BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Can ESL Instructors Spot AI Translation? Evidence from Arabic-English Classroom

Anas Alkhofi

Department of English, College of Arts, King Faisal University, Hofuf 36362, Saudi Arabia

ABSTRACT

This study investigates the ability of ESL instructors to differentiate between AI machine-generated and student-generated translations and assesses their confidence in doing so. Twenty instructors evaluated 44 translations (22 student-generated, 22 machine-generated), classifying each as either machine- or student-produced. In total, 434 valid responses were analyzed using a Generalized Linear Mixed Model (GLMM) in R. The responses were coded as 0 (incorrect/unsure) and 1 (correct) to determine whether instructors' correct identification of machine-generated translations was statistically significant. The results revealed a low identification rate, with instructors having only a 28% probability of correctly distinguishing machine translations, which was significantly below the 50% expected by random chance. Despite this low success rate, 90% of instructors expressed confidence in their ability to detect machine translation. The findings suggest a gap between instructors' perceived and actual abilities, indicating that reliance on instructors' judgments may no longer be sufficient for detecting machine translation use. Moreover, instructors may need to prioritize in-class assessments, especially when the focus is on tapping into students' raw abilities, while take-home assignments should incorporate active post-editing to foster critical engagement with machine translation. These adjustments are crucial for maintaining academic integrity, equipping students with the skills needed to navigate both human and machine-generated translations, and preparing them for the evolving translation marketplace.

Keywords: AI in TESOL; Machine Translation in Teaching English; Google Translate in the Classroom; English-Arabic Translation

*CORRESPONDING AUTHOR:

Anas Alkhofi, Department of English, College of Arts, King Faisal University, Hofuf 36362, Saudi Arabia; Email: amalkhofi@kfu.edu.sa

ARTICLE INFO

Received: 4 February 2025 | Revised: 20 February 2025 | Accepted: 24 February 2025 | Published Online: 2 March 2025

DOI: https://doi.org/10.30564/fls.v7i3.8652

CITATION

Alkhofi, A., 2025. Can ESL Instructors Spot AI Translation? Evidence from Arabic-English Classroom. Forum for Linguistic Studies. 7(3): 340–350. DOI: https://doi.org/10.30564/fls.v7i3.8652

COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

The integration of artificial intelligence (AI) with machine translation around 2014 led to the development of neural machine translation, which has significantly improved the quality of machine-generated translations [1]. Neural machine translation utilizes deep neural networks, a form of AI, to learn linguistic patterns from large corpora and predict translations. Neural machine translation functions as a fully integrated system, autonomously acquiring semantic meanings and translation knowledge directly from its training data [1]. Since the transition to neural machine translation, translation quality has been steadily improving. For instance, Hassan et al. [2] found that their Chinese-to-English neural machine translation achieved human-level quality in sentence-level translations. Similarly, Wang et al. [1] noted that such translations more closely mimic human-like fluency. Google Translate (GT) adopted neural machine translation in 2016, making its translations increasingly natural and accurate [3].

As improvements in machine translation technology have progressed, there has been an increasing reliance on these tools by English as a second language (ESL) students across various educational contexts and for diverse purposes [4, 5]. Additionally, research shows that students frequently use machine translation tools, even when instructed not to do so [6, 7]. This raises important questions as to how educators can navigate this evolving reality. AI-based machine translation is a relatively new and rapidly evolving field, and its integration into ESL education remains a fertile ground for research. A key issue is whether instructors can effectively detect AI-based machine translations in ESL students' work. While several studies have assessed instructors' ability to identify AI-generated writings in general [8–10], few have focused specifically on instructors' ability to distinguish between AI-generated and student-generated translations.

Although AI detector tools exist, they are not always reliable as AI-based machine translation systems continue to evolve and produce increasingly human-like outputs [11]. Research has shown that AI detectors frequently yield false positives and negatives [12–14]. Furthermore, the general assumption among instructors is that they can detect AI-generated text, but what if they cannot? If instructors are unable to detect AI content, this raises concerns about the validity of current assessment methods. In such cases, integrating AI tools into the classroom for guided instruction, rather than banning their use, may be a more effective approach. Also, a shift toward in-class assessments rather than take-home assignments could help ensure that students' language proficiency is being accurately evaluated.

Given that GT is both the most widely used AI-based machine translation tool among students [15] and the most commonly researched platform [4], the current study focuses on investigating instructors' ability to distinguish between machine-generated and student-generated translations using GT as the primary tool for comparison. The present research is guided by the following research questions: 1) Can ESL instructors detect whether a translation has been generated by a student or by a machine? 2) How confident are instructors in their ability to differentiate between student-generated and machine-generated translations?

2. Literature Review

2.1. Instructors' Perceptions of Machine Translation Tools

Given that the objective of this study is to investigate instructors' ability to differentiate between machine and ESL student translations, it is important to consider the existing literature on instructors' perceptions of and engagement with machine translation tools in the context of second language education, as well as their familiarity with these tools.

In terms of use and familiarity, the literature suggests that a significant portion of instructors regularly use machine translation tools for personal or teaching purposes, though the frequency of use varies. Ata and Debreli [16] found that 26% of Turkish EFL instructors never used machine translation, and among those who did use the tools, a majority (82.4%) used machine translation only a few times per month, indicating limited engagement. In contrast, Jolley and Maimone [17] reported higher usage rates among Spanish instructors, with 82.05% using free online machine translation tools for personal or teaching purposes, though very few (7.69%) assigned tasks requiring their students to use machine translation tools, and a majority of instructors (73%) had never received formal machine translation-related training in the past. Meanwhile, Liu et al. [18] highlighted that 53.4% of instructors understood the machine translation mechanism, and 60% could use it proficiently.

When examining perceptions of machine translation quality, studies suggested that instructors viewed machine translation as particularly accurate for shorter segments of text such as individual words, and less reliable for longer segments [17]. Moreover, both Liu et al. [18] and Ata and Debreli [16] reported higher instructor satisfaction with machine translation when translating from English, as opposed to translating into English (such as English to Turkish/Chinese versus Turkish/Chinese to English), perhaps due to machine translation systems being better equipped to decode English source texts, given the larger amount of training data available in English.

As for ethical concerns, a consistent finding across the studies was that ethicality largely depends on the context in which machine translation is used. Ata and Debreli [16] found that 70% of instructors believed that the ethicality of machine translation depended on how it was used in assignments, a sentiment echoed by Jolley and Maimone [17], with 82% of instructors agreeing that context determined whether machine translation use constituted cheating. Meanwhile, Liu et al. [18] showed that 46% of instructors were concerned about the ethical implications of using machine translation in graded assignments, while 47% believed that ethicality depended on the extent of modifications made to the machine-translated text, with greater post-editing being associated with higher ethical acceptability.

Despite these ethical concerns, there was broad agreement that machine translation use in translation education is unavoidable and should be embraced with appropriate guidelines. Liu et al. [18] found that 80% of instructors supported integrating machine translation and post-editing skills into translation curricula to prepare students for the evolving translation technology. While machine translation was generally viewed as a valuable tool, instructors cautioned against over-reliance, warning that it could hinder learning if students use it solely to "spoon-feed" answers without critical engagement.

In conclusion, the studies indicate a general consensus among instructors that machine translation tools, while useful, have limitations in terms of accuracy for longer texts and are better suited for shorter segments or specific translation directions. This demonstrates that instructors are familiar with these tools and acknowledge their usefulness, even though many have not received formal training in their use. Instructors remain divided on the ethicality of machine translation use, largely depending on how the tools are deployed in academic contexts. However, there is broad recognition by instructors that machine translation tools are increasingly essential in translation education, as reflected by the widespread agreement on the need to integrate these technologies into curricula.

2.2. Students' Perceptions of Machine Translation Tools

Students across multiple contexts exhibit a strong inclination towards using machine translation tools in their translation tasks, particularly due to the ease and efficiency with regard to using these tools. For example, over 60% of Chinese students [18] and 69% of Turkish students [19] consistently used machine translation for assignments, with many even recommending its use. Almusharraf and Bailey [4] reported that both Saudi and Korean students valued machine translation for its availability and quick translation processes, particularly for short segments. Additionally, nearly half of the students in Liu et al. [18] preferred to post-edit the outputs rather than use them directly. Similarly, Aslan [19] found that students often used machine translation to verify their own translations. Almusharraf and Bailey [4] also highlighted the tool's role in helping students with vocabulary acquisition and language comprehension, especially for beginner and intermediate learners.

Students generally perceive machine translation quality to be more reliable for shorter texts such as individual words or sentences, but less accurate when it comes to longer or more complex texts [18, 19]. While machine translation is valued for its efficiency and convenience, students recognize its limitations, particularly in handling advanced linguistic structures, metaphors, and religious or literary texts [4, 18]. Despite these accuracy concerns, many students continue to view machine translation as useful in supporting their translation tasks, especially when they use it critically for verification and when they post-edit [4, 18].

