Universitat Oberta de Catalunya

Faculty of Arts and Humanities

MA in Technology-Mediated Language Teaching and Learning

Raising university students' awareness of Automated Writing Evaluation tools in the English as a Foreign Language writing classroom to promote uptake of feedback and learner autonomy

Matthew Schlosser

mschlosser@uoc.edu

16/06/2023

Tutor/Supervisor: Jennifer Rose Ament

Final Master's Thesis

Year 2022-2023, semester 2

Abstract

If EFL students are expected to improve their writing over the course of a term at university, they must be provided with formative assessment opportunities; when it comes to writing, this involves establishing multi-step feedback loops that allow learners to respond to instructor's feedback, rewrite and resubmit for further feedback. Providing truly effective feedback of this nature has been proven to help improve students' writing scores, but it also requires more commitment from instructors, both in terms of effort and time, the latter of which is often lacking, especially in a university context. Incorporating Automatic Writing Evaluation (AWE) tools may be part of the solution to this problem; this Final Master's Project (FMP) puts forward the notion that AWE tools, when incorporated effectively in the university EFL classroom, can help instructors establish feedback loops as they provide students with a kind of initial proofreader that can clean up some of their more mechanical errors. In this FMP, first-and-second-year university students are introduced to the Artificial Intelligence-powered tools *Write and Improve*, *Grammarly* and *Chat GPT* to promote their uptake of feedback in general, raising their awareness of the type of errors they make and making them more independent learners. The study, involving two groups of students at a private Spanish university, outlines how these tools can be used in the EFL classroom and attempts to gather students' perceptions of them and to what extent they took the computer-generated feedback on board.

Keywords: AWE tools; student uptake of feedback; learner autonomy; feedback loops; Artificial Intelligence in EFL

Resumen

Si se espera que los estudiantes de inglés como lengua extranjera mejoren su escritura a lo largo de un trimestre en la universidad, se les deben brindar oportunidades de evaluación formativa; en lo que respecta a la escritura, esto implica establecer bucles de retroalimentación de varios pasos que permitan a los estudiantes responder a los comentarios del instructor, reescribir y volver a enviar para obtener más retroalimentación. Proporcionar retroalimentación verdaderamente efectiva de esta naturaleza se ha demostrado que ayuda a mejorar las calificaciones de escritura de los estudiantes, pero también requiere un mayor compromiso por parte de los instructores, tanto en términos de esfuerzo como de tiempo, este último a menudo escasea, especialmente en el contexto universitario. La incorporación de herramientas de Evaluación Automática de Escritura (AWE) puede ser parte de la solución a este problema; este Trabajo Fin de Máster (TFM) plantea la idea de que las herramientas de AWE, cuando se incorporan de manera efectiva en el aula universitaria de inglés como lengua extranjera, pueden ayudar a los instructores a establecer bucles de retroalimentación al proporcionar a los estudiantes una especie de corrector inicial que puede corregir algunos de sus errores más mecánicos. En este PFM, se introduce a estudiantes universitarios de primer y segundo año en una universidad privada española en herramientas impulsadas por Inteligencia Artificial, como *Write and Improve*, *Grammarly* y *Chat GPT*, con la intención de promover su aceptación de la retroalimentación en general, aumentar su conciencia sobre el tipo de errores que cometen y convertirlos en estudiantes más independientes. El estudio, que involucra a dos grupos de estudiantes en una universidad española privada, describe cómo se pueden utilizar estas herramientas en el aula de inglés como lengua extranjera y trata de

recopilar las percepciones de los estudiantes sobre ellas y en qué medida tomaron en cuenta la retroalimentación generada por la computadora.

Palabras clave: Herramientas de Evaluación Automática de Escritura (AWE); Aceptación de retroalimentación por parte del estudiante; Autonomía del estudiante; Bucles de retroalimentación; Inteligencia Artificial en el aprendizaje del inglés como lengua extranjera

TABLE OF CONTENTS

1. INTRODUCTION 7

1.1. Justification 9

2. THEORETICAL FRAMEWORK 11

2.1. The importance of WCF: the how and the why 11

2.2. The potential and limitations of AWE tools 13

2.3. Common uses of AWE tools in the classroom 15

2.4. Student perception of AWE tools 16

3. OBJECTIVES 16

4. METHODOLOGY 17

4.1. A comparative analysis of the feedback provided by the three AWE tools (RQ1) 17

4.2. Learner autonomy and student uptake of feedback provided by the three AWE tools (RQs 2 & 3) 23

4.2.1. Learning context and participants 23

4.2.2. Design 23

4.2.3. Procedure 24

4.2.4. Instruments used for data collection 26

5. RESULTS / EVALUATION OF THE INTERVENTION 26

5.1. Strengths and weaknesses of each AWE tool used with students (Q1) 26

5.2. Learner autonomy (Q2) 28

5.3. Student uptake of feedback (Q3) 29

6. CONCLUSIONS 29

6.1. Discussion 29

6.2. Limitations 30

6.3. Recommendations 30

REFERENCES 32

APPENDICES 37

LIST OF ABBREVIATIONS

Abbreviation	Definition
AEE	Automated Essay Evaluation
AWE	Automated Writing Evaluation
CGF	Computer Generated Feedback
MKO	More Knowledgeable Other
RQ	Research Question
TGF	Teacher Generated Feedback
WCF	Written Corrective Feedback

INDEX OF FIGURES AND TABLES

Number	Title	Page
Figure 1	The writing process	8
Figure 2	Typical direct feedback from Grammarly	10
Figure 3	Part of a chat with Chat GPT	10
Figure 4	Teacher feedback on student's letter of application	18
Figure 5	*Grammarly* feedback on student's letter of application	19
Figure 6	*Grammarly*'s short report on user performance	20
Figure 7	Write & Improve's feedback on student's letter of application	21
Figure 8	List of errors flagged and corrected by Chat GPT	22
Table 1	Methodological/pedagogical design of FMP	23
Table 2	Summary of features of TGF and AWE tools	27

1. INTRODUCTION

"Learning to write is a very long process" (Polio & Park, 2016, p. 298). Many language teachers have come to embrace this idea and now put an emphasis on writing as a process rather than simply as a final product. Perhaps unsurprisingly, the so-called process approach -- a term that stems from research done four decades ago that found that cognitively speaking, writing is "a highly complex process, made up various subprocesses that occur not one after another in a strict linear sequence, but cyclically and in varying patterns" (Caudery, 1995, p. 1)-- has established itself as an important teaching/learning resource in the EFL as well as L1 writing classroom. The process approach to this productive skill acknowledges that "writing happens as a recursive progression with different stages that range from pre-writing to editing and where writers exercise different thinking skills in order to shape their work" (Marulanda Ángel & Martínez García, 2017, p. 53). Coffin et al. (2003) outline the ideal writing process (see Figure 1); here, the emphasis is less on the final piece of writing and more on formative assessment, "with teachers providing feedback to pupils on how well they have achieved particular objectives […], and what else they might need to do in order to improve" (Torrance & Pryor, 1998, p. 8). Indeed, an important part of the writing process involves student-teacher interaction (in the form of feedback and revision), and studies reveal "a positive relationship between process-related classroom activities and higher writing scores" (Deane et al., 2008, p. 33).

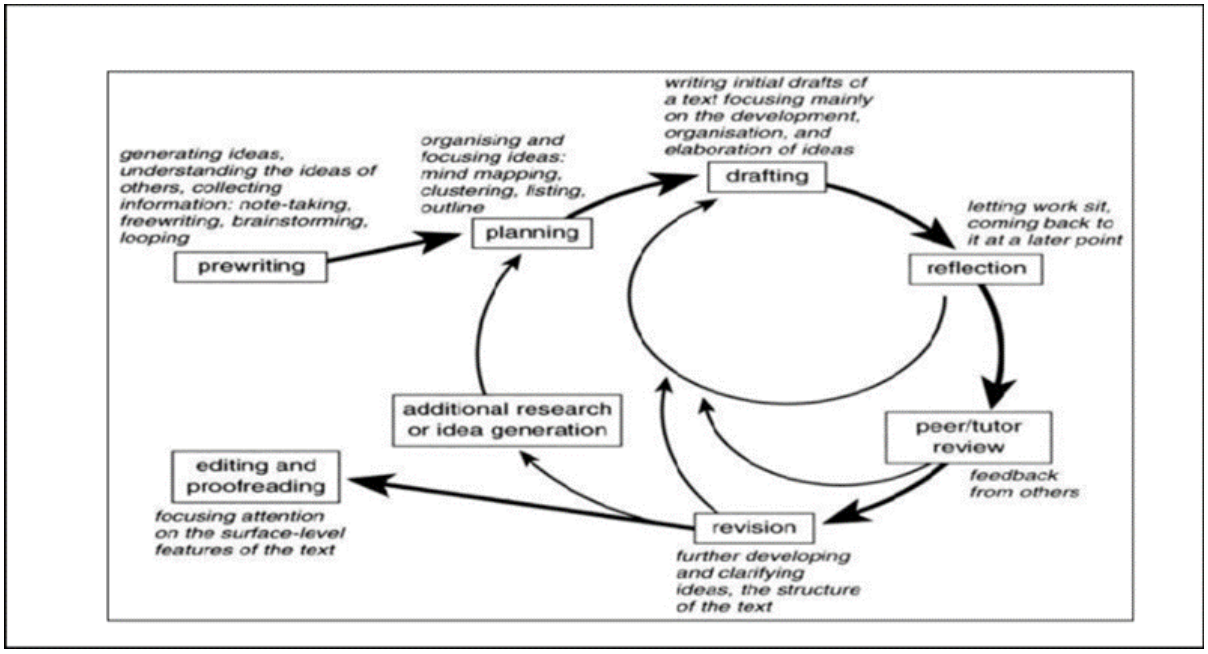
If writing is a process, teachers providing students with effective feedback is "an essential part of the learning journey in foreign language learning" (Harper et al., 2018); research suggests that when students are given and engage with constructive feedback, their writing skills improve significantly (William, 2011). It should come as no surprise then that some researchers go so far as to say that high-quality feedback is the most important influence on students' learning (Hattie, 1987; Black & William, 1998). Studies also suggest that students are aware of the influence feedback has on their progress as language learners; when it comes to their writing, for example, "language learners still strongly request written corrective feedback from their teachers as a way to improve" (Mao & Crosthwaite, 2019, p. 47).

To promote the idea that writing is a process, as well as to help students make the most of the feedback they are given, they can—or should—be given the opportunity to engage with the feedback, make modifications to their writing and resubmit it, either for further feedback or as an evaluative assessment. Allowing students to submit several drafts of their writing "supports a feedback loop, in which students feel seen and supported" (Lee, 2020). Establishing active

feedback loops in the classroom could also be an integral part of promoting what Carless and Boud (2018) refer to as students' feedback literacy, one of the core concepts in this Final Master's Project (FMP).

Figure 1

The writing process



Note: Retrieved from Coffin et al., 2003.

But for as much effort as well-intentioned teachers may make to provide their students with quality—be it direct, indirect, simple or more elaborate—feedback, two nagging questions must be answered:

1) While students *say* they want feedback, to what extent are students actually interacting with the feedback that teachers provide them with? Carless & Boud (2018) put forward the logical notion that teachers must ensure that their students have the feedback literacy to understand and engage with it; feedback literacy is defined as "the understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies" (p. 1316). Carless & Boud (2018) advocate for the use of in-class awareness-raising activities, which have been included as intended learning outcomes in the Methodology section of this master's project.

2) Where should university instructors, who have up to or more than 30 students on any given course and perhaps four or five courses a semester, find the time to provide students with the kind of quality feedback experts advocate for? Very few researchers address the issue of workload constraints when it comes to giving feedback, but this FMP will recognize it as one of the most important issues facing instructors today, as they often find themselves pressed for time and giving feedback to all their students can pose a real challenge (Harper et al., 2018).

1.1. Justification

What can busy teachers do to establish the kind of feedback loops that promote process writing as well as feedback literacy, while not spending an exorbitant amount of time on error-correcting students' work? This is where Artificial Intelligence (AI)-driven Automated Writing Evaluation (AWE) tools like *Write & Improve*, *Grammarly* and *Chat GPT* can be used in the classroom to help provide students with both timely, innovative and personalized corrective feedback on formative assessments, which might in turn free up teachers to focus on giving students direct, descriptive feedback that focuses on *feeding forward*, or what to do next. To be clear, these AWE tools are not meant to be seen as "a teaching aid, not a replacement for teachers" (Harrison, n.d.). When used by learners with the support of a teacher who understands what kind of feedback is provided by the three AWE tools, these resources can help make learners "notice, think and self-reflect" (American English Live, 2020) on the type of errors that they typically make. A brief summary of the tools follows:

1) ***Write & Improve*** is a free website that supplies learners of English with feedback on Cambridge Exam-type writing tasks. English Language iTutoring, in collaboration with Cambridge Assessment, is the company behind the project launched in 2016 that aims to, as the company's co-founder and CTO Paul Butcher writes, "give learners pedagogically sound feedback [on their writing] which allows them to focus their effort where there's the most to be gained, without telling them all the answers" (Harrison, n.d.). Indeed, the feedback that the website gives is *not* very detailed; however, what it does indicate to students is the CEFR level that the piece of writing they submit demonstrates, which does seem to motivate some students to keep using the tool to work on their writing (Mehtrova, 2021). That seems to be a sound reason—after all, experts agree that motivation is "crucial for language learning" (Mackey, 2014)-- to keep encouraging English language learners to use it.

2) According to its website**, *Grammarly*** is used by more than 30 million people around the world every day to "strengthen their writing and say what they really mean" (*About Us* |

Grammarly, n.d.). Like *Write & Improve*, the AWE is powered by AI and, if not downloaded onto a user's computer, in its free version can be used as a kind of writing consultant that proofreads a piece of written text; what sets it apart from *Write & Improve*, is that it gives more detailed direct feedback on errors—while not correcting the error automatically-- and even supplies some kind of explanation for the error (see Figure 2).

Figure 2

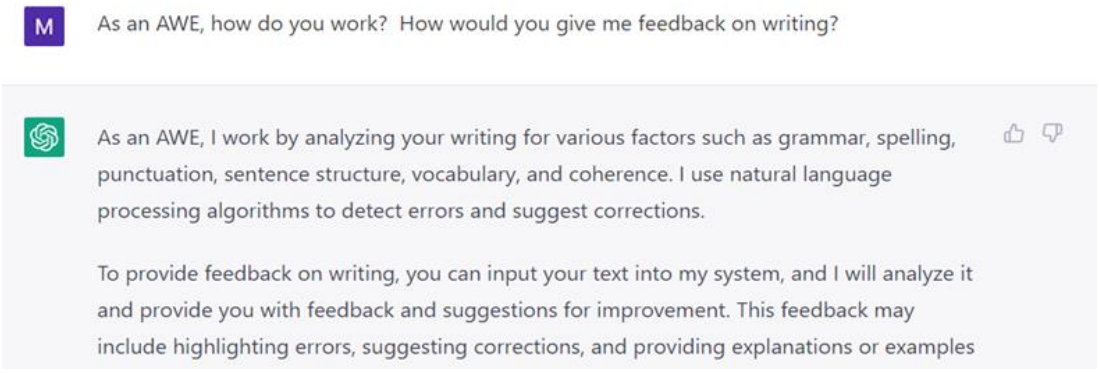
Typical direct feedback from Grammarly

Note: Screenshot taken of feedback provided by Grammarly's Writing Assistant in February 2023

3) ***Chat GPT*** is likely already better known than the other two tech tools by students and teachers alike. When asked how it might be used as an AWE by users, the AI-powered chatbot responded by explaining that it could help identify errors and recommend improvements (see Figure 3). The fact that users can chat with this AWE tool makes it obviously different and arguably more exciting than the other two tools.

Figure 3

Part of a chat with Chat GPT



Note: Screenshot taken of chatopenai.com, February 2023

An in-depth comparative analysis of all three tools is found in Section 4.1. Current themes around AWE tools in the research literature shall be discussed in the next section.

2. THEORETICAL FRAMEWORK

As an increasing number of AI-driven AWE resources like *Grammarly* and *Write & Improve* have hit the market in the last decade, it may come as no surprise that the number of articles and studies devoted to the use of AWE tools in the EFL classroom has also increased significantly since 2015 (Fu et al., 2022); having said that, important studies on Automated Essay Evaluation (AEE) and AWE tools and their impact on student writing and test scores were already being conducted in the 2000s (Chen & Cheng, 2008; Attali, 2004), laying the groundwork for subsequent research. Some commonly recurring themes in the literature include the importance of written corrective feedback (WCF) in the development of students' writing skills, the potential and limitations of AWE tools as test scorers and feedback providers and how teacher-generated feedback (TGF) compares to computer-generated feedback (CGF), how AWE and AEE tools tend to be used in the classroom, and students' perception of feedback supplied by the tools.

2.1. The importance of WCF: the how and the why

Whether WCF should be direct or indirect has been a constant source of debate for instructors and researchers for years (O'Neill & Russel, 2018). While there are studies that show that *direct* feedback—where teachers simply provide students with the correction—can be effective (Kloss & Quintanilla, 2022), the consensus among researchers appears to be that *indirect* error correction—where teachers indicate that an error is present without providing the correction—is generally more effective because, among other reasons, "it forces learners to engage in guided learning and problem-solving, leading to self-reflection and 'noticing', which foster increased acquisition" (Mao & Crosthwaite, 2019, p. 47). Moser concurs, acknowledging that when learners engage cognitively with the indirect feedback they have been given, they are helped "to recognize certain error patterns", which "deepens their existing knowledge of the foreign language" (2020, p. 54). Recent SLA studies advocate for more elaborate indirect feedback: feedback seems to be *most* effective when accompanied by metalinguistic explanations, or "rule reminders" (Ferris, 2010, p. 193). Hartshorn et al. (2010) advocate for what they call "dynamic WCF", which consists of an ambitious list of best practices, from timeliness to constancy, that create a kind of multi-step "feedback cycle" with meaningful, frequent (and time-consuming) teacher correction and student revision (p. 90). While many teachers may wonder if the time invested in supplying students with meaningful feedback makes much impact on students, William claims that "students who only received scores made no progress from one task to the next, while those students who received comments improved about 30 percent" (2011). Indeed, it seems commonsensical that learners would get more out

of feedback when told why something they have produced is incorrect, or when indications by their teacher would lead them to understand the reason for their error.

