIAT on EFL students in terms of fluent language use, syntactic complexity, and lexical density, as well as the extent to which these gains were retained in a follow-up assessment. The results indicated that the application of AIAT was effective to some extent over a long period; however, its impact varied across different linguistic features.

Based on these findings, several directions for future research are proposed. While the present study measured effects over a 2–4-month period, further research could investigate whether improvements persist over longer durations (e.g., 6–12 months). Although Google Translate was the focus here, comparative studies involving multiple AIAT tools could determine whether differences in accuracy, feedback, or features influence learning outcomes. Exploring how teachers can incorporate AI tools such as Google Translate into classroom practices to balance assistance with independent learning, without fostering over-reliance, also merits investigation. Additionally, examining whether providing corrective feedback alongside AIAT outputs enhances long-term improvements in writing accuracy would offer valuable insights.

The current study observed a post-treatment decrease in performance without AIAT, suggesting the need for research on whether over-reliance on such tools hinders students' ability to self-correct. Given the observed decrease in lexical density, further studies could examine whether AIAT tools promote genuine vocabulary acquisition or encourage repetitive language use. Investigating how effects vary across different proficiency levels could identify which learner groups benefit most. Since this study focused on academic writing, future research could also explore the impact of AIAT tools on other genres, such as creative, descriptive, or argumentative writing. Beyond linguistic proficiency, examining how AI tools influence learner psychology, such as motivation, self-efficacy, or writing apprehension, could provide additional perspectives. Finally, to mitigate dependency, research should explore instructional strategies that teach students to use AIAT tools effectively while fostering autonomy in writing.

## References

Arifin, S. (2017). L2 writing strategies used by EFL graduate students. *Journal of ELT Research*, 2(2), 115–129. [https://doi.org/10.22236/JER\\_Vol2Issue2pp115-129](https://doi.org/10.22236/JER_Vol2Issue2pp115-129)

Bitchener, J., & Koch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37(2), 322–329. <https://doi.org/10.1016/j.system.2008.12.006>

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Multilingual Matters.

Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses. *JALT CALL Journal*, 14(1), 2–24.

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.

Chang, P., Chen, P.-J., & Lai, L.-L. (2022). Recursive editing with Google Translate: The impact on writing and error correction. *Computer Assisted Language Learning*, 37(7), 2116–2141. <https://doi.org/10.1080/09588221.2022.2147192>

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>

Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine translated texts: Raters' assessment, linguistic complexity, and errors. *System*, 96, 1–12. <https://doi.org/10.1016/j.system.2020.102408>

Chung, E., & Ahn, S. (2022). The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer Assisted Language Learning*, 35(9), 2239–2264 <https://doi.org/10.1080/09588221.2020.1871029>

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>

ETS. (2021). Writing scoring guide. Retrieved from <https://www.ETS.org/pdfs/toefl/toefl-ibt-writingrubrics-enhanced.pdf>

Ferris, D. (2011). *Treatment of error in second language student writing*. The University of Michigan Press.

Ferris, D. F., & Hedgcock, J. S. (2014). *Teaching L2 composition: Purpose, process, and practice* (3rd Ed.). Routledge.

Fredholm, K. (2015). Online translation use in Spanish as a foreign language essay writing: Effects on fluency, complexity and accuracy. *RevistaNebrija de LingüísticaAplicada a la Enseñanza de las Lenguas*, 9, 1–18.

Fredholm, K. (2019). Effects of Google translate on lexical diversity: Vocabulary development among learners of Spanish as a foreign language. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*, 13, 98–117. <http://doi.org/10.26378/rnlael1326300>

Fredholm, K. (2021). Google Translate search strategies among learners of Spanish L3: A complex lexico-morpho-syntactic weave of trial-and-error. *Estudios de Lingüística Aplicada*, 39(72), 9–48. <http://doi:10.26378/rnlael918248>

Garcia, I., & Peña, M. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning*, 24(5), 471–487. <http://doi.org/10.22201/enallt.01852647p.2021.72.926>

Graesser, A., McNamara, D., Louwerse, M., & Gai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers*, 36(2), 193–202. <https://doi.org/10.3758/BF03195564>

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>

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.

Kol, S., Schcolnik, M., & Spector-Cohen, E. (2018). Google Translate in academic writing courses? *The EuroCALL Review*, 26(2), 50–57. <http://doi:10.4995/eurocall.2018.10140>

Koltovskaia, S. (2022). Postsecondary L2 writing teachers' use and perceptions of Grammarly as a complement to their feedback. *ReCALL*, 35(3), 1–15. <https://doi.org/10.1017/S0958344022000179>

Lee, I. (2019). Teacher written corrective feedback: Less is more. *Language Teaching*, 52(4), 524–536. <http://doi:10.1017/S0261444819000247>

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>

Lee, S.-M. (2022a). 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>

Lee, S.-M. (2022b). L2 learners' strategies for using machine translation as a personalized writing assisting tool. In J. Colpaert & G. Stockwell (Eds.), *Personalization, contextualization, and socialization* (pp. 184–206). Castledown Publishers.

Lee, S.-M., & Briggs, N. (2021). Effects of using machine translation to mediate the revision process of Korean university students' academic writing. *ReCALL*, 33(1), 18–33. <https://doi.org/10.1017/S0958344020000191>

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press.

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>

Mirzaeian, V. (2020). Machine Translation Output Assessment and its Impact on Reading Comprehension. *Technology of Education Journal*, 14(2), 697–710.  
<https://doi.org/10.22061/ite.2019.5473.2227>

Mirzaeian V. (2022). Scientific study of the application of Machine Translation in language teaching and meta-analysis of its effect. *Technology of Education Journal*, 16(2), 384-400.  
<https://doi.org/10.22061/tej.2022.8267.2642>

Mirzaeian, V. (2024). The Effect of Artificial Intelligence Generated Translation versus Human Translation on Reading Comprehension of the Speakers of Less Commonly Taught Languages. *Iranian Journal of Applied Language Studies*, 16(1), 63-74.  
<https://doi.org/10.22111/ijals.2024.47092.2395>

Mirzaeian V. and Maghdoudi, M. (2024). Benchmarking Machine Translation Efficacy for Teaching EAP Reading Comprehension Skills. *Technology of Education Journal*, 18(1), 1-18.  
<https://doi.org/10.22061/tej.2023.9892.2913>

Mirzaeian V. and Oskoui, K. (2022). Investigating Iranian EFL Student Teachers' Attitude toward the Implementation of Machine Translation as an ICALL Tool. *Journal of English Language Teaching and Learning*, 14(30), 165-179. <https://doi.org/10.22034/elt.2022.52038.2496>

Mirzaeian V. and Oskoui, K. (2023) GTALL: A GNMT model for the Future of Foreign Language Education. *Issues in Language Teaching*, 11(2), 129-159.  
<https://doi.org/10.22054/ilt.2023.69268.724>

Mirzaeian V., & Oskoui, K. (2023) Google Translate in Foreign Language Learning: A Systematic Review. *Applied Research on English Language*, 12(2), 51-84.  
<https://doi.org/10.22108/are.2023.134264.1949>

Niño, A. (2004). Recycling MT: A Course on Foreign Language Writing via MT post. In *Proceedings of 7th Annual CLUK Research Colloquium* (pp. 179–187). <https://mt-archive.net/CLUK-2004-Nino.pdf>

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>

Niño, A. (2020). Exploring the use of online machine translation for independent language learning. *Research in Learning Technology*, 28, 1–32. <https://doi.org/10.25304/rlt.v28.2402>

O'Neill, E. M. (2016). Measuring the impact of online translation on FL writing scores. *The IALLT Journal of Language Learning Technologies*, 46(2), 1–39.  
<https://doi.org/10.17161/iallt.v46i2.8560>

O'Neill, E. M. (2019). Training students to use online translators and dictionaries: The impact on second language writing scores. *International Journal of Research Studies in Language Learning*, 8(2), 47–65. <http://doi:10.5861/ijrsll.2019.4002>

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29(1), 67–100.  
<https://doi:10.1017/S0272263107070039>

Shadiev, R., Sun, A., & Huang, Y.-M. (2019). A study of the facilitation of cross-cultural understanding and intercultural sensitivity using speech-enabled language translation technology. *British Journal of Educational Technology*, 50(3), 1415–1433. <https://doi.org/10.1111/bjet.12648>

Shin, D., & Chon, Y. V. (2023). Second language learners' post-editing strategies for machine translation errors. *Language Learning & Technology*, 27(1), 1–25. <https://hdl.handle.net/10125/73523>

Shintani, N., & Aubrey, S. (2016). The effectiveness of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. *The Modern Language Journal*, 100(1), 296–319. <https://doi.org/10.1111/modl.12317>

Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, 72, 118–32. <https://www.jstor.org/stable/43104426>

Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15, 165–79. <https://www.jstor.org/stable/44487616>

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>

Tang, C., & Liu, Y. (2018). Effects of indirect coded corrective feedback with and without short affective teacher comments on L2 writing performance, learner uptake, and motivation. *Assessing Writing*, 35, 26–40. <https://doi.org/10.1016/j.asw.2017.12.002>

Tourmen, C., & Hoffmann, D. (2022). A "hands-on" approach to raise awareness of technologies: A pilot class and its lessons. *L2 Journal*, 14(1), 237–257.  
<http://repositories.cdlib.org/uccllt/l2/vol14/iss1/art10>

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>

Tsai, S.-C. (2022). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*, 35(5–6), 1250–1272. <https://doi.org/10.1080/09588221.2020.1799412>

Valijärvi, R.-L., & Tarsoly, E. (2019). Language students as critical users of Google Translate: Pitfalls and possibilities. *Practitioner Research in Higher Education*, 12(1), 61–74.

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://www.jstor.org/stable/unteteacgerm.46.2.230>

Yang, Y., Wei, X., Li, P., & Zhai, X. (2023). Assessing the effectiveness of machine translation in the Chinese EFL writing context: A replication of Lee (2020). *ReCALL*, 35(2), 1–14. <https://doi.org/10.1017/S0958344023000022>

Zhang, L., & Cheng, X. (2021). Examining the effects of comprehensive written corrective feedback on L2 EAP students' linguistic performance: A mixed-methods study. *Journal of English for Academic Purposes*, 52, 1-15. <https://doi.org/10.1016/j.jeap.2021.101043>