Students generally recognize the ethical considerations surrounding machine translation use, with many believing that the degree of post-editing influences whether or not its use is ethical [18, 19]. Some students expressed the view that as long as they actively improve machine translation outputs and apply their own judgment, using machine translation is acceptable [18]. However, there are concerns about over-reliance, with students acknowledging that machine translation should not be a substitute for critical thinking, especially in educational contexts [4].

A common call for better integration of machine translation training into curricula was noted. Liu et al. [18] and Aslan [19] highlighted the lack of formal machine translation training among students, advocating for its inclusion in translation courses. Almusharraf and Bailey [4] echoed this sentiment, suggesting that guided practice would help students use machine translation more effectively without becoming overly reliant on it.

2.3. Previous Empirical Studies on Instructors' Ability to Detect Machine Translation Use

The present study aims to investigate instructors' ability to distinguish between ESL student-generated translations and those produced by GT. Since there has been substantial improvement in GT quality after Google adopted neural machine translation in 2016 [3], we review relevant studies published after 2016. Studies directly comparing instructors' detection of machine translation in students' translations are limited. Most existing research focuses on instructors' detection of machine translation use in students' writings rather than in translations.

To the best of our knowledge, the Master's Thesis by Nygård [20] and the work of Innes [21] are the only studies directly aligned with the present research, as they specifically examined instructors' ability to distinguish between student translations and those produced by a machine. In Nygård's [20] study, 18 Norwegian upper secondary school teachers were tasked with identifying whether Norwegian-English translations were produced by students or by machine translation tools such as GT and ChatGPT. The study found that teachers correctly identified machine translations 53% of the time. Teachers struggled the most with identifying translations made by ChatGPT (46% accuracy). The qualitative data revealed that teachers often relied on features such as organization, grammar, and idiomaticity when making correct judgments, but were misled by word choice and register, leading to incorrect assessments.

Innes [21] provides another direct comparison of translations by students and machine translation systems. In this study, 17 native English teachers were asked to distinguish between human translations and machine translations of Japanese news stories. Teachers were able to correctly identify the machine translations 74.04% of the time. The study highlighted linguistic features such as passive voice and inappropriate pronoun use as markers of machine translation use. The overall conclusion was that increasing sophistication in machine translation could make detection more challenging in the future.

Maimone and Jolley [22] explored a similar area but with a focus on detecting machine translations within student-written texts rather than students' translations. In their study, 26 intermediate-level Spanish learners produced texts, either with or without the aid of GT. Thirty-one college instructors of Spanish were then tasked with distinguishing between machine-assisted and non-machine writing samples produced by the Spanish learners. The results showed that instructors could accurately distinguish machine-assisted texts 73% of the time, with higher accuracy in terms of identifying non-machine writing (82.76%) compared with machine-assisted writing (63%). Notably, the instructors relied on vocabulary and grammatical structures that they deemed beyond the typical level of L2 learners in detecting machine translation use in students' writings.

Similarly, Stapleton and Leung Ka Kin [23] compared instructors' evaluations of machine translations and students' writings in the Chinese-to-English context. The study involved collecting 26 English and 22 Chinese compositions from 6 primary students in Hong Kong. The Chinese compositions were translated into English using GT and then evaluated alongside the original English compositions. The instructors often rated the GT output higher than the students's generated compositions, with only two out of twelve instructors suspecting that machine translation had been used. The preference by teachers with regard to the GT versions might be explained by the fact that the participants are primary-level students who are writing in their L2.

In terms of assessing the ability of educators to differentiate between AI-generated and human-written texts, particularly when the task involves writing on a specific topic rather than translating into a specific language, the literature indicates notable challenges in detecting AI use. For instance, Hostetter et al. [10] demonstrated that both college students and faculty struggled to detect AI-generated writing, with detection rates no better than chance. Similarly, Avila-

Chauvet and Mejía[8] found that both teachers and students were unable to reliably identify the origin of written texts. Waltzer et al.[24] further noted that as the quality of student-written texts improves, it becomes increasingly difficult to distinguish them from AI-generated writings.

To sum up, it seems that the current literature suggests that when students write in or translate into their L2, instructors are generally more successful at detecting machine translation use. This is likely due to the fact that producing language in the L2 tends to result in non-native-like errors, which are more easily distinguishable from machine translations, as machine translation models do not typically make such mistakes. Features such as vocabulary and grammatical structures that are beyond the typical level of L2 learners also make it easier for instructors to detect machine translation use[21, 22]. In contrast, the present study explores instructors' ability to detect machine use in L2 to L1 translation, where participants produce writings in their native language, providing a different challenge for instructors. In addition, most of the reviewed studies compare machine translations to students' writings, not to their translations. This is typically done by having one group of students write in their L1 and another group write on the same topic in their L2. Researchers then translate the L1 version into L2 using machine translation and compare it to the L2 version produced by the students. Conversely, the current study asks participants to translate an L2 text into their L1 and then uses machine translation to translate the same L2 text into L1. Instructors are then tasked with detecting machine use in the L1 translations.

3. Methodology

3.1. Participants

The study involved 22 university professors. They were asked to evaluate translations and determine whether each one was student-generated or machine-generated. This group comprised 12 females and 10 males, with 15 holding PhDs and 7 holding master's degrees, all in language-related fields. Each instructor had at least six years of teaching experience in ESL education. However, 69% of the instructors indicated that they had never received any prior training in machine translation.

Additionally, the translations assessed by the instructors were produced by 20 university students enrolled in a final-year translation project course. These participants, 16 males and 4 females, were majoring in English and Translation at a Saudi university. Their ages ranged from 22 to 25 years. The students were selected through convenient sampling, as they were the available students registered for the translation project class in that semester. Students enrolled in this advanced course had already completed essential introductory courses in translation and related language subjects. This background ensured they had a solid understanding of both the source and target languages.

3.2. Materials

The materials for this study consisted of 22 human-translated sentences from 20 different students. These translations originated from in-class assignments where students were required to translate entire paragraphs from English to Arabic. These paragraphs varied in their features.

For instance, one of the texts, titled "Chinese Immigrants during the Industrial Era," begins with: "During the industrial era, immigrants from various parts of Asia and Eastern and Southern Europe came to the U.S. in even greater numbers than those from Western Europe." This text is historical and informational, characterized by descriptive language and a chronological structure.

Another text narrates the story of Scott Douglas, an individual who struggled with addiction. It begins with: "Two years ago, Scott Douglas died of a heroin overdose. Why, then, is his father, who is a persistent opponent of drugs, calling for all drugs to be legalised?" This text adopts a conversational and argumentative style, incorporating persuasive language and rhetorical questions.

In addition to the translations produced by the students, the same sentences were translated using GT, creating a dataset of 44 translations: 22 human-generated and 22 machine-generated. Therefore, each instructor-participant made 44 translation judgments, resulting in a total of 968 possible responses.

3.3. Procedure

To facilitate translation evaluations, the materials were divided into three separate Google Forms. This approach helped minimize fatigue among the instructor participants. The forms were distributed electronically, allowing the instructors to complete them at their own pace. After submitting one form, an instructor would receive the next, continuing until all three forms were completed. To prevent order effects from influencing the instructors' judgments, the sequence of the forms and the order of translations within each form—human versus machine—were randomized. After the instructors read the translations, they were presented with this question:

Which of the following is true:

a) The first translation seems like it's machine-generated;

b) The second translation seems like it's machine-generated;

c) Not sure.

Upon completion of the evaluation forms and to assess the instructors' self-perceived confidence in distinguishing between machine-generated and student-generated translations, they were asked a follow-up question. The question read:

In general, when reading translated sentences from English to Arabic, how confident are you that you can usually tell whether it was translated by a machine or by a student?

a) I think I can usually tell the difference;

b) I think I usually cannot tell the difference;

c) I am unsure about my ability to tell the difference.

4. Results

Participants had the option to respond to or skip any trial. Therefore, out of 968 possible trials, 434 responses were recorded. Of these, 32% were correct identifications of machine translation, 44% were incorrect identifications, and 25% indicated being unsure (see **Figure 1**). The responses were then coded as 0 (unsure/incorrect) and 1 (correct) to test whether the proportion of correct identifications was statistically significant. The analysis was conducted using a Generalized Linear Mixed Model (GLMM) in R, employing the glmer function from the lme4 package with a binomial family. This statistical method was chosen to account for the random effects associated with the repeated measures from individual instructors.

As **Table 1** shows, the estimated intercept was -0.948, with a standard error of 0.229. The negative value of the intercept suggests that, on average, instructors were significantly less likely than not to correctly identify machine translations ($z = -4.127$, $p < 0.001$). Converting the log-odds of −0.948 to a probability using the *exp* function indicates that instructors, on average, have a 28% chance of correctly identifying machine translations, which is significantly below the 50% expected as a result of random guessing.

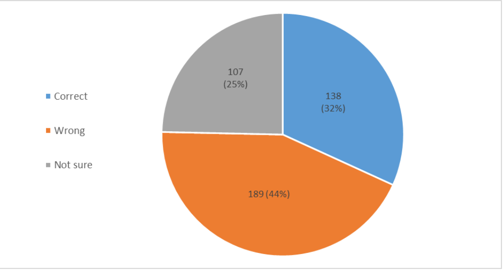

Figure 1. Number and percentage of each identification type.

On the question of what instructors think of their ability to detect machine translations compared with human translations, 20 instructors responded to the question (see **Figure 2**). Of these, 18 (90%) believed that they can usually tell the difference, and only 2 (10%) were unsure about their ability to tell the difference.

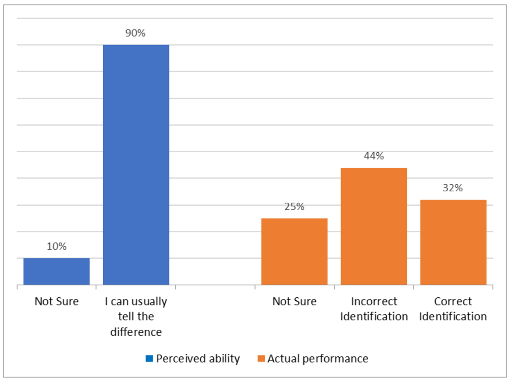

Figure 2. Percentage of Perceived Ability vs. Actual Performance.

5. Discussion

The goal of this research was to evaluate the ability of ESL instructors to accurately distinguish between AI-based machine-generated and student-generated translations. Additionally, it aimed to determine how confident instructors were in their ability to make such distinctions. This inquiry was guided by two research questions: 1) Can ESL instructors accurately detect whether a translation has been generated by a student or by a machine? and 2) How confident are instructors in their ability to differentiate between student-generated

Table 1. Model Estimates.

Term	Estimate	Standard Error	z-Value	p
Machine	−0.948	0.229	−4.127	< 0.001*

*p < 0.001 indicates statistical significance, meaning the accuracy rate of identifying machine translations was significantly lower than would be expected by random chance (50%).

and machine-generated translations?

The findings indicated that instructors generally struggled when it came to accurately identifying machine-generated translations, with a success rate significantly below the 50% expected by random chance. Specifically, instructors were able to correctly identify machine translations in only 32% of cases, with a 28% probability of accurate identification. This shows a substantial difficulty in distinguishing between the two types of translation. Despite their low success rate, the majority of instructors (90%) reported a high level of confidence in their ability to discern the difference between machine-generated and human-generated translations.

Instructor participants in previous studies demonstrated higher accuracy rates in detecting machine translations. For example, Innes [21] reported an identification rate of 74%, and Nygård [20] reported a slightly lower but still relatively high rate of 53%. Although these findings may initially appear to contrast with those of the current study, they actually reinforce a broader understanding of machine translation.

In the current study, students translated from L2 to L1, meaning that the final output was in their native language. In contrast, both Innes [21] and Nygård [20] had participants translate into their L2. Since machine translation systems often produce outputs with human-like fluency [1], it is easier to detect machine use when comparing it to students' L2 writing than when comparing it to their L1 writing. Machine-generated translations are less prone to the types of non-native errors that instructors typically rely on for detection in students' L2 writings. In fact, when students wrote in their L2, instructors in Stapleton and Leung Ka Kin [23] rated machine-generated texts higher than student-generated ones. Therefore, it is expected that the success rate in identifying machine translation tools will be lower when the task involves students writing in their L1.

Another explanation lies in the length of the translation segments. In addition to having participants translate from L1 to L2, in Innes [21] and Nygård [20], the segments were paragraph-length, whereas in the current study, students translated shorter segments of only two to three sentences. Previous research suggests that machine translation errors become more apparent with longer contexts, as longer segments provide more opportunities for errors in terms of cohesion and coherence to appear. For instance, Hassan et al. [2] found that machine translation could achieve near-human quality at the sentence level. However, Läubli et al. [25] attempted to replicate Hassan et al.'s findings with larger segments and discovered that human translations were rated more favorably than machine translations. These results were attributed to machine translation difficulties in maintaining textual cohesion and coherence across longer segments [25]. Consequently, instructors may find it easier to distinguish L2-L1 machine translations from human ones when dealing with longer passages.

The current study revealed that most instructors were unable to accurately differentiate between machine-generated and student translations despite their confidence in doing so. This aligns with the results of Waltzer et al. [24], who found that instructors' confidence in detecting AI-generated texts was such a poor indicator of actual performance that it was essentially useless. This might highlight a gap in instructors' awareness of the advancements in machine translation quality, suggesting that machine translation is improving at a rate that surpasses teachers' expectations. Indeed, previous studies have suggested that machine translation systems are nearing human-professional parity [2, 26], so if machine translation is arguably becoming indistinguishable from professional translators, it will easily catch up to the students' level.

The discrepancy between perception and actual performance raises questions about how such tools should be integrated into translation education. Reliance on instructors' judgments to detect machine translation use might not be a reliable means of addressing these ethical concerns, further supporting the argument advocated by Liu et al. [18] that machine translation and post-editing skills should be formally integrated into translation curricula. Instructors have traditionally relied on various observable indicators to distinguish machine translation outputs from L2 learner writing. Early studies, such as that of Anderson [27], identified frequent machine translation errors, including homograph confusion, mistranslated prepositions, untranslated words, and improper verb tenses. Similarly, Luton [28] noted that machine translation systems struggled with idiomatic expressions, translating proper nouns incorrectly, and leaving some words untranslated. Steding [29] categorized machine translation errors into four main areas: spelling, vocabulary, grammar, and style, highlighting mistakes such as untranslated words and unexpected lexical choices. Correa [30] synthesized a list of common machine translation errors from earlier research, including literal translations, grammatical inaccuracies, awkward phrasing, and the inability to handle idioms or cultural references. More recently, Innes [21] found that raters could identify machine translation use by noticing issues such as improper use of passive voice and prepositional errors.

However, as machine translation systems continue to advance, these once-reliable indicators are becoming outdated. In fact, as Maimone and Jolley [22] and Ducar and Schocket [6] suggested, current machine translation can sometimes be identified not by its flaws, but by how too advanced the writing appears to be compared to what is typically expected from student writing. The level of sophistication in machine translation outputs can exceed learners' assumed capabilities, raising suspicion when the text seems beyond the student's proficiency level. Ducar and Schocket [6] identified several indicators of machine translation use, such as excessive use of sophisticated vocabulary, correct usage of complex structures, the absence of common learner errors such as prepositional mistakes, and producing verb tenses that have not yet been studied. Along these lines, Yamada [11] further suggested that it is now harder to recognize the subtle errors in the fluent-sounding outputs of machine translation systems. This shift challenges traditional detection methods and highlights the growing complexity of distinguishing human from machine translations.

The low success rate of 32% in identifying machine translations raises concerns about the reliability of take-home translation assignments, where students may use machine translation without detection. This finding necessitates a rethinking of traditional assessment methods with regard to translation, with the need for a greater emphasis on in-class assessments that limit students' access to machine translation tools. Furthermore, since research suggests that students are likely to use machine translation even when explicitly instructed not to do so [6], take-home assignments should be designed to integrate machine translation effectively, such as by requiring students to post-edit such translations. Despite this, very few instructors (7.69%) assign tasks requiring students to use machine translation tools [17], underscoring the need to promote active engagement with these technologies. Such integration allows students to learn to critically engage with and refine machine-generated texts rather than blindly relying on them.

6. Conclusion and Pedagogical Recommendations

This study sought to evaluate ESL instructors' ability to accurately distinguish between machine-generated and student-generated translations and to gauge their confidence in doing so. The findings suggest that instructors struggled to accurately identify machine translations, with a success rate significantly lower than expected by chance. Despite this, instructors expressed high confidence in their ability to discern between the two, highlighting a disconnect between their perceived and actual performance. This discrepancy points to the growing sophistication of machine translation systems, which are increasingly achieving human-like fluency, making it difficult to detect machine translation use based on traditional indicators.

The implications of these findings are important for translation education. Relying solely on instructors' judgment to detect machine translation use may no longer be an effective strategy. Additionally, instructors should not attempt to ban machine translation tools entirely, as students are increasingly relying on them and are likely to use them regardless of restrictions [6]. Instead, there is a need to formally integrate machine translation and post-editing skills into the curriculum, ensuring that both students and instructors are equipped to navigate this evolving landscape. Specifically, instructors can implement dedicated machine translation sessions that teach students how to assess and refine machine translations. Instructors should display samples of machine-generated translations in class, guiding students in refining the output collaboratively. This approach allows instructors to highlight shortcomings of machine translation, such as literal translations, incorrect collocations, and a lack of contextual awareness. Additionally, instructors can integrate comparative analysis exercises, where students compare human and machine translations, discuss errors, and reflect on the role of post-editing in professional translation settings.

From an assessment perspective, traditional take-home translation assignments, where students may have unrestricted access to machine translation tools, may no longer serve as effective assessment methods. To address this, educators should consider incorporating in-class assessments, where students translate short passages without machine assistance, allowing instructors to assess their raw translation abilities. Also, take-home tasks should focus on post-editing machine translations which foster critical engagement with the tools rather than passive reliance on them. As machine translation continues to advance, adapting teaching practices and assessment strategies will be crucial for maintaining academic integrity and ensuring that students develop the skills necessary to navigate both human and machine-generated translations.

7. Limitations and Future Research

One limitation of the current study is the use of convenience sampling, as the participants were drawn from a pool of instructors and students who work and study at the same university as the researcher, potentially leading to a sample with shared characteristics. Future research should consider recruiting a broader and more diverse sample to achieve more generalizable findings.

Additionally, previous studies, such as those of Innes[21] and Nygård[20], examined L1 to L2 translation and reported high success rates in teachers' identification of machine-generated translations. In contrast, the current study found lower identification rates in the L2 to L1 direction, suggesting that translation into one's native language presents distinct challenges for detection. More research is needed to explore this L2 to L1 direction further and to verify whether the findings of this study are consistent across different language pairs and contexts.

Another limitation relates to the length of the translation segments. Participants in this study translated relatively short texts (two to three sentences). Previous research suggests that machine translation errors become more evident in longer segments, where issues of cohesion and coherence arise. Future studies should investigate how instructors perform with longer passages, particularly when students are translating into their dominant language (L2 to L1). Furthermore, the study did not analyze the specific linguistic markers that may have misled instructors when classifying translations. Exploring why instructors misidentify translations—such as particular lexical, syntactic, or stylistic features that contribute to misclassification—could strengthen future analyses.

Moreover, instructors' familiarity with machine translation tools was not deeply explored in this study, leaving it unclear whether prior experience enhances detection accuracy. Future research should examine whether instructors with greater exposure to MT systems perform better in distinguishing human from machine translations.

Finally, since the majority of instructors in the current study (69%) had not received formal machine translation-related training, it could be beneficial to explore whether more experienced and trained instructors would perform differently. Factors such as prior training or experience with machine translation are fertile ground for investigating the ability to detect machine translation.

Funding

This work was funded and supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [KFU250727].

Institutional Review Board Statement

Ethical review and approval were waived for this study. This study did not involve direct interaction with human participants or the collection of personal or identifiable data. It was based solely on pre-existing student translations, analyzed without any intervention or influence on the students. Professors' involvement was limited to making anonymous online judgments on these translations without providing personal or identifiable information. No interventions, experiments, or sensitive data collection were conducted.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data that supports the findings of this study is available from the author upon request.

Acknowledgments

Sincere appreciation is given to the reviewers for their constructive feedback on the initial draft of this paper.

Conflicts of Interest

The author declares no competing interests.

References

[1] Wang, H.F., Wu, H., He, Z.J., et al., 2022. Progress in Machine Translation. Engineering. 18(2), 143–153. DOI: https://doi.org/10.1016/j.eng.2021.03.023

[2] Hassan, H., Aue, A., Chen, C., et al., 2018. Achieving Human Parity on Automatic Chinese to English News Translation. DOI: https://doi.org/10.48550/arXiv.1803.05567

[3] Zhang, J.J., Zong, C.Q., 2020. Neural machine translation: Challenges, progress and future. Sci China Technol Sci. 63, 2028–2050. DOI: https://doi.org/10.1007/s11431-020-1632-x

[4] Almusharraf, A., Bailey, D., 2023. Machine translation in language acquisition: A study on EFL students' perceptions and practices in Saudi Arabia and South Korea. J Comput Assist Learn. 39(6), 1988–2003. DOI: https://doi.org/10.1111/jcal.12857

[5] Bin Dahmash, N., 2020. 'I Can't Live Without Google Translate': A Close Look at the Use of Google Translate App by Second Language Learners in Saudi Arabia. Arab World English Journal. 11(3), 226–240. DOI: https://dx.doi.org/10.24093/awej/vol11no3.14

[6] Ducar, C., Schocket, D.H., 2018. Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. Foreign Lang Ann. 51(4), 779–795. DOI: https://doi.org/10.1111/flan.12366

[7] Kazemzadeh, A., Kashani, A.F., 2014. The effect of computer-assisted translation on L2 learners' mastery of writing. Int J Res Stud Lang Learn. 3(3), 29–44. DOI: https://doi.org/10.5861/ijrsll.2013.396

[8] Avila-Chauvet, L., Mejía D., 2023. Can professors and students detect ChatGPT essays? Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4373643 (cited 29 September 2024).

[9] Fleckenstein, J., Meyer, J., Jansen, T., et al., 2024. Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. Comput Educ Artif Intell. 6(5), 100209. DOI: https://doi.org/10.1016/j.caeai.2024.100209

[10] Hostetter, A.B., Call, N., Frazier, G., et al., 2023. Student and Faculty Perceptions of Artificial Intelligence in Student Writing. DOI: https://doi.org/10.31234/osf.io/7dnk9

[11] Yamada, M., 2019. The impact of Google neural machine translation on post-editing by student translators. The Journal of Specialised Translation. 31(1), 87–106. Available from: https://www.researchgate.net/profile/Masaru-Yamada/publication/364994185_art_yamada_newpdf/data/63620fc037878b3e87755cf7/art-yamada-new.pdf (cited 19 August 2024).

[12] Elkhatat, A.M., Elsaid, K., Almeer, S., 2023. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. Int J Educ Integr. 19(1), 1–16. DOI: https://doi.org/10.1007/s40979-023-00140-5

[13] Pan, W.H., Chok, M.J., Wong, J.L.S., et al., 2024. Assessing AI Detectors in Identifying AI-Generated Code: Implications for Education. Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET '24); April 14–April 20, 2024; Lisbon, Portugal. pp. 1–11.

[14] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., et al., 2023. Testing of detection tools for AI-generated text. Int J Educ Integr. 19(1), 26. DOI: https://doi.org/10.1007/s40979-023-00146-z

[15] 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. DOI: https://doi.org/10.1017/S0958344020000191

[16] Ata, M., Debreli, E., 2021. Machine Translation in the Language Classroom: Turkish EFL Learners' and Instructors' Perceptions and Use. IAFOR J Educ. 9(4), 103–122. DOI: https://doi.org/10.22492/ije.9.4.06

[17] Jolley, J.R., Maimone, L., 2015. Free online machine translation: Use and perceptions by Spanish students and instructors. In: Moeller, A.J. (eds.). Learn languages, explore cultures, transform lives. Publisher: Johnson Litho Graphics of Eau Claire, Ltd., Eau Claire, WI, USA. pp. 181–200.

[18] Liu, K.L., Kwok, H.L., Liu, J.W., et al., 2022. Sustainability and influence of machine translation: perceptions and attitudes of translation instructors and learners in Hong Kong. Sustainability. 14(11), 6399.

DOI: https://doi.org/10.3390/su14116399

[19] Aslan, E., 2023. Machine Translation: Perception of Translation and Interpreting Students in Turkey. Curr TRENDS Transl Teach Learn E. 10(1), 185–216. DOI: https://doi.org/10.51287/cttl20237

[20] Nygård, A., 2024. Testing teachers' abilities to distinguish between translations produced by students, Google Translate and ChatGPT [Master Thesis]. Trondheim: Norwegian University of Science and Technology. pp. 1–53.

[21] Innes, A., 2019. Differentiating between machine translation and student translation: red flags and salient lexicogrammatical features. Lub Stud Mod Lang Lit. 43(4), 1–13. DOI: https://doi.org/10.17951/lsmll.2019.43.4.1-13

[22] Maimone, L., Jolley, J., 2023. Looks like google to me: Instructor ability to detect machine translation in L2 Spanish writing. Foreign Lang Ann. 56(3), 627–644. DOI: https://doi.org/10.1111/flan.12690

[23] Stapleton, P., Kin, B.L.K., 2019. Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. Engl Specif Purp. 56, 18–34. DOI: https://doi.org/10.1016/j.esp.2019.07.001

[24] Waltzer, T., Cox, R.L., Heyman, G.D., 2023. Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students. Hum Behav Emerg Technol. 2023(1), 1–9. DOI: https://doi.org/10.1155/2023/1923981

[25] Läubli, S., Sennrich, R., Volk, M., 2018. Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing; October 31–November 4, 2018; Brussels, Belgium. pp. 4791–4796.

[26] Graham, Y., Federmann, C., Eskevich, M., et al., 2020. Assessing Human-Parity in Machine Translation on the Segment Level. In: Cohn, T., He, Y.L., Liu, Y. (eds.). Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics: Online. pp. 4199–4207.

[27] Anderson, D.D., 1995. Machine translation as a tool in second language learning. CALICO J. 13(1), 68–97. DOI: https://doi.org/10.1558/cj.v13i1.68-97

[28] Luton, L., 2003. If the Computer Did My Homework, How Come I Didn't Get an "A"? Fr Rev. 76(4), 766–770.

[29] Steding, S., 2009. Machine Translation in the German Classroom: Detection, Reaction, Prevention. Die Unterrichtspraxis/Teaching Ger. 42(2), 178–189.

[30] Correa, M., 2014. Leaving the "peer" out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. Lat Am J Content Lang Integr Learn. 7(1), 1–20. DOI: https://doi.org/10.5294/laclil.2014.7.1.1