Like earlier research done on TGF, more recent research on AWE tools and the feedback they provide focuses on the way they highlight student errors and whether they supply direct or indirect feedback; Liao (2016) presents a useful checklist of questions that should be considered when analyzing CGF, the first of which focuses on the *type* of error that AWE tools tend to focus on. Again, like the feedback supplied by teachers (Hartshorn et al., 2010), studies looking at CGF show that the tendency of AWE tools is to focus attention on identifying grammatical errors (Crusan, 2015), or what researchers refer to as "formal attributes" (Li et al., 2014), "local" issues (Liao, 2016) or "surface features" (Chen & Cheng, 2008). This is likely the reason why a considerable amount of AWE-related research has looked at how–or if–CGF impacts students' grammatical accuracy, and how the tools can be used to help students and teachers identify typical grammatical errors (Liao, 2016; Dikli & Bleyle, 2014; Calaveri & Dianati, 2016; O'Neill & Russel, 2018). It also seems worth pointing out that a great deal of that grammar-focused research has focused primarily on one of the first popular AWE tools made available to educators, *Criterion*, which was launched by the American firm that is behind the TOEFL & TOEIC exams, Educational Testing Service (ETS) (Hassanzadeh & Fotoohnejad, 2021; Li et al., 2014; Dikli & Bleyle, 2014; Wilson & Andrada, 2016; Link et al., 2014); the AWE tool *My Access* appears to be another popular subject of research (Chen & Cheng, 2008; Mohsen & Alshahrani, 2019), while fewer studies have been carried out on *Grammarly* (O'Neill & Russel, 2018; Calaveri & Dianati, 2016; Fahmi & Cahayono, 2021)*, Write & Improve* (Karpova, 2020; Kostikova & Miasoiedova, 2019; Wali & Huijser, 2018) and *Chat GPT*– though that trend seems to be changing as the tools have been gaining notoriety in the last few years.

Even if CGF tends to fixate on grammatical issues in writing, the use of AWE tools in the classroom is often justified by researchers because they provide the kind of corrective feedback that may lead to students being more aware of their errors, which is often considered "an essential element in L2 acquisition" (Li et al., 2014). Indeed, Hassanzadeh and Fotoohnejad's recent study of university students in Iran linked their use of *Criterion* with the "triggering of noticing of errors" (2021, p. 1500), which demonstrated that the CGF supplied by the AWE tool over the course of a semester helped in improving students' writing scores. But their study is not alone in extolling the benefits of corrective feedback supplied by teachers and AWE tools alike.

Generally speaking, research has determined that the effect that AWE's corrective feedback makes on student writing and test scores is positive (Wali & Huijser, 2018; Fu et al., 2022; Karpova, 2020; Kostikova & Miasoiedova, 2019). One of the first studies at scale that looked at the impact that the corrective feedback supplied by *Criterion* had on thousands of students in the U.S. determined that the AWE tool was effective in helping students improve their writing quality (Attali, 2004). The findings from these studies, both large and small, have likely justified the continued use and study of AWE-generated WCF and its impact on student writing.

2.2. The potential and limitations of AWE tools

Most researchers seem to agree that AWE tools like *Criterion* and *Grammarly* have enormous potential as fast and effective feedback generators, even if they tend to identify more basic errors at word level. First and foremost, they save already-busy teachers precious time (Wilson & Adrada, 2016; Fu et al., 2022; Wali & Huiser, 2018), and allow students to receive personalized error correction with a click of a mouse, liberating teachers from the toils of being simply "proofreading slaves" (Laio, 2016). It is not surprising that some researchers see AWE tools as a welcome resource for teachers seeking "relief from the staggering amount of grading they do" (Crusan, 2015, p.21). However, there seems to be consensus among most proponents of AWE tools and the error correction they provide that they should be used *in conjunction with* teacher guidance and instructor feedback (Chen & Cheng, 2008)– *not as a replacement for* these vital parts of students' learning. Indeed, Fu et al.'s comprehensive review (2022) of 45 recent studies on AWE tools found that a mix of AWE feedback with TGF was the teaching strategy adopted most frequently; *Criterion* and *Grammarly*, advocates say, should be used as a kind of "learning facilitator" (Liao, 2016), an initial proofreader (Fahmi & Cahayono, 2021), or "surrogate writing coach" (Chen & Cheng, 2008). When used as a supplementary but important first part of a feedback loop, or in the early drafting of a written text (Li et al., 2014; Crusan, 2015), AWE can "heighten learner autonomy" and become an integral part of students' "toolbox for the writing process" (Link et al., 2014) and lead to improvements in students' writing (Hassanzadeh & Fotoohnejad, 2021). Wilson and Adrada (2016) go so far as to compare AWE tools with Zygotsky's concept of the More Knowledgeable Other (MKO), who engages with the learner and guides them through a process of educational development.

While some might argue that AWE tools can never replace the guidance of a human instructor, others have found additional reasons– beyond speed and its potential to become a kind of teaching assistant– to sing its praises. CGF, while not always as comprehensive as a human,

is, if nothing else, consistent in how it gives feedback, which most would argue is an essential part of effective WCF. Indeed, "humans are prone to foibles – they get tired, distracted, and hungry. These issues, proponents argue, interfere with objectivity" (Crusan, 2015, p. 24). Whereas teachers may forget, say, what color they usually use to highlight a preposition error, an AWE tool never will. In addition, the consensus among researchers and students alike is that AWE tools– especially *Grammarly*– provide explanatory feedback that is easy to understand (Fahmi & Cahayono, 2021; O'Neill & Russel, 2018; Calaveri & Dianati, 2016). Another important element to mention is that most AWE tools like *Grammarly* are now generally perceived to be easy to use (Calaveri & Dianati, 2016), thus "promoting learning through learner engagement" (Fu et al., 2022).

AWE tools are not without their detractors, and their limitations have been spelt out in several studies. One of educators' most common complaints is related to using tech resources like *Criterion* as a scoring tool in summative assessment, as too much emphasis seems to be put on the length of students' writing, grammar and vocabulary and formulaic answers (Chen & Cheng, 2008; Li et al., 2014) and not enough importance is given to content and creativity. This may be true, and those who have found too much attention is paid by AWE tools to grammatical errors are numerous (Liao, 2016; Dikli & Bleyle, 2014; Calaveri & Dianati, 2016; O'Neill & Russel, 2018), which is why many advocate for combining AWE tool feedback with instructor feedback, which could focus its attention on aspects like content, creativity and text organization. This feedback combination of grammar-focused error correction provided by AWE tools and instructor-generated formative assessment may help prevent the kind of "negative washback" among students that some researchers believe could come about from focusing too much on grammatical form (Li et al., 2014). In fact, it seems that many of the most common grievances that detractors of AWE tools have could be remedied by instructor guidance and explanation. If, as some researchers say, the CGF provided by AWE is difficult to understand or flawed (Fu et al., 2022; Liao, 2016; Grimes & Warschauer, 2010), then students must ask for help and instructors must step in wherever possible; this process could encourage heightened learner autonomy and is dealt with in more detail in Section 4.2.

One limitation of AWE tools, as pointed out by Liao (2016), is related to how comprehensive the CGF is-- that is, to what extent tools like *Grammarly*, *Write & Improve* and *Chat GPT* exhaustively identify all the errors a student makes. This issue is dealt with in more detail in Section 4.1, in response to one of the three research questions proposed.

2.3. Common uses of AWE tools in the classroom

One of the most common ways that teachers regard AWE tools is as teaching assistants or proofreaders. While some researchers only briefly mention how they promote the independent use of the tools in or outside of the classroom (Karpova, 2020), others go into varying degrees of detail as to how they integrate the tools into classroom writing instruction. Some researchers claim that further studies must be done on using AWE tools in the classroom (Stevenson, 2016). Although there is as of yet little published academic research on using *Chat GPT* in the classroom, experts have acknowledged its potential as a tool to "improve writing quality" ("Chat GPT and Artificial Intelligence in Higher Education: A Quick Start Guide," 2023). Most research has looked at how AWE-generated feedback can be used at early or regular intermittent stages of the writing process, thus resulting in the kind of feedback loop that seems to encourage more active uptake of error correction, learner independence and even higher writing scores.

In one study that analyzed *Write & Improve*, Kostikova & Miasoiedova (2019) used the tool to raise B2 students' awareness of the four evaluation criteria (Content, Communicative Achievement, Organization and Language) used by Cambridge Assessment to award marks on their writing exams. As the tool currently only generally gives indirect feedback on a variety of sub-skills, instructors in this study supplemented the *Write & Improve* output with a self-evaluation question list that encouraged students to notice aspects like following instructions, register, paragraphing and cohesive devices (Kostikova & Miasoiedova, 2019). Through this process of writing, engaging with the feedback and re-writing, the authors (2019) report seeing a 36% improvement in students' writing scores over the course of an academic year.

Liao's study (2016) on the AWE tool *Criterion* looked specifically at how it could be used to help improve students' knowledge and execution of grammatical aspects; while the study concluded that improvement was seen, Liao (2016) acknowledged that aspects like students' level, learning style, agency and metacognitive strategies also had an influence on how much students improved. The researcher also recognized that the writing feedback loop was complex– including planning, first drafts, reacting to initial AWE feedback, rewriting, and obtaining instructor feedback– and required time commitment from both students and teachers; in the study, students were asked to submit four essays based on *Criterion* rubrics and noticeable improvement was not observed until the third submission (Liao, 2016).

In a more recent study on using *Grammarly* in the writing classroom, Fahmi and Cahyono integrated the tool into their 7-step feedback loop as an initial in-class, supplementary proofreader (2021). Quite simply, students were given instructions on their writing task (a

"cause-and-effect paragraph") and then told to submit the paragraph to *Grammarly* for feedback. The third step involved students making corrections based on the feedback before submitting it to their instructor for further drafting (Fahmi & Cahyono, 2021). After four weeks of regularly using *Grammarly* in class, students were asked to complete a survey on the AWE tool and teacher feedback, with 100% of them responding positively to the statements "*Grammarly* has helped me to improve my writing" and "*Grammarly* was easy to use" (Fahmi & Cahyono, 2021).

2.4. Student perception of AWE tools

Students' perception of the AWE tools discussed in this section is generally positive, as the tools tend to be useful and easy to use, perhaps the two most important factors that influence uptake of new technologies (Calaveri & Dianati, 2016). For the most part, the CGF provided by these tools are considered easy to understand and/or helpful (Calaveri & Dianati, 2016; O'Neill & Russel, 2018; Wali & Huiser, 2018; Dikli & Bleyle, 2014) and in one case, even "magical" and "powerful" (Liao, 2016). According to researchers, students who find AWE tool-generated feedback difficult to comprehend tend to be those learners with lower levels of English (Liao, 2016), or from older studies (Grimes & Warshauer, 2010; Chen & Cheng, 2008), which may suggest that the CGF being provided to students has improved over time, or that students have become more accustomed to the feedback provided by AWE tools. Another factor that seems to influence students' opinion of these tools is teachers' opinions of them (Li et al., 2014); if instructors integrate the use of AWE tools into the university curriculum as trustworthy grammar checkers, students will likely also see them as such.

3. OBJECTIVES

Taking into account the issues related to AWE tools outlined above, the following research questions have been selected for this FMP:

1. What are the strengths and weaknesses of the three AWE tools (*Grammarly, Write & Improve, Chat GPT*) as feedback providers to first-and-second-year university students?
2. To what extent does using these AWE tools as feedback providers in the classroom lead to increased learner autonomy among first-and-second-year university students?
3. To what extent does using AWE tools as feedback providers in the first-and-second-year university classroom increase student uptake of feedback in general?

4. METHODOLOGY

In this section, the aim is to outline the process put forward to find the answers to the three research questions listed in 3. As my first research question addresses the strengths and weaknesses of the AWE tools themselves and the second two questions are concerned with how their use in the university classroom may affect learner autonomy and uptake of feedback, it has been deemed appropriate to divide this section into two parts, with Part 1 (4.1) addressing the first research question and serving as a kind of introduction to Part 2 (4.2).

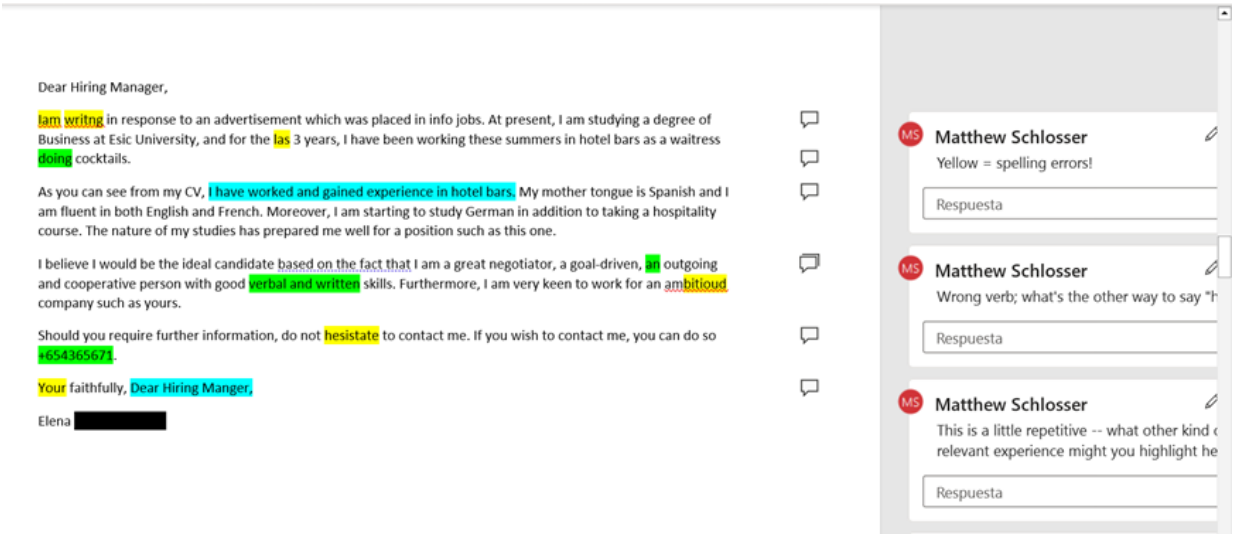
4.1. A description of the three AWE tools: a comparative analysis of the feedback provided by *Write & Improve*, *Grammarly* and *Chat GPT* (RQ1)

What the three AWE tools chosen for this project—*Grammarly*, *Write & Improve* and *Chat GPT*—have in common is that they are all readily accessible online for free while requiring users to register and log in for seamless usage; more importantly, all three provide, to varying degrees of detail, summative and formative feedback on submitted writing. However, the differences between the three are numerous, from how and to whom they are marketed, their popularity amongst students, teachers and the general public, to the extent to which they provide effective, detailed corrective feedback on users' writing. The latter is the focus of this section, which aims to highlight the strengths and weaknesses of each AWE tool as a useful feedback-provider for students as they proceed through the writing process.

In their comparative analysis of two AWE tools' feedback provided to student writing, Ranalli & Yamashita (2022) employ the criteria of *access* (the ease to which users can access the tool), *delivery mode* (synchronously and/or asynchronously), *analysis* (the techniques used to identify and/or correct errors) and *focus* (the type of error most commonly identified). For this comparative analysis, which will be limited to comparing the feedback given by the instructor (see Figure 4) and the three AWE tools, the latter two criteria will be taken primarily into account, as the former two are basically the same for all three: the tools are, as mentioned earlier, readily accessible for free and will be used in their asynchronous mode, albeit arguably so fast that they might be considered near-synchronous.

Figure 4

Teacher feedback on Student response to Write & Improve Letter of Application task



Note: Teacher's feedback is divided into color categories: blue for content, yellow for spelling, green for vocabulary. Suggestions and further explanations provided in the comment boxes to the right. In total, 12 errors were identified.

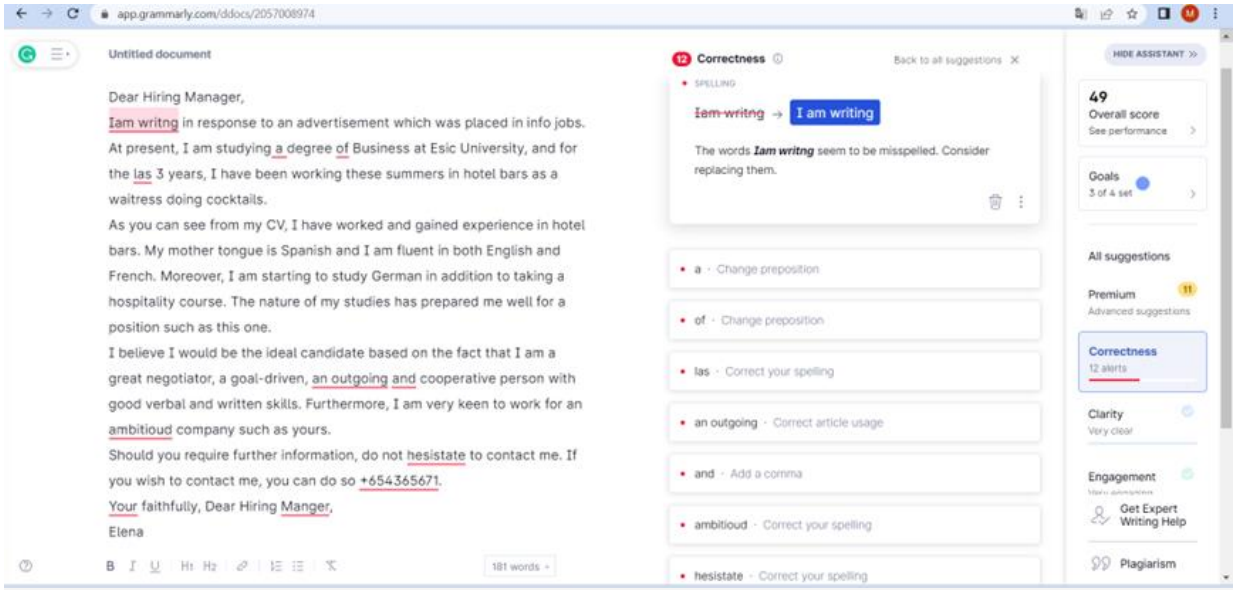
Grammarly

After accessing the *Grammarly* website, registered users looking for corrective feedback simply copy and paste their text into the text box. Within seconds, the website's Assistant tool generates a kind of instant report on "Correctness" (see Figure 5). In terms of drawing errors to the user's attention, *Grammarly* underlines what it considers incorrect in the text. For further explanation, users may hover over the error and obtain a correction with further explanation, which appears in a box to the right of the text. Alternatively, users may click on one of the boxes to the right of the text, which expands the box and displays the correction with a short explanation.

The focus of *Grammarly*'s more formative feedback is, primarily on grammar and spelling. The website's Assistant tool identified the same five spelling errors that the professor had in my feedback as well as an article error in the third paragraph and a missing preposition issue in the fourth paragraph. While the Assistant tool does not provide any feedback on content (which the professor did in two cases of repeated or inappropriate information), *Grammarly's* feedback did point out two other errors (both involving prepositions) that the instructor had simply missed in the second paragraph.

Figure 5

Grammarly feedback on Student's Letter of Application task

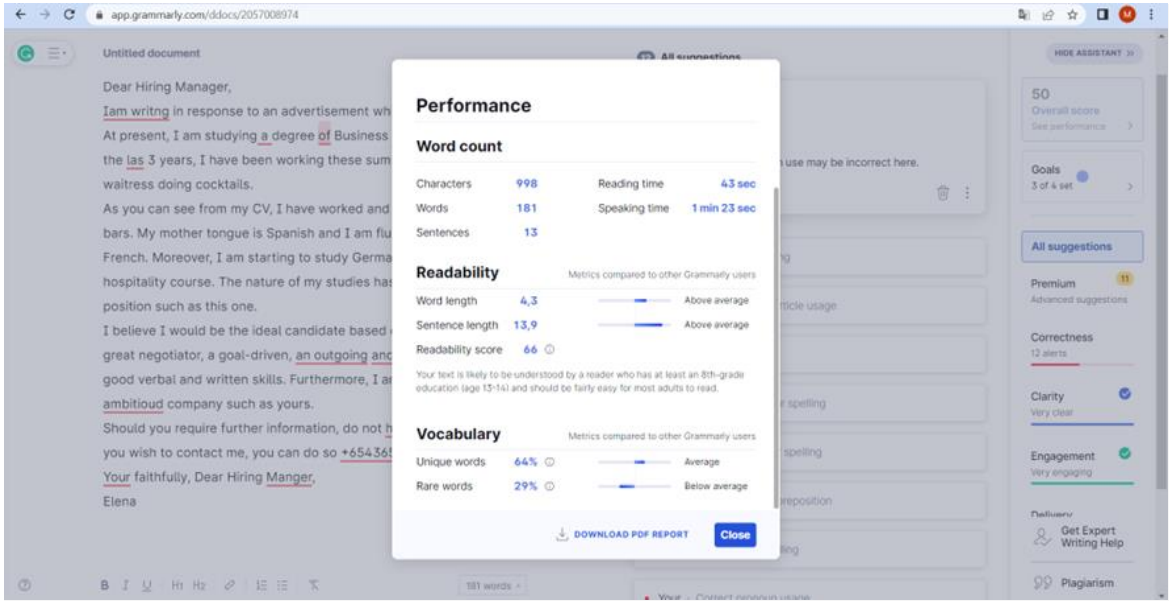


It may be important to point out that *Grammarly* is programmed in American English, so the feedback provided by the AWE tool is based on conventions and spellings found in the United States. An example of this is found in the feedback given by *Grammarly* on what it considers a missing comma (before an "and" at the end of a short list of adjectives); commonly referred to as an Oxford comma, users should know that it is "grammatically optional" despite the fact that *Grammarly* marks it as an error and seems to go against conventions accepted in British English and at the UOC itself (Universitat Oberta de Catalunya, n.d.; *What Is the Oxford Comma (or Serial Comma)?*, 2023).

Finally, it should be noted that *Grammarly*'s Assistant also awards users an overall score for the writing performance demonstrated in their texts, which is based on the following categories: "readability", which takes "word length" and "sentence length" into account, and vocabulary, which analyses how "unique" or "rare" the words in the text are (Grammarly: Free Writing AI Assistance, n.d.). A short report accompanying this score can be easily accessed by clicking on the number/overall score box in the top right corner of the screen (see Figures 5 and 6).

Figure 6

Grammarly's Writing Assistant's short report on user performance



Write & Improve

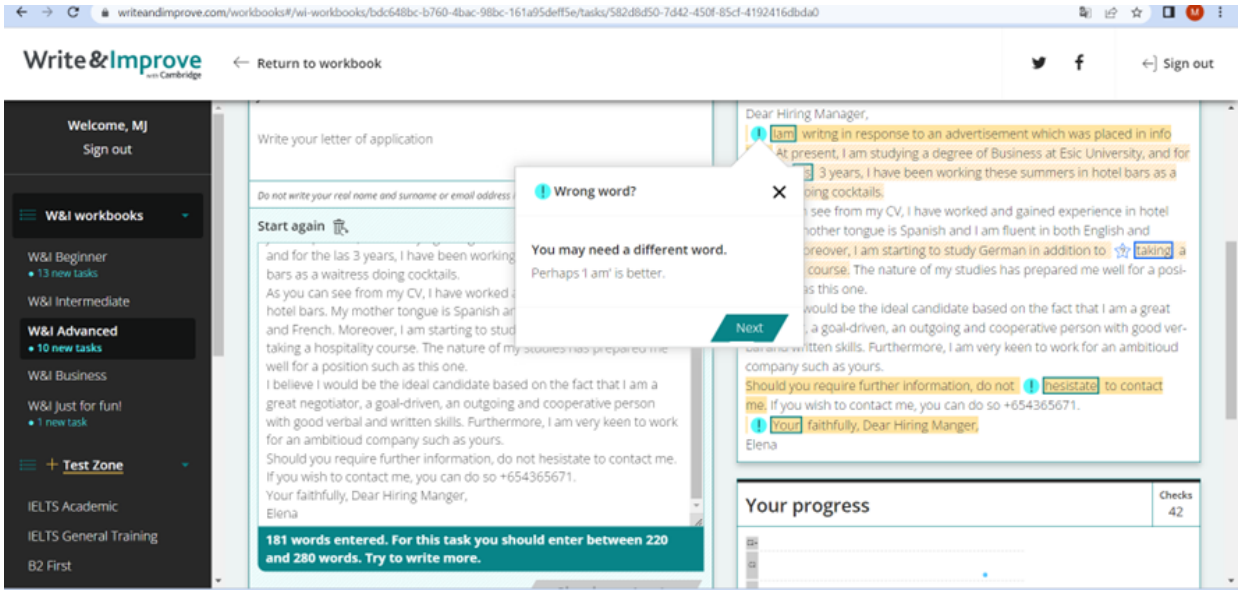
This website, affiliated with Cambridge Assessment, provides users with writing tasks typically found on official Cambridge Writing exams, so the target user is considerably more specific than *Grammarly,* which is used by native speakers as a proofreading tool. Like *Grammarly*, students submit their texts by typing directly into a text box and then clicking on the "Check" button at the bottom of the screen. Within seconds, users receive both summative and formative feedback on their writing; the former, in terms of a CEFR level, and the latter, in the form of highlighted sections of text, exclamation points and question marks (see Figure 7). As its name implies, *Write & Improve* has been designed to be used regularly, over time, to help users enhance their writing skills. In addition to providing users with an overall mark and a mark out of 5 for content, *Write & Improve* encourages users to continue working on their writing; this kind of positive feedback that feeds forward is not present in Grammarly's feedback reports.

However, when it comes to providing users with effective explanatory feedback on their errors, *Write & Improve* is not nearly as comprehensive or detailed as *Grammarly*. Whereas the latter alerted the user to 12 correctness issues, *Write & Improve* provides specific feedback in only five cases, four of which being spelling errors ("Iam", "las", "hesistate" and "Your faithfully"). Users may click on the exclamation points that appear in the text so as to obtain more information on the error: the AWE tool asks if the word is correct and makes a suggestion

for improvement (see Figure 7), ultimately allowing the user to decide if the change should be made.

Figure 7

Formative corrective feedback provided by Write & Improve



Out of the five errors that *Write & Improve* flagged, one appears to be what Ranalli & Yamashita would call a false positive (2022), as the word ("taking", preceded by "in addition to" and followed by "a hospitality course") has been identified by the AWE tool as "suspicious". It is not clear if the program has found an issue with the form of the verb (in this case, a gerund followed by "to"), or if the word take has been identified as part of an incorrect collocation, but as is, the phrase "in addition to taking a hospitality course" should be considered grammatically correct.

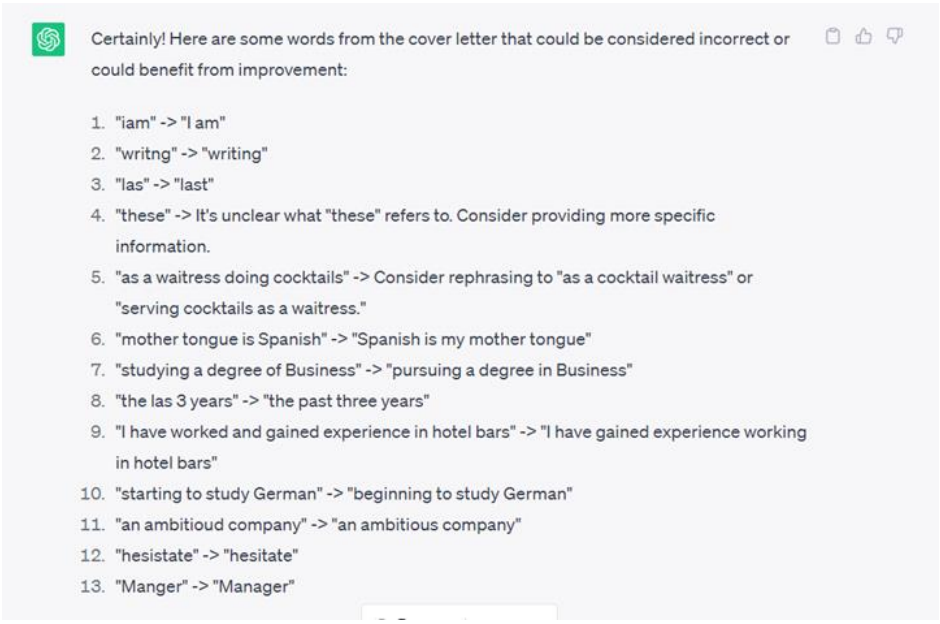
Chat GPT

Beyond a doubt, *Chat GPT* is the best-known of the three AWE tools selected for this project, likely since it can be used as a consulting tool on nearly any topic: from travel itineraries to cooking recipes, from real-time weather reports to relationship advice. The artificial intelligence powering this tool has been in development for years and offers users a conversational experience that traditional search engines do not (Goode, 2023). It is precisely because of this conversational element that users must determine effective ways to ask *Chat GPT* for the kind of effective written corrective feedback they seek.

After explicitly asking the tool to only "suggest and list changes" and submitting the text in question, *Chat GPT* responded—within seconds-- by making changes directly to the main text instead of listing them. Upon being requested to list the corrections one more time, the tool responded appropriately and provided the list found in Figure 8. In terms of error-flagging, this appears to be the most effective way to request and obtain clear identification of errors and suggested changes.

Figure 8

List of errors flagged and corrected by Chat GPT



The list of corrections includes 13 flagged errors, 7 of which can be categorized as spelling errors, all of which were also identified by *Grammarly*. However, five words and phrases that were flagged by *Chat GPT* were not identified by *Grammarly* as errors or in need of improvement. Of these five, three errors could very well be considered false negatives. For example, in correction 6, *Chat GPT* suggests changing "my mother tongue is Spanish" to "Spanish is my mother tongue", while correction 10 consists of changing "starting" to "beginning" before the phrase "to study German." Although the latter could be interpreted as a way of trying to provide the text with a more formal register, it can be argued that the use of "start" should not be considered an error as it does not impede understanding or come across as particularly informal. Still, the majority of the 13 errors flagged should not be considered false positives, as the tool does point out that its aim is to "correct potential errors or improve clarity" and that "some words […] could benefit from improvement" (Chat GPT, 2023).

4.2. Learner autonomy and student uptake of feedback provided by *Write & Improve, Grammarly* and *Chat GPT* (RQs 2 & 3)

4.2.1. Learning context and participants

The study will involve two groups of 1st and 2nd-year university students currently enrolled in a private university—specializing in the academic fields of Business, Marketing and Advertising—in the outskirts of Madrid, Spain. The mandatory subject and the corresponding course material being taught as well as the activities carried out in both groups are essentially the same, as all three groups are, on paper, at C1.2 level of English. Having said that, individual levels of English vary between a strong B2 and a low C2, though most would likely officially test at C1. Still, the two key Intended Learning Outcomes in both courses are the same: improve presentation/public speaking skills and hone formal writing skills.

In a 15-week semester-long course, students are usually assigned one in-class writing exam and one research-based essay, the former using a prompt similar to those used by Cambridge Assessment in its official English exams. Generally speaking, students' writing skills tend to be one of their weaker areas, which is one of the important justifications for this project.

4.2.2. Design

In Table 1, a timeline is proposed as to when and how the three AWE tools can be used with both groups of learners, as well as when and how students' perceptions of the AWE tools can be collected. Further details on the procedure can be found in 4.2.3.

Table 1

Methodological/pedagogical design of the FMP

Lesson	**When?**	**Objectives**	**How?**
Lesson 1: *Write & Improve*	The first class of the term (Week 1)	(1) To help determine students' writing level (2) To introduce students to *Write & Improve*	Allow students to engage with *W&I's* summative and formative feedback
Lesson 2: *Grammarly*	Immediately after the in-class writing test (Week 4 or 5)	(1) To introduce students to *Grammarly*	- Allow students to engage with *Grammarly's* CF

		(2) To encourage a feedback loop and students' uptake of feedback	- Give students the chance to turn in an improved version of writing for a higher mark
Lesson 3: *Chat GPT*	Immediately after returning students' 2nd writing assignment (Week 11 or 12)	(1) To demonstrate to students that *Chat GPT* can be used as an AWE tool (2) To encourage another feedback loop & uptake of feedback (3) To immediately obtain students' opinions of *Chat GPT*	- Allow students to engage with *Chat GPT* as an AWE tool - Ask students to turn in assignment reviewing *Chat GPT's* feedback - Ask students to participate in Canvas Discussion
End-of-term survey	Week 14 or 15 of 15-week term	Gather students' perceptions of three AWE tools used in class	Ask students to complete a 7-question survey on AWE tools and learner autonomy

4.2.3. Procedure

In this section, a step-by-step description of how the three AWE tools can be used over the course of a semester will be outlined. In addition, an explanation will be provided as to how students' perceptions of the resources in question can be surveyed as well as how student uptake of the feedback could possibly be measured.

Lesson 1: Write & Improve

Students are asked to write a letter of application through an assignment made available to them through the LMS Canvas. Students are told that the objective of the task is to properly gauge their level and find out what grammatical and lexical areas might need the most attention.

After the 30-minute time limit ends, students' attention is then drawn to *Write & Improve* and after being given a brief introduction to its purpose and target user, are asked to submit their writing to the website's level-checker. Students are encouraged to look at any highlighted

sections of their own text, as well as any exclamation and/or question marks flagging errors and are told to make any changes to the letter they find appropriate and asked to resubmit the task. To conclude this first-day activity, students are told to take a screenshot of their *Write & Improve* feedback (with corresponding CEFR level) and told to submit it along with the previously uploaded text. Informally, students are asked about their first impression of the website, and whether they found it useful and/or could imagine using it in the future.

Lesson 2: Introduction to Grammarly

In the lesson immediately after writing a motivational or cover letter for an in-class writing exam, students are given feedback on their errors: first, the instructor explains how errors are flagged and then students are given back their letters, which will have been written using pen and paper. Students are encouraged to look closely at the feedback given to them, to ask their instructor for further feedback or clarification where necessary, and to create an improved draft of their letters on a Word document, which they will be encouraged to resubmit for a higher mark. However, before allowing students to resubmit, their attention is drawn to *Grammarly* and they are shown how to use the AWE tool to receive feedback. Like in Lesson 1, students who resubmit their writing must also attach a screenshot of the *Grammarly* feedback. Informally, students could be asked what their impressions of the tool are and to keep using both tools to help improve their writing over the semester.

Lesson 3: Chat GPT as an AWE tool

After students' research-based essays have been marked and returned, a feedback lesson similar to the one outlined in the previous section is conducted. However, rather than asking students to submit their writing to *Grammarly,* students are asked to submit their essays to *Chat GPT*. Ideally, the instructor would have a sample student essay ready to use for class and illicit commands from students in order to obtain a list of flagged errors with corrections and suggestions for improvement (see Figure 8). Students are given time in class to experiment with *Chat GPT* as a feedback tool and are told to copy and paste the list of flagged errors generated by the program. Students are asked to analyze the list closely and evaluate the feedback provided and upload their ideas to a Canvas assignment. As an incentive, students are informed that their essay marks will improve by half a point (out of 10) if they complete the assignment.

4.2.4. Instruments used for data collection

In addition to the AWE tools themselves, the Learning Management System (LMS) Canvas and Google Forms can be used to determine the answers to research questions 2 and 3. Data can be collected in the following ways:

1) In the last week or two of the terms, students complete a 7-question survey administered through Google Forms to determine students' perception—both quantitatively and qualitatively-- of the AWE tools used in the classroom and if the tools helped them become more autonomous learners. See *Appendix A* or access the survey here: https://forms.gle/xusHB9SSmJSQzNVE9. The last three questions are designed to measure students' level of learner autonomy:

   Which of the tools do you think you're most likely to use in the future to help you improve your writing?

   To what extent did using these tools in class make you more aware of the grammatical and vocabulary errors you make?

   How likely do you think these tools could make you a more independent learner?

2) Before the end of the 3rd lesson outlined in 4.2.4, students are also asked to participate in an informal, qualitative LMS Canvas Discussion that asks their opinion of the feedback supplied by *Chat GPT* (see Appendix B).

3) In this project, students' uptake of the feedback provided to them by the AWE tools will be measured by the extent to which students take advantage of opportunities to revise and resubmit their writing and/or submit the *Chat GPT* assignment discussed in 4.2.4. Both assignments, administered through the LMS Canvas, are optional but award improved marks on the two writing assignments which precede each.

5. RESULTS / EVALUATION OF THE INTERVENTION

In this section, the research questions proposed for this TFM will be answered based on the data gathered, as outlined in the Methodology section.

5.1. Strengths and weaknesses of each AWE tool used with students (Q1)

A summary of the strengths and weaknesses of all three CGF as well as TGF on the same student submission can be found in the table below (see Figure 10).

While Grammarly's Assistant tool does not provide any feedback on content, it did point out two other errors (both involving prepositions) that the instructor had simply missed in the second paragraph. This would seem to align with some research that suggests AWE tools might provide better feedback than a teacher (Hassanzadeh & Fotoohnejad, 2021; Crusan, 2015). Generally speaking, Grammarly rates high amongst researchers in terms of identifying errors that are indeed errors, and "false positives"—errors flagged by an AWE tool that aren't actually errors—appear to be relatively uncommon (Ranalli & Yamashita, 2022). After analyzing the feedback provided by Grammarly, the instructor involved in this study would concur; Grammarly's feedback is user-friendly and offers at least some explanation to users.

While Chat GPT actually flagged more errors than the other two AWE tools, three of them could be considered false positives, which students must be made aware of. Another factor that could be considered a weakness of Chat GPT is that an errors list must be requested, whereas with the other two AWE tools formative feedback is provided automatically.

Table 2

Summary of features of TGF and AWE tools

	TGF	Grammarly	Write & Improve	Chat GPT
Total errors flagged	12	12	5	13
Flagged errors related to grammar, vocabulary or spelling	10	12	5	13
Feedback comments on content	2	0	0	0
False positives	0	0	1	3
How errors flagged	Color-coded	Underlined	Highlighted	Listed, *when commanded to do so*
Summative feedback	A score out of 10	"Readability Score" out of 100	CEFR level	CEFR level, *when requested*
Time taken to generate feedback	8 minutes	Less than 10 seconds	Less than 10 seconds	After requesting twice, less than 10 seconds

When students were asked specifically about the perceived strengths and weaknesses of each tool, students made a wide range of observations. Regarding Write & Improve, one respondent found it particularly useful that the AWE tool provided a great way to "practice all

kinds of essays as well as an approximate English level, but also perceived the feedback given as superficial: "It doesn't correct it for you."

As for Grammarly, the perceived strengths included quickness and ease of accessing the feedback, while the weaknesses brought up by respondents were that the feedback might not provide a proper explanation as to why something was wrong; one student found that Grammarly's feedback, in its free version, was as limited as the automatic feedback provided in a Word document; indeed, two of the four respondents believed Grammarly's feedback lacked depth.

Finally, the perceptions of Chat GPT as a feedback tool were generally the most positive of the three; one respondent said it provided the best feedback of the three, another highlighted the variety of its feedback responses and another pointed out the speed with which it provided the feedback. Students' contributions to the Canvas discussion about Chat GPT were also mostly positive; adjectives used to describe the feedback included "accurate, useful, detailed, structured and insightful." Two students pointed out that the tool helped them make their writing sound more formal.

5.2. Learner autonomy (Q2)

Unfortunately, only four students from Group 2 took part in the Feedback Tools survey, so only very limited conclusions can be made. When asked to select one AWE tool that they would most likely use in the future, three of the four respondents chose Grammarly, which would suggest that they might independently seek out the tool in the future.

Still, the quantitative data that supports the idea that the three AWE tools discussed make learners more independent is limited at best. While it is true that the four students who took part in the survey did believe that it was "highly likely" or "likely" that they would use the tools independently in the future, those results cannot be considered representative of all students. Having said that, personal observation in the classroom does demonstrate that at least half of the students in Group 2 either downloaded the Grammarly plug-in onto their personal computers or informed me that they did use it as an editing tool for their assignments.

Perhaps one student summed up his future intentions best when responding to the end-of-term survey:

"I would use both, use Grammarly first, and then put it on ChatGPT. Chat GPT did not correct the grammar that Grammarly did but it did help with the change of wording to make it sound more professional."

It should be pointed out that the same student also believed that neither AWE tool in question could replace the feedback provided by a teacher; in fact, he went on to say, "It would be interesting to have a mixture of the 3."

5.3. Student uptake of feedback (Q3)

Students' uptake of the feedback provided to them by the AWE tools was measured by the extent to which they took advantage of opportunities to submit two assignments discussed in 4.2. Due to logistical constraints, while both groups were given the chance to resubmit their first writing assignment, only Group 1 could submit the second assignment evaluating the feedback provided by Chat GPT.

Of the 15 students from Group 2 that took the in-class writing exam and were allowed to resubmit, 13 took advantage of the opportunity, which is the equivalent of nearly 87%. In Group 1, of the 27 students that were eligible to resubmit their in-class writing text, 16 actually did so, which is approximately 59%, considerably lower than the first group.

Uptake among students in Group 1 was also low when it came to analyzing the feedback given to them by Chat GPT in order to improve their second writing mark. Of the 18 students that could submit the assignment for a higher mark, only 3 (16.7%) did so.

6. CONCLUSIONS

6.1. Discussion

Perhaps surprisingly, few comparative analyses of the feedback supplied by different AWE tools have been carried out, and on the three chosen for this project, none have been found to date (15 June). One recent study (Ranalli & Yamashita, 2022) contrasting the feedback generated by *Grammarly* and *Microsoft Word's* editing tool found that the former flagged significantly more errors than the latter. Indeed, when weighing the strengths and weaknesses of the feedback provided by the three AWE tools in this study, *Grammarly* comes out on top for its user-friendly website, explanatory feedback and lack of false positives when it comes to flagged errors. Having said that, both *Write & Improve* and *Chat GPT* can be used by language learners and teachers in an EFL classroom as a way of promoting learner autonomy and uptake of feedback, though neither was clearly demonstrated to do so in this particular study. *Write & Improve's* strength lies in its seemingly accurate ability to gauge a student's level on the CEFR scale, while *Chat GPT* stands out for its capacity to make suggestions regarding register and organization, which is something the other two tools are much less apt to do.

Promoting the use of all three tools may be a way to help teachers save some time when correcting students' writing, may help students self-correct their more basic errors and may also even lower students' affective filters by not being subjected to teachers' "red pen" (Semke, 1984). Of course, it must be acknowledged that encouraging the use of *Chat GPT* in the classroom may trigger its fair share of objections from fellow educators, as it might be seen as a way of promoting the use of the tool to generate texts rather than receiving feedback on them.

6.2. Limitations

The limitations involved in this TFM are numerous.

Firstly, in the comparative analysis of the three AWE tools, only one student submission was used to contrast the feedback generated by *Grammarly*, *Write & Improve* and *Chat GPT*. To draw sounder, data-driven conclusions about the three tools, surely more samples must be analyzed. In addition, other instructors as well as a broader pool of students could also be surveyed to help determine where each website's strengths and weaknesses lie. Another important factor to be taken into account is the fact that AI is evolving by the minute; what is true about these tools today could very well seem like ancient history in a month or two. New products, versions and updates are being released every day, and keeping up with all their potential uses in the classroom is particularly challenging for educators today.

In terms of measuring learner autonomy and student uptake of feedback, this study is similarly marred by its limited number of student participants and survey respondents. Indeed, with only 7 students participating in the Canvas Discussion and 4 students responding to the survey on all three AWE tools, the data collected for this study is extremely limited and cannot be considered representative.

6.3. Recommendations

Dismissing the use of AI-driven tools as simply cheating machines for students is, without a doubt, simply reductive; many students even believe that using Chat GPT for assignments is a form of cheating (Nietzel, 2023). But AI's potential in the language classroom—especially when seen as an MKO, a proofreader, a feedback generator-- is too great to be ignored, feared or regarded as cheating. When asked about how it could be used by language instructors and students in the classroom, *Chat GPT* (15 June, 2023) responded:

> Teachers can integrate an AI language model like me into language learning platforms or tools used in the classroom. Students can access me for instant translation, vocabulary expansion, or language practice exercises. Teachers can incorporate interactive lessons and activities using the AI's capabilities to enhance language learning in an engaging and interactive manner.

Indeed, educators must learn how to tap into the potential of AI-driven tools like the ones discussed in this study and promote their use as supporting learning. Of course, instructors must proceed cautiously, but more exhaustive, more representative action research must be done on how to help language learners use these tools to become better, more autonomous writers of English. Is having students take note of the CGF provided to them by these tools the way to make students more aware of their errors? Will CGF help make students more appreciative of TGF, and help them become more feedback literate? Will encouraging the use of AWE tools like *Grammarly* be an effective way of making students more independent learners? Ideally, these are questions that educators should take the time to consider, research, to ask themselves and their students in order to be better prepared for what may be a very different-looking future.

REFERENCES

About Us | Grammarly. (n.d.). https://www.grammarly.com/about

American English [American English Live]. (2020, March 21). AE Live 9.3 - Strategic Corrective Feedback in the EFL Classroom [Video]. YouTube. https://www.youtube.com/watch?v=M25Wc57Zd7M

Attali, Y. (2004). *Exploring the feedback and revision features of Criterion*. Paper presented at the National Council on Measurement in Education (NCME), San Diego, CA.

Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education 5, no.1: 7-74.*

Carless, D. & Boud, D. (2018) The development of student feedback literacy: enabling uptake of feedback, Assessment & Evaluation in Higher Education, 43:8, 1315-1325, DOI: 10.1080/02602938.2018.1463354

Caudery, T. (June 1995). What the "process approach" means to practising teachers of second language writing skills. TESL-EJ, 1(4), 1-17. https://tesl-ej.org/~teslejor/ej04/a3.html

Cavaleri, M. R. & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning*, *10*(1). http://journal.aall.org.au/index.php/jall/article/viewFile/393/246

ChatGPT. (2023, March 30). Conversation with user. [Chat transcript]. Retrieved from https://chat.openai.com/chat/ec1be910-a859-4ca4-bec2-350b89afea97

ChatGPT. (2023, May 23). Conversation with user. [Chat transcript]. Retrieved from https://chat.openai.com/c/d1453269-40de-401e-bacc-bacc3c379034

ChatGPT. (2023, June 15). Conversation with user. [Chat transcript]. Retrieved from https://chat.openai.com/share/b817e4cd-6c1f-4407-a9a7-8abbbfeeb64d

Chat GPT and Artificial Intelligence in higher education: A quick start guide. (2023). In *UNESCO International Institute for Higher Education in Latin America and the Caribbean (IESALC)*. UNESCO. Retrieved April 24, 2023, from https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf

Chen, C.E. & Cheng, W.E (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.

Coffin, C., Curry, M.J., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2003). Teaching Academic Writing. Routledge (1st edition), Routledge.

Crusan, D. (2015). And then a miracle occurs: The use of computers to assess student writing. *International Journal of TESOL and Learning*, *4*(1), 20-32.

Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C. & Bivens-Tatum, J. (2008). Cognitive Models of Writing: Writing Proficiency as a Complex Integrated Skill. ETS Research Report Series. 10.1002/j.2333- 8504.2008.tb02141.x.

Developing an ELT product based on machine learning: Write & Improve. (2016, March 6). *Learn Jam*. Retrieved April 5, 2023, from https://learnjam.com/developing-an-elt-product-based-on-machine-learning-write-improve/

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22(1), 1–17.

Fahmi, M.A. and Cahyono, B.Y. (2021). EFL students' perception on the use of Grammarly and teacher feedback. *JEES (Journal of English Educators Society), 6*(1), 18-25

Ferris, D.R. (2010). Second language writing research and written corrective feedback in SLA. Studies in Second Language Acquisition, 32, 181– 201. doi:10.1017/S0272263109990490

Fu, Q., Zou, D., Xie, H., & Cheng, G. J. (2022). A review of AWE feedback: types, learning outcomes, and implications. *Computer Assisted Language Learning*, 1–43. https://doi.org/10.1080/09588221.2022.2033787

Grimes, D. & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *Journal of Technology, Learning, and Assessment*, *8*(6).

Harper, F., Green, H., and Fernandez-Toro, M. (2018). Using screencasts in the teaching of modern languages: investigating the use of Jing® in feedback on written assignments. *The Language Learning Journal*. 46(3), 277-292.

Harrison, L. (n.d.). Developing an ELT product based on machine learning: Write & Improve. *LearnJam*. Retrieved February 25, 2023, from https://learnjam.com/developing-an-elt-product-based-on-machine-learning-write-improve/

Hartshorn, K.J., Evans, N.W., Merrill, P.F., Sudweeks, R.R., Strong-Krause, D. & Anderson, N.J. (2010), Effects of Dynamic Corrective Feedback on ESL Writing Accuracy. *TESOL Quarterly*, 44: 84-109. https://doi.org/10.5054/tq.2010.213781

Hassanzadeh, M., & Fotoohnejad, S. (2021). Implementing an automated feedback program for a foreign language writing course: A learner-centric study. *Journal of Computer Assisted Learning*, 37(5), 1494– 1507. https://doi.org/10.1111/jcal.12587

Hattie, J.A. (1987). Identifying the salient facets of a model of student learning: a synthesis of meta-analyses. International Journal of Educational Research, 11: 187-212.

Karpova, K. (2020). Integration of "Write and Improve" AWE tool into EFLT at Higher Educational Establishment: Case Study *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics*, *7*(2), 137-150. https://doi.org/10.22219/celtic.v7i2.14036

Kloss, S. & Quintanilla, A. (2022). The Effectiveness of Direct and Metalinguistic Written Corrective Feedback to Deal with Errors in the Use of Information-Structuring Connectors. *Journal of Language and Education*, *8*(4), 88–101. https://doi.org/10.17323/jle.2022.15906

Kostikova, I., & Miasoiedova, S. (2019). Supporting post-graduate students writing skills development with the online machine learning tool: Write & Improve. *Information Technologies and Learning Tools*, *74*(6), 238–249. https://doi.org/10.33407/itlt.v74i6.2600

Lee, C (2020, February 20). The importance of feedback loops within the writing process. Turnitin. https://www.turnitin.com/blog/the-importance-of-feedback-loops-within-the-writing-process

Li, J., Link, S., & Hegelheimer, V. (2014). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18.

Liao, H. C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62, 77–92.

Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards Best ESL Practices for Implementing Automated Writing Evaluation. *CALICO Journal*, *31*(3), 323–344.

Mackey, A. (2014, June 25). Wanting it enough: why motivation is the key to language learning. *The Guardian*. https://www.theguardian.com/education/2014/jun/26/motivation-key-language-learning

Marulanda Ángel, N. and Martínez García, J. (January 2017). Improving English Language Learners' Academic Writing: A Multi-Strategy Approach to a Multi-Dimensional Challenge. GIST - Education and Learning Research Journal, 14, 49-67, https://doi.org/10.26817/16925777.367

Mao, S.S. & Crosthwaite, P. (2019). Investigating written corrective feedback: (Mis)alignment of teachers' beliefs and practice. Journal of Second Language Writing, 45, 46-60, https://doi.org/10.1016/j.jslw.2019.05.004

Mehrotra, P. (2021). Improve Student Motivation in High-Stakes Environments. *edCircuit | Improve Student Motivation in High-Stakes Environments*. https://edcircuit.com/improve-student-motivation-in-high-stakes-environments/

Mohsen, M. & Alshahrani, A. (2019). The effectiveness of using a hybrid mode of automated writing evaluation system on Efl students' writing. *Teaching English with Technology*, *19*(1), 118-131.

Moser, A. (2020). Written corrective feedback: the role of learner engagement: a practical approach (1st ed.). Springer. https://doi.org/10.1007/978-3-030-63994-5

Nietzel, M. T. (2023, March 20). More Than Half Of College Students Believe Using ChatