



Efficacy of Instatext for Improving Persian-English Freelance Translators' Language Quality: From Perception to Practice

Ali Beikian

*English Department, Chabahar Maritime University, Chabahar,
Sistan and Baluchestan, Iran
a_beikian@cmu.ac.ir*

Mansoor Ganji*

*English Department, Chabahar Maritime University, Chabahar,
Sistan and Baluchestan, Iran
ganjimansoor1980@gmail.com*

Abstract

There is growing agreement among researchers on the advantages of using automated feedback programs (AFPs), but most of the previous studies have evaluated more well-known AFPs like Grammarly, Ginger, etc. in English writing classes. None of the previous studies on AFPs evaluated the effectiveness of and users' perception of InstaText. Thus, this study was aimed at examining the effects of InstaText on improving the language quality (i.e., grammar, spelling, and style) of Persian-English freelance translators using InstaText for editing their English translations of Persian academic papers, which are considered technical translations. In addition, it was conducted to investigate how the said users perceived this InstaText. This quantitative study was conducted in two phases: a one-group pretest-posttest phase, where the effect of using InstaText on improving the language quality of translated technical texts was examined with 75 participants; and a survey phase, where the participants' perception toward InstaText was measured using Usefulness, Satisfaction, and Ease of Use (USE) questionnaire. InstaText did not help the participants make significant progress in grammar and spelling, but its effect on improving their style was significant. Further, the participants perceived the tool as intuitive, user-friendly, efficient, time-saving, and satisfactory.

Keywords: Automated Feedback Program (AFP); Automated Writing Evaluation (AWE); InstaText; Language Quality; Perception; Persian-English Freelance Translators

Received: 2022-11-03

Accepted: 2022-12-17

Available Online: 2022-12-17

DOI: 10.22034/efl.2022.368332.1205

* Corresponding Author

1. Introduction

1.1 Automated Writing Evaluation Tools

As an important, yet difficult-to-acquire productive skill, writing can act as a means for learning other receptive and productive skills (Hyland & Hyland, 2006) and is considered an essential skill for students at the tertiary level (Narita, 2012). It makes thought available for reflection, encourages students to communicate, and, thus, brings about thinking and learning (Mekheimer, 2005). The sovereign position of writing is further elaborated on by Olshtain (2001), who argues "...the skill of writing enjoys special status—it is via writing that a person can communicate a variety of messages to close or distant known or unknown readers" (p. 207).

Despite the importance of writing skill, Learners of English as a Foreign Language (EFL), international students studying in English-speaking countries, and their teachers experience multifarious unfavorable emotions (Tynan & Johns, 2015), and writing is among the most demanding competences they need to acquire (Afshari & Salehi, 2017). Besides, even native speakers of English exhibit difficulties with grammar and form in their writing in academic settings (Koroglu, 2014). Acknowledging the burdensome nature of writing, Nunan (1999, as cited in Afshari & Salehi, 2017) argues that "...producing a coherent, fluent, extended piece of writing is probably the most difficult thing there is to do in language" (p. 2).

Other serious challenges for writing teachers are to involve students in the writing process actively and to help them master grammar, punctuation, capitalization, and spelling, which are essential elements in delivering the message clearly and precisely (Perdana & Farida, 2019). Therefore, providing enough instruction and feedback on the said elements becomes highly important, especially in the initial stages of writing (Williams, 2004), when direct feedback is needed for improving accuracy (Dikli & Bleyle, 2014). This requires intervention and explicit input provided by teachers (Muller et al., 2017). However, teachers are unwilling to provide sufficient language support for all students (Murray, 2010), teachers' feedback is not consistent, and the students do not understand them (Ranalli et al., 2017), and teachers might not find the time and have the patience to correct all the essays written by students and provide them with accurate and holistic feedback (Wilson & Czik, 2016).

In order to help EFL students improve their abilities in writing, EFL teachers and researchers have adopted different initiatives and taken different steps such as offering corrective feedback, providing them with model essays, and employing computer software and online resources, technically called computer-assisted language learning (CALL) (Daniels & Leslie, 2013). Integrative CALL



started in the last decade of the 20th century, integrating technology more fully into language teaching and learning process through using authentic language in meaningful contexts (Rinaldo et al., 2011).

Automated writing evaluation (AWE) has been designed to help language learners with both input and correction on grammar, spelling, mechanics of language, and style among others. AWE tools have become increasingly ubiquitous and powerful (Daniels & Leslie, 2013). In its list of online AWE tools, *Just Publishing Advice* (2020) mentions 50 free writing software applications under seven categories of which one main category is that of automated feedback programs (AFPs). AFPs, including *Grammarly*, *Ginger*, *WhiteSmoke*, *ProWritingAid*, *GrammarCheck*, *Hemingway Editor*, *Slick Write*, *InstaText*, and *Language Tool*, mainly aim to help writers check, correct, and improve their writing components such as grammar, spelling, and mechanics of language.

1.2 Statement of the Problem

Previously, it was believed that success in technical translation projects depended only on two skills: command of the source language, which would allow the translator to comprehend the source material accurately, and proficiency in the target language, which would enable them to produce a text that clearly expressed the information contained in the source material. The importance of knowing the target language increased when the purpose of translation turned more toward communication. Herman (1993, as cited in Duběda & Obdržálková, 2021) noted that technical translation has the same stylistic objectives as technical writing: correctness, clarity, and concision. Byrne (2010) argues that technical translators would be able to produce translations that are up to par as long as they have the language abilities necessary to pass for the writing of a true expert in the subject matter. To put it another way, they must be very proficient writers in the target language, be able to write creatively in a range of text types, and have a knack for words in order to provide high-quality technical translations. Herman's maxim that one must first be a good technical writer in order to be a good technical translator has taken on new significance and dimensions as a result of various environmental changes that have an impact on the very nature of technical documentation and how it is produced and translated. The role of the technical translator has changed as a result of new research, increased collaboration at the professional and academic levels, new technologies and media for dissemination, as well as various legal considerations (see, for instance, Byrne, 2007; 2010). Technical translators these days not only translate but also write, create, and, more often, design and engineer texts (Byrne, 2010).

A translator translating into their mother tongue enjoys the perfect mastery of the target language with all its subsystems and rules and finds it simple to produce a quality target text because they draw on a natural linguistic repertoire that they

have amassed from birth and that flows easily and naturally when needed. A translator translating into a foreign language, on the other hand, may not have complete command of the grammar and vocabulary of the target language, which inevitably leads to flouting target language norms and creating odd grammatical structures. [Campbell \(1998, p. 57\)](#) maintains that translators translating into the mother tongue avoid "the problem of lack of textual competence in the target language- in other words, native writers can manipulate all the devices that go to make up natural-looking texts." Similarly, [Newmark \(1995, p. 180\)](#) emphasizes the importance of natural writing in translation, which can only be generated by the native speaker by arguing that a non-native translator "...will be caught every time, not by his grammar, which is probably suspiciously 'better' than an educated native's, not by his vocabulary which may well be wider, but by his unacceptable collocation". [Baker \(2002\)](#) also considers the perfect mastery of the target language a decisive factor by asserting that translators should only translate into their language of habitual use.

No doubt, native English speakers are the best resources for translating technical Persian texts into English. However, this is not the case in the Iranian context, where a majority of technical texts, including academic papers, are translated from Persian to English by Iranian freelance translators whose native language is Persian. Therefore, AFPs could be considered a good resource to help them achieve the stylistic objectives of technical writing mentioned by Herman (1993, as cited in [Duběda & Obdržálková, 2021](#)), i.e., correctness, clarity, and concision.

1.3 Purpose of the Study

There is growing agreement among researchers on the benefits of using AFPs in EFL writing classes ([Afshari & Salehi, 2017](#); [Rahimi et al., 2020](#)). As will be shown in the next section, most of the previous studies have evaluated more well-known software applications such as *Grammarly*, *Ginger*, *iWrite*, *MI Write*, *PEG Writing*, *NC Write*, *Writing Roadmap*, *EssayCritic*, and *Summary Street* in English writing classes. The present study, however, is intended to investigate the effectiveness of *InstaText* in improving the language quality of texts translated from Persian into English by freelance translators whose mother tongue is not English. In other words, the aim of the current study is twofold and as follows: 1. to see how the Persian-English freelance translators who used *InstaText* for editing their English translations of Persian academic papers perceive this online tool, and 2. to investigate the effects of the said tool on the language quality of the said translators' final translations through comparing the language quality of their translations before and after using *InstaText*.



2. Review of the Related Literature

The development of writing expertise entails sustained deliberate practice (Kellogg & Whiteford, 2009), which will help students both develop fluency and automaticity in lower-level writing skills (Kellogg, 2008) and have complete control of the central cognitive processes which are at work in composing (Graham et al., 2019). In addition to sustained deliberate practice, effective and frequent feedback also gets the lion's share in the development of writing expertise (Hattie & Timperley, 2007), and such feedback can be beneficial to students, especially when it is provided immediately and in a localized, specific, and detailed manner by addressing both surface-level and content features of writing (Hattie & Timperley, 2007; Patchan et al., 2016). The problem, however, is that on the one hand students can hardly engage in sustained deliberate practice due to the limited time allocated to writing courses included in most curricula (Brindle et al., 2016). On the contrary, providing high-quality feedback in the manner described above is demanding both in terms of teachers' knowledge (Parr & Timperley, 2010) and their time (Dikli, 2010). Previous research (e.g., Matsumura et al., 2002) concluded teachers' feedback often focuses on low-level writing skills with little effect on students' writing performance. Hence, improving students' writing outcomes necessitates that alternative methods capable of helping enhance both students' sustained deliberate practice and teachers' feedback frequency and quality be identified.

Automated writing evaluation (AWE), defined as using all the computer technologies to evaluate and score a written text, is one of such alternative methods. AWE tools were developed with three aims in mind: improving the practice-feedback circle (Kellogg et al., 2010); reducing teachers' evaluation demands (Palermo & Wilson, 2020); and creating fair scorings without any human imperfection (Stevenson & Phakiti, 2019; Wang et al., 2020). A central feature of any typical AWE tool is the provision of automated feedback aimed to help learners improve their writing quality by increasing students' access to the cycles of practice and feedback (Palermo & Wilson, 2020). AWE tools use a scoring engine named automated essay scoring (AES) (Shermis & Burstein, 2003). AES, in turn, comprises a set of computerized methods for evaluating the text and providing a general numeric score; and it can be based on either latent semantic analysis (LSA) or natural language processing (NLP). LSA-based AES evaluates content, while NLP-based AES evaluates a broad array of aspects such as mechanics, usage, grammar, sentence variety, and the like (Nunes et al., 2021). In addition to the AES engine, AWE tools include a qualitative feedback engine providing feedback aimed at improving the quality of writing (Allen et al., 2016).

Previous studies have shown that using AWE tools has some benefits for supporting writing skills acquisition. As regards students, it has been found that

AWE tools help reduce the time spent on the writing task, increase the number of revisions students made, and improve their writing content (Franzke et al., 2005). In addition, it has been revealed through several studies that feedback provided by AWE tools has been beneficial to students in that they help them better both their writing attitudes and their opinions about their writing potential (Roscoe et al., 2018; Camacho et al., 2020), increase their autonomy and self-efficacy (Wilson & Czik, 2016; Wilson & Roscoe, 2020), enhance their writing quality, especially the mechanical aspects such as grammar and spelling, and accuracy due to the corrective feedback provided (Franzke et al., 2005; Kellogg et al., 2010; Wilson & Czik, 2016), improve their performance on state ELA exams (Wilson & Roscoe, 2020), and receive more accurate and objective scores (Warschauer & Grimes, 2008).

With respect to user perception, students using AWE tools tend to perceive them as valuable, since they incite the users to reflect on and become more aware of the writing process (Franzke et al., 2005). Furthermore, the feedback provided by AWE tools helps the users feel more confident and motivated, enjoy writing, rewriting, and revising more, and remain focused for longer (Tang & Rich, 2017; Palermo & Thomson, 2018).

On the negative side of the coin, some studies have pointed out that AWE tools are afflicted with several general disadvantages and limitations as follows: unclear feedback and judgments resulting from limited quantitative information on such issues as word repetition, word distribution, sentence length, etc. (Chou et al., 2016); limitedness to evaluating the content based on the specific prompts of the program and, thus, discriminating against students unfamiliar with technology (Khoii & Doroudian, 2013); and sufficing to common recommendations, which is a weakness in formative learning, profound meaning negotiation, and rich content development (Chen & Cheng, 2008). In addition, some studies show that students found some AWE tools' feedback too extensive and overwhelming (Ranalli, 2018) and that they cannot interpret the automated feedback without additional attention and support from the teachers (Palermo & Thomson, 2018; Wilson & Roscoe, 2020).

Although the findings of the previous studies show that the outcomes of using AWE tools are generally positive, such studies are mostly focused on investigating the effects of and/or users' perceptions toward AWE tools such as *Grammarly* (Koltovskaia, 2020), *Ginger* (Lastari, 2021), *Criterion* (Casal, 2016); *iWrite* (Qian et al., 2020), *MI Write* (Palermo & Wilson, 2020; Wilson et al., 2022) *PEG Writing* (Wilson & Czik, 2016; Wilson & Roscoe, 2020), *NC Write* (Palermo & Thomson, 2018), *Writing Roadmap* (Tang & Rich, 2017), and *Summary Street* (Franzke et al., 2005). In light of the foregoing, the present study can be novel in terms of its choice of the AWE, namely *InstaText*, which has not been studied to date to the best of our knowledge, and its participants' use of the



software, i.e. for improving the quality of texts translated into English rather than originally written in English.

3. Method

3.1 Design

The present quantitative study was conducted in two phases. In the first phase, being of a one-group pretest-posttest design, the effect of using *InstaText* on improving the language quality of texts translated from Persian into English was investigated in a single group of participants. In the second phase, which was of a survey design and immediately followed the first, a perception questionnaire was administered to the participants to collect their impressions about using *InstaText* as an aid for improving the language quality of texts translated from Persian into English.

3.2 Participants

The participants of the study included 75 Persian-to-English translators who took a freelance translator admission exam following a call for freelance translators of academic papers by a reputable translation agency in Tehran, Iran. The participants' demographic information is presented in Table 1. The data for the first phase of the study were collected from the responses of all 75 participants to the questionnaire, but the data for the second phase were collected through the comparison of the translations submitted by 15 participants before and after using *InstaText*.

Table 1
The Participants' Demographic Information

Gender	Age Range	Degree	Field of Study	Translation Experience
Female: 58	20-37 Years	B.A./B.S. Student: 3	Translation: 28	0-3 Years: 45
Male: 17		B.A./B.S.: 28	English Literature: 9	4-6 Years: 20
Total: 75		M.A./M.S. Student: 1	TEFL: 8	7-9 Years: 4
		M.A./M.S.: 37	Linguistics: 3	10-12 Years: 5
		Ph.D. Candidate: 2	Other: 27	Over 12 Years: 1
		Ph.D.: 3		
		M.D.: 1		

3.3 Instrumentation

3.3.1 Persian-to-English Translation Pretest

Given the aims of the exam, the translation test was designed in such a way that it would cover the three main areas of Arts and Humanities, Engineering Sciences, and Life Sciences and represent the three challenging sections of academic papers, i.e., Abstract, Introduction, and Discussion. The reason for the latter decision was that as per our experience at the translation agency, our existing freelance staff's translations were afflicted with the largest number of issues in the said three sections. Considering the foregoing concerns, we designed a Persian-English translation test adding up to 502 words in total and comprising three paragraphs as follows: 1. the abstract of a paper in the area of life sciences (166 words); 2. the first paragraph of the introduction of a paper in the area of engineering sciences (138 words); and 3. the first paragraph of the discussion section of a paper in the area of arts and humanities (198 words).

3.3.2 Persian-to-English Translation Posttest

For the posttest, which followed the treatment immediately, the participants were required to use InstaText for editing their translations produced in the pretest phase. Adopting such a strategy, the researchers tried to minimize the typical threats to the internal validity of one-group pretest-posttest designs such as history, testing, maturation, instrumentation, regression to mean, and spontaneous remission (Campbell & Stanley, 2015).

3.3.3 InstaText

InstaText is an artificial intelligence-based online tool developed to help both native and non-native speakers of the English language with writing, editing, and revising academic papers, business proposals, marketing materials, and translations among others. Using artificial intelligence and language technologies, it generates recommendations and ideas on how to improve the text. For the purposes of the present study, we had the participants use the free trial version of *InstaText*.

The motivation behind choosing *InstaText* was that it claims to be a "personal writing assistant" (InstaText, 2021, Para. 3) aimed at simulating a native speaker (Para. 7) and capable of helping the user to produce high quality and efficient sentences and, thus, write clearly and accurately (Para. 5). It further claims that "one of the benefits that *InstaText* users so often praise is generating ideas" (Para. 5). In addition, *InstaText* claims to go much further than *Grammarly* by improving styling and word choice, correcting grammatical errors, and enriching the content to make it more readable and understandable (Para. 8). Furthermore, the researchers being practicing translators and translator trainers thought that *InstaText* could be very useful in the context of Iran's translation market, where



there is high demand for translation of academic papers from Persian into English in a native-like manner, while the translators are native speakers of the source language rather than the target language.

3.3.4 Quality Assessment Rubric

In order to score the pretests and posttests, the researchers used a quality assessment rubric derived from *LISA Quality Assurance Model 3.1*, which LISA recommends for anyone needing an objective measure of translation and localization quality, including reviewers who want to evaluate translated texts (Martínez, 2014). The said rubric is comprised of several error categories, i.e., accuracy, terminology, language, country standards, formatting, and project manager instructions, and weighs errors as minor, medium, major, duplicate, and preferential.

However, out of the various error categories contained in the Quality Assessment Rubric, only quality issues under the *Language Category* were spotted and scored in this study. These issues include 1. Grammar, i.e., failure to follow target-language-specific rules of grammar, syntax, and punctuation; 2. Spelling, namely, misspellings, typographic errors, and incorrect accentuation and capitalization, and 3. Style, i.e., wrong register, inappropriate level of formality, not following style conventions, and unidiomatic usage of the target language.

3.3.5 USE Questionnaire: Usefulness, Satisfaction, and Ease of Use

Introduced by Lund (2001), the *USE Questionnaire* consists of three dimensions, namely, *Usefulness* (items 1-8), *Satisfaction* (items 24-30), and *Ease of Use* (items 9-23). The questionnaire is made up of 30 seven-point Likert Scale items, where point 1 stands for *strongly disagree* and point 7 stands for *strongly agree*. In addition, each item is provided with an *NA* option which means the item does not apply to a specific case. It must be noted that *Ease of Use* is further subdivided into *Ease of Use* (items 9-19) and *Ease of Learning* (items 20-23). Other than that, there are two open-ended questions, i.e., *List the most noticeable weakness(es)* and *list the most significant strength(s)*. Last but not least, the questionnaire is available as an online form, which respondents can fill in and email directly.

Faria et al. (2016) maintain that the evaluation aspects presented in the *USE Questionnaire* are the most important factors to be used for evaluating software usability, and Lund (2001) states that the items are so easy-to-understand that the respondents can understand them with little training. The questionnaire has been used widely by several researchers (e.g., Faria et al., 2016; Salameh, 2017). Last but not least, the questionnaire has a public domain license, which means researchers do not have to pay for purchasing it.

3.4 Data Collection and Analysis Procedure

A total number of 98 applicants had registered for the above-mentioned admission exam, which was administered online via *Google Forms* and comprised of different sections. However, the data from the following sections were collected for this study: 1. exam instructions; 2. applicant's details and demographic information; 3. a Persian-English translation test of academic papers (502 words and 100 minutes); 4. editing the Persian-English translation test of academic papers with *InstaText* (50 minutes); and 5. answering the *USE Questionnaire* (20 minutes). It must be noted that, in the interest of respecting research ethics, it was explained that sections 4 and 5 were optional and their data would be used for a research project. Further, to encourage and compensate the volunteers' time and efforts spent on these optional sections, one of the researchers held a free 4-hour workshop on the use of parallel corpora in translation.

The actual exam was held on Friday, January 15, 2021, from 9:00 A.M to 2:20 P.M. The applicants took the exam online at home. In the end, out of the 98 people who had initially registered for the exam, only 92 applicants took the exam. Upon reviewing the responses, we found that only 75 applicants out of the 92 had completed the optional sections 4 and 5, which provided the data for the present study.

Following the exam, the relevant sections of the exam which provided the required data for the present study were exported in CSV format. These included demographic information, the pretest, the posttest, and the *USE Questionnaire*. First of all, all of the participants were anonymized with P1, P2, P3, etc. codes, and their demographic information was extracted.

In order to extract the data for the one-group pretest-posttest phase, we needed to evaluate the pretest and the posttest. Due to the fact evaluating 75 pretests and 75 posttests would prove too arduous a task, we decided to suffice to 15 samples. Choosing these samples was done using *Research Randomizer*, an online randomizing tool, which gave us the following numbers: P4, P10, P12, P23, P26, P40, P45, P46, P54, P58, P60, P62, P63, P66, and P68. Then the authors themselves evaluated the aforementioned participants' pretests and posttests in terms of the *Language Category* of the *Quality Assessment Rubric*. Subsequently, each participant's scores on the pretest and posttest were compared to determine to what extent *InstaText* might have aided that participant in improving their Persian-English translation in terms of grammar, spelling, and style. This was done using a t-test.

In order to investigate the participants' views about *InstaText*, each individual seven-point Likert scale item was analyzed separately by measuring the



percentage for each scale. Then, the overall means for all four sections of the questionnaire are presented, followed by the total mean of the questionnaire.

As regards the second phase of the study, the scores of all 15 participants in both the pretest and posttest and their mean differences are presented, as well as their scores on the subsections of style, grammar, and spelling. Following these, the data were checked for normality through the Shapiro-Wilk test. Finally, paired-sample t-tests and Wilcoxon Signed Ranks test were employed to find out if InstaText had influenced the participants' performance in the grammar section, style section, spelling section, and total test.

4. Results

4.1 Users' Questionnaire Results

As mentioned before, to examine the attitudes of the participants toward *InstaText*, the responses of 75 participants who used this online tool during the course were recorded. The questionnaire consisted of 4 parts, namely, *usefulness* (1-8), *ease of use* (9-19), *ease of learning*, (20-23), and *satisfaction* (24-30). The results of each section are presented separately. First of all, the descriptive statistics of the *Usefulness* section, comprising 8 items, are presented in Table 2.

Table 2
Descriptive Statistics of Usefulness

Number	Item	Mean
1.	It assists you to be more effective.	5.52
2.	It assists you to be more productive.	5.42
3.	It is a useful app.	5.84
4.	Using it gives me more control over my activities.	4.44
5.	I accomplish the thing I want more easily.	5.54
6.	I save time when using it.	5.67
7.	It fulfills my needs.	5.30
8.	It meets all my related expectations.	5.12
Mean of the category		5.37

Since the number of points on the scale ranged from 1 (strongly disagree) to 7 (strongly agree), it can be seen that all seven items have been ranked more than average. This shows that the participants believed the software was really useful (5.84, the highest mean) and could help them be more productive and effective in their translation, save their time, reach their goals; and, above all, it met all their needs. The highest mean belonged to item number 3, which is at the heart of this section. However, the lowest mean was related to item number 8, stating that it could do everything I expected, which is quite acceptable since every online tool has some limitations and problems. All in all, the mean for this section was 5.37 out of 7, which shows the participants were happy with this aspect of the software.

The next category in the questionnaire was *Ease of Use*, consisting of 11 items. This aspect, having the largest number of items, can be considered one of the very important aspects of every language learning tool. The descriptive results of this section are shown in Table 3 below.

Table 3
Descriptive Statistics of Ease of Use

Number	Item	Mean
9.	I can easily use it.	6.00
10.	I can use it with no problem.	6.08
11.	It is really user-friendly.	5.98
12.	A few steps are needed to achieve what I want.	5.47
13.	It is flexible.	5.50
14.	Using it is effortless.	5.49
15.	No written instructions are needed for using it.	5.43
16.	There is no inconsistency when I use it.	5.24
17.	Both occasional and regular users like it.	5.38
18.	Recovering from mistakes is quick and easy.	5.68
19.	Every time, I use it successfully.	5.60
Mean of the category		5.63



As the results clearly show, the mean values of all the items are more than 5, with the lowest mean belonging to item 16 at 5.24, which is not low in its turn. This item is related to the inconsistencies while using InstaText. However, the results show that it is both easy to use, simple to use, and user-friendly (6.00, 6.08, and 5.98 respectively). Items 12, 13, 14, 15, 18, and 19 can be considered middle-ranked items in this category, with their means ranging from 5.43 to 5.68. The total mean for this category is 5.63, which is the second-highest mean among the four categories.

The category *Ease of Use* is closely linked with another category called *Ease of Learning*, forming the *Ease* aspect of the questionnaire. The category *Ease of Learning* had the lowest number of items (only 4 items), but it enjoyed the highest total mean among the 4 categories (6.03).

Table 4
Descriptive Statistics of Ease of Learning

Number	Item	Mean
20.	I quickly learned how to use it.	6.04
21.	I can remember easily how I should use it.	6.14
22.	I could easily learn to use it.	6.13
23.	I became a skilled user quickly.	5.78
Mean of the category		6.03

As the statistics in Table 4 show, the participants did not experience any difficulty in learning how to use *InstaText*, became skillful with it, and could easily remember how to use it. Thus, it is safe to say that besides ease of use, the application was easy to learn too. These two aspects can make it one of the best options for writing and translation teachers while teaching. The last seven items in the questionnaire belonged to the *Satisfaction* category. Table 5 presents the data related to this category and the total mean of the questionnaire.

Table 5
Descriptive Statistics of Satisfaction

Number	Item	Mean
24.	It is satisfying.	5.74
25.	I recommend it to my friends.	5.81

26.	Using it is fun.	5.57
27.	It works exactly in the way I like.	5.45
28.	This app is wonderful.	5.48
29.	It is necessary to have it.	5.60
30.	Using it is pleasant.	5.59
Mean of the category		5.60
Total Mean of the questionnaire		5.61

The analysis of the responses revealed that the participants were highly satisfied with *InstaText*. Although this category was ranked third among the four categories, the total mean of the category was quite high (5.60). The participants were satisfied with the application (5.74), felt the need to have it (5.60), and were even ready to recommend it to their friends (5.81). The last point related to the questionnaire is that the total mean of the items was 5.61, which shows that this application was easy to use and easy to learn, proved useful, and was regarded as satisfactory.

4.2 Pretest-Posttest Results

As mentioned before, the participants translated a text from Persian to English (pretest) and then used *InstaText* for editing the same text (posttest). In order to see to what extent *InstaText* had been effective in improving the quality of the participants' second text, the participants' scores in the pretest (including grammar, spelling, and style) were compared with their scores in the posttest with the same sections. Besides, the mean difference for each of these sections is shown in Table 6. All these details are shown for each of the fifteen participants separately. It must be noted that due to the nature of the AFPs, in general, and *InstaText*, in particular, the users are provided with corrective suggestions that they can choose to accept or dismiss. Therefore, the results having been reported here do not indicate which suggestions they have accepted and which ones they have dismissed. Our judgments of the efficacy of *InstaText* for improving the three sections are based on the scores which the participants obtained on the pretest and the post-test.

Table 6

Participants' Scores on the Three Sections of the Pretest and Posttest



Participant Code	Section of the Test	Scores on Pretest	Scores on Posttest	Mean Difference
P4	Grammar	57	86	+29
	Spelling	100	100	0
	Style	57	68	+11
P10	Grammar	77	76	-1
	Spelling	100	100	0
	Style	49	56	+7
P12	Grammar	62	76	+14
	Spelling	100	99	-1
	Style	52	56	+4
P23	Grammar	39	57	+18
	Spelling	91	90	-1
	Style	64	45	-19
P26	Grammar	77	87	+10
	Spelling	100	100	0
	Style	53	57	+4
P40	Grammar	36	49	+13
	Spelling	87	98	+11
	Style	59	59	0
P45	Grammar	87	77	-10
	Spelling	100	100	0
	Style	53	62	+9

P46	Grammar	36	43	+7
	Spelling	90	99	+9
	Style	40	46	+6
P54	Grammar	36	67	+31
	Spelling	89	88	-1
	Style	51	60	+9
P58	Grammar	26	33	+7
	Spelling	100	100	0
	Style	39	43	+4
P60	Grammar	77	77	0
	Spelling	100	100	0
	Style	46	57	+11
P62	Grammar	87	68	-19
	Spelling	100	100	0
	Style	48	56	+8
P63	Grammar	18	32	+14
	Spelling	99	98	-1
	Style	52	58	+6
P66	Grammar	10	17	+7
	Spelling	81	100	+19
	Style	39	43	+4
P68	Grammar	87	57	-30
	Spelling	100	100	0



Style

53

62

+9

The results show that the scores of most of the participants, 10 out of 15, in the grammar section improved. The amount of mean difference ranged from 7 to 31. One of the participants, P60, did not have any improvement in his writing score. On the downside, three participants had a lower mark in their grammar posttest, with their scores decreasing by 1, 19, and 30 scores respectively. Thus, it seems that InstaText could help most participants write more accurate sentences in the posttest. With regard to spelling, one can clearly see that more than 50% of the participants did not experience much change in their spelling scores, while the scores of four participants decreased. Only three participants, P40, P46, and P66, had better scores in their spelling posttest. Finally, it was revealed that the biggest change was related to the scores of participants in the style section of the test. Twelve participants out of fifteen took advantage of InstaText in improving their style, with mean differences ranging from 4 to 11. One participant, P40, did not change at all in his style score, while another participant, P23, used InstaText to deteriorate the style of his writing.

In order to find out if using InstaText helped participants improve the quality of their writing in general and in three specific areas of grammar, spelling, and style; the researchers compared the results of the pretest and posttest. Before doing so, the data were checked for normality. Using the Shapiro-Wilk test which is suitable for samples fewer than 50, it was found that pretest scores, posttest scores, grammar pretest, grammar posttest, style pretest, and style posttest were all normal with significance values of .16, .61, .09, .25, .41, and .07 respectively. However, the spelling pretest and spelling posttest scores were not normally distributed. Thus, a series of paired-sample t-tests were run to find out if using InstaText had been effective in the aspects displaying normal data. The results of these t-tests are presented in Table 7.

Table 7
Paired-sample T-test Results for Grammar Section, Style Section, and Totals Score

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig.
				Lower	Upper			
Pair 1	-4.40	5.74	1.48	-7.58	-1.21	-2.96	14	.01

PairGrammar	pretest--6.00	16.37	4.22	-15.07	3.07	-1.41	14	.17	
2 Grammar	posttest								
PairStyle	pretest-	Style-4.86	7.26	1.87	-8.89	-.84	-2.59	14	.02
3	posttest								

As the results in Table 7 show, a paired-sample t-test was run to compare the overall quality of the texts written by participants before and after using InstaText. On average, the learners performed worse before ($M = 66.75$, $SD = 11.30$) than after using InstaText ($M = 71.15$, $SD = 9.13$). This improvement, 4.40, was statistically significant, $t(14) = -2.96$, $p = 01$. Therefore, it was concluded that InstaText could significantly change the scores of participants from the pretest to the posttest. Considering the fact that the spelling section, which accounts for one-third of the performance, did not change much as a result of treatment, this improvement must be regarded as really significant.

As regards the grammar section, the results from the pre-test ($M = 54.13$, $SD = 26.91$) and post-test ($M = 60.13$, $SD = 21.37$) indicated that using InstaText did not result in a significant improvement in the accuracy of the texts, $t(14) = -1.41$, $p = .17$. Finally with regard to style, the results from the pre-test ($M = 50.33$, $SD = 7.18$) and post-test ($M = 55.20$, $SD = 7.54$) indicated that using InstaText led to a significant improvement in the style of the texts, $t(14) = -2.59$, $p = .02$. Thus, it can be concluded that using InstaText helped participants improve the overall quality of their texts and their style, but it could not help them write more accurate sentences.

The last point about the effect of using InstaText is related to the spelling section. Since the data of the spelling section was not normally distributed, a Wilcoxon signed-rank test was run to check for the possible difference between the spelling pretest and posttest. The results of the Wilcoxon signed-rank test showed that using the said tool did not elicit a statistically significant change in the spelling scores of the learners ($Z = -.689$, $p = 0.491$).

5. Discussion

The current study was conducted with the purpose of investigating both the effect of *InstaText* on improving language quality and users' perceptions toward it. The results revealed that the participants perceived the tool as really useful in that it helped them be more productive and effective in their translation, save their time, and reach their goals. Besides, it was found that *InstaText* was not only easy to use but also easy to learn since the participants did not experience any difficulty in learning how to use it. Added to that, the participants were so satisfied with the tool that they said they would highly recommend it to their colleagues. These findings, which are in line with those of most previous studies (e.g., [Franzke et](#)



al., 2005; Graham et al., 2015; Tang & Rich, 2017; Palermo & Thomson, 2018), may be attributed to three factors: 1. *InstaText* is an easily accessible cloud-based tool; 2. both its editing process and the relevant toolkit are very intuitive; and 3. its editing environment is quite similar to that of *Microsoft Word*, meaning that translators, being already adept at using *Microsoft Word*, can easily migrate to. In addition, the participants' satisfaction with the tool lends support to *InstaText's* claim that it goes much further than *Grammarly* by improving style and word choice, correcting grammatical errors, and enriching the content to make it more readable and understandable (*InstaText*, 2021, Paras. 5,8). It must be noted that the foregoing findings are in contrast to those of Palermo and Thomson (2018), Ranalli (2018), and Wilson and Roscoe (2020).

In the pretest-posttest phase of the study, we found that *InstaText* did not help the participants make significant progress in grammar and spelling, which contradicts the findings of most previous studies (e.g., Franzke et al., 2005; Kellogg et al., 2010; Wilson & Czik, 2016; Wilson, 2017). In fact, *InstaText* did help improve 10 participants' scores tangibly (from +7 to +31); still, such improvement was traded off for the decrease in 4 participants' scores (-1 to -30). Further, such a finding can be attributed to the fact that the participants were freelance translators, which means they were already good at grammar. The reason why we could not observe any significant improvement in spelling is that the participants typed their translations in *Google Forms*, which suggests corrections for misspelled words. This is supported by 8 participants' full spelling scores in their pretests, meaning that no spelling improvement was needed. Having a closer look at Table 6, we can see for 3 participants, *InstaText* improved their spelling scores noticeably (from +9 to +19). As regards the remaining 4 participants whose spelling scores were reduced slightly as a result of using *InstaText*, the reduced scores were related to extra blank spaces, included in our *TQA Rubric* but not amended by *InstaText*. On the other hand, the participants made remarkable progress in terms of style in the wake of editing their translations with *InstaText*, which not only supports Wilson and Czik (2016), Palermo and Thomson (2018), and Wilson and Roscoe (2020) but also corroborates *InstaText's* claim that it is capable of helping the user produce clear and efficient sentences with improved style and word choice (*InstaText*, 2021) by suggesting lots of amendments to the user to accept or reject. Overall, using the tool resulted in remarkably higher scores in the posttest as compared to the pretest, which is consistent with Graham (2006), Graham and Perin (2007) Graham et al. (2012), and Wilson and Roscoe (2020). Such significant improvement in overall scores could be because AWE tools prove more fruitful with more proficient users (Koltovskaia, 2020) and because they help better users' writing attitudes (Roscoe et al., 2018; Camacho et al., 2020).

6. Conclusion

The present study showed that the participants perceived *InstaText* as a useful, easy-to-use, and satisfactory assistant for editing their Persian-to-English translations. This was also confirmed by the positive effect of the tool on improving language quality in the pretest-posttest phase of the study. Therefore, *InstaText* can be regarded as an AFP capable of helping both students and freelance translators to exert metacognitive control over the central cognitive processes involved in writing through the provision of effective and frequent feedback immediately and in a localized, specific, and detailed manner.

The findings of this study have to be seen in the light of some limitations as follows. First, we conducted the study using the one-group pretest-posttest design, not being free from threats to the internal and external validity; hence, it is suggested that future studies on *InstaText* use the true experimental design. Second, the participants of the present study were selected using volunteer convenience sampling, so the study can be replicated using more systematic sampling methods. Third, we studied *InstaText* from the perspective of freelance Persian-English translators using a questionnaire. Hence, we recommend that the tool be evaluated from students, teachers, and professional writers' viewpoints by using in-depth interviews.

The findings of the study are expected to contribute to our knowledge by signifying translators' attitudes toward AFPs and helping discover the strengths and weaknesses of *InstaText*. The study can also enable researchers as well as teachers to have a better picture of how AFPs, in general, and *InstaText*, in particular, could contribute to improving writing quality and ability. With such knowledge, researchers, curriculum designers, and teachers would be able to develop specific, appropriate, and creative pedagogical methods for making effective use of online AFPs.

Acknowledgments: The researchers are really thankful to all the participants who stayed with us until the end of the research and helped us with data collection.

References

- Afshari, S., & Salehi, H. (2017). Effects of using Inspiration software on Iranian EFL learners' prewriting strategies. *International Journal of Research*, 6(2), 1-11. 10.5861/ijrset.2017.1670
- Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 316–329). Guilford.
- Baker, M. (2002). *In other words: A coursebook on translation*. London: Routledge.



- Brindle, M., Graham, S., Harris, K. R., & Hebert, M. (2016). Third and fourth grade teacher's classroom practices in writing: A national survey. *Reading and Writing*, 29(5), 929–954. <https://doi.org/10.1007/s11145-015-9604-x>
- Byrne, J. (2007). Caveat translator: Understanding the legal consequences of errors in professional translation. *Journal of Specialised Translation*, 7, 2-24.
- Byrne, J. (2010). Are technical translators writing themselves out of existence? In I. Kemble (Ed.) *The translator as a writer* (pp. 14–27). Portsmouth: University of Portsmouth.
- Camacho, A., Alves, R. A., & Boscolo, P. (2020). Writing motivation in school: A systematic review of empirical research in the early twenty-first century. *Educational Psychology Review*, 33, 213–247. <https://doi.org/10.1007/s10648-020-09530-4>
- Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio books.
- Campbell, S. (1998). *Translation into the second language*. London and New York: Longman.
- Casal, J. E. (2016). Criterion online writing evaluation. *calico journal*, 33(1), 146-155.
- Chen, C. F. E., & Cheng, W. Y. E. C. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94-112.
- Chou, H. N. C., Moslehpour, M., & Yang, C. Y. (2016). My access and writing error corrections of EFL college pre-intermediate students. *International Journal of Education*, 8(1), 144-161. <https://doi:10.5296/ije.v8i1.9209>
- Daniels, P., & Leslie, D. (2013). Grammar software ready for EFL writers? *OnCue Journal*, 9(4), 391–401.
- Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal*, 28(1), 99–134. <https://doi.org/10.11139/cj.28.1.99-134>
- Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22, 1–17. <https://doi.org/10.1016/j.asw.2014.03.006>

- Duběda, T., & Obdržálková, V. (2021). Stylistic competence in L2 translation: stylometry and error analysis. *The Interpreter and Translator Trainer*, 15(2), 172-186.
- Faria, T. V., Pavanelli, M., & Bernardes, J. L. (2016, July). Evaluating the usability using USE questionnaire: Mindboard system use case. In *International Conference on Learning and Collaboration Technologies* (pp. 518-527). Springer, Cham. https://doi.org/10.1007/978-3-319-39483-1_47
- Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., and Dooley, S. (2005). Summary Street®: Computer support for comprehension and writing. *Journal of Educational Computing Research*, 33, 53–80. <https://doi.org/10.2190/DH8F-QJWM-J457-FQVB>
- Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York: Guilford. <https://doi.org/10.1007/s11145-008-9121-2>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Graham, S., Harris, K. R., Fishman, E., Houston, J., Wijekumar, K., Lei, P. W., & Ray, A. B. (2019). Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *The Elementary School Journal*, 119(3), 487-510. <https://doi.org/10.1086/701720>
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, 115(4), 523-547.
- Graham, S., McKeown, D., Kiuvara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, 104(4), 879–896. <https://doi.org/10.1037/a0029185>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. <https://doi.org/10.3102/003465430298487>
- Hyland, K., & Hyland, F. (2006). *Feedback in second language writing: Contexts and issues*. Cambridge University.



- InstaText (2021). Write like a native speaker (Webpage). Retrieved from <https://instatext.io/>, Date of Access: 02 January 2021
- Just Publishing Advice (2020). 50 Free Writing Software Tools and the Best Free Writing Apps (Webpage). Retrieved from https://justpublishingadvice.com/free-writing-software-and-the-best-free-writers-tools/#6_Free_writing_apps_for_accuracy, Date of Access: 20 October 2020
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1, 1–26. <https://doi.org/10.1080/00461520903213600>
- Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, 44(4), 250–266. <https://doi.org/10.1080/00461520903213600>
- Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students to write? *Journal of Educational Computing Research*, 42, 173–196. <https://doi.org/10.2190/EC.42.2.c>
- Khoii, R., & Doroudian, A. (2013). Automated scoring of EFL learners' written performance: a torture or a blessing. In *Conference proceedings. ICT for language learning* (p. 367). libreriauniversitaria.it Edizioni.
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450.
- Koroglu, Z. C. (2014). An analysis of grammatical errors of Turkish EFL students' written texts. Retrieved from http://www.turkishstudies.net/Makaleler/374851580_8%C3%87etinK%C3%B6ro%C4%9FluZeynep-edb-101-111.pdf
- Lastari, D. S. (2021). The effect of ginger software on the tenth grade students' writing skill. *Globish: An English-Indonesian Journal for English, Education, and Culture*, 10(2), 12-18.
- Lund, A. M. (2001). Measuring usability with the use questionnaire. *Usability interface*, 8(2), 3-6.
- Martínez, R. (2014). A deeper look into metrics for translation quality assessment (TQA): A case study. *Miscelánea: A Journal of English and American Studies*, 49, 73-93

- Matsumura, L. C., Patthey-Chavez, G. G., Valdés, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *The Elementary School Journal*, 103, 3–25. <https://doi.org/10.1086/499713>
- Mekheimer, M. (2005). *Effects of Internet-based Instruction, Using Webquesting and emailing on developing some essay writing skills in student teachers* (Unpublished doctoral dissertation). Cairo University.
- Muller, A., Gregoric, C., & Rowland, D. R. (2017). The impact of explicit instruction and corrective feedback on ESL postgraduate students' grammar in academic writing. *Journal of Academic Language and Learning*, 11(1), A125-A144. Retrieved from <https://journal.aall.org.au/index.php/jall/article/view/442>
- Murray, N. (2010). Conceptualizing the English language needs of first-year university students. *The International Journal of the First Year in Higher Education*, 1(1), 55–64.
- Narita, M. (2012). Developing a corpus-based online grammar tutorial prototype. *Language Teacher*, 36(5), 23-31.
- Newmark, P. (1995). *Approaches to translation*. New York: Phoenix ELT.
- Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2021). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, 1–22. <https://doi.org/10.1111/jcal.12635>
- Olshtain, E. (2001). Functional tasks for mastering the mechanics of writing and going just beyond. In M. Celce-Murcia (ed.), *Teaching English as a second or foreign language* (pp. 207-217). Boston, MA: Heinle & Heinle.
- Palermo, C., & Thomson, M. M. (2018). Teacher implementation of self-regulated strategy development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology*, 54, 255–270. <https://doi.org/10.1016/j.cedpsych.2018.07.002>
- Palermo, C., & Wilson, J. (2020). Implementing Automated Writing Evaluation in Different Instructional Contexts: A Mixed-Methods Study. *Journal of Writing Research*, 12(1).
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for



- teaching and learning and student progress. *Assessing Writing*, 15, 68–85. <https://doi.org/10.1016/j.asw.2010.05.004>
- Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology*, 108(8), 1098–1120. <https://doi.org/10.1037/edu0000103>
- Perdana, I., & Farida, M. (2019). Online grammar checkers and their use for EFL writing. *Journal of English Teaching, Applied Linguistics and Literatures (JETALL)*, 2(2), 67-76. DOI: <http://dx.doi.org/10.20527/jetall.v2i2.7332>
- Qian, L., Zhao, Y., & Cheng, Y. (2020). Evaluating China's automated essay scoring system iWrite. *Journal of Educational Computing Research*, 58(4), 771-790.
- Rahimi, A., Jahangard, A., & Norouzizadeh, M. (2020). Students' attitudes towards computer-assisted language learning and its effect on their EFL writing. *International Journal of Learning and Teaching*, 12(3), 144-152. DOI: 10.18844/ijlt.v12i3.4767
- Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 31, 653–674. <https://doi.org/10.1080/09588221.2018.1428994>
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1). <https://doi.org/10.1080/01443410.2015.1136407>.
- Rinaldo, S. B., Tapp, S., & Laverie, D. A. (2011). Learning by Tweeting: Using Twitter as a pedagogical tool. *Journal of Marketing Education*, 33(2), 193-203. <https://doi.org/10.1177/0273475311410852>
- Roscoe, R. D., Allen, L. K., Johnson, A. C., & McNamara, D. S. (2018). Automated writing instruction and feedback: Instructional mode, attitudes, and revising. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2089–2093. <https://doi.org/10.1177/1541931218621471>
- Salameh, Z. (2017). Attitudes towards Facebook and the Use of Knowledge and Skills among Students in the English Department at the University of Hail. *Journal of Education and Practice*, 8(8), 1-6.

- Shermis, M. D., & Burstein, J. (2003). Introduction. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. xiii–xvi). Lawrence Erlbaum Associates.
- Stevenson, M., & Phakiti, A. (2019). Automated feedback and second language writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing* (pp. 125–142). Cambridge University Press. <https://doi.org/10.1017/9781108635547.009>
- Tang, J., & Rich, C. S. (2017). Automated writing evaluation in an EFL setting: Lessons from China. *JALT CALL Journal*, 13(2), 117–143. <https://doi.org/10.29140/jaltcall.v13n2.215>
- Tynan, L., & Johns, K. (2015). Piloting the post-entry language assessment: Outcomes from a new system for supporting research candidates with English as an additional language. *Quality in Higher Education*, 21(1), 66–78. <http://dx.doi.org/10.1080/13538322.2015.1049442>
- Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A., & Quintana, R. (2020). eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing*, 44, 100449. <https://doi.org/10.1016/j.asw.2020.100449>
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36. <https://doi.org/10.1080/15544800701771580>
- Williams, J. (2004). Tutoring and revision: Second language writers in the writing center. *Journal of Second Language Writing*, 13(3), 173–201. <https://doi.org/10.1016/j.jslw.2004.04.009>
- Wilson, J. (2017). Associated effects of automated essay evaluation software on growth in writing quality for students with and without disabilities. *Reading and Writing*, 30(4), 691–718.
- Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers and Education*, 100, 94–109. <https://doi.org/10.1016/j.compedu.2016.05.004>
- Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58(1), 87–125. <https://doi.org/10.1177/0735633119830764>



Wilson, J., Myers, M. C., & Potter, A. (2022). Investigating the promise of automated writing evaluation for supporting formative writing assessment at scale. *Assessment in Education: Principles, Policy & Practice*, 1-17.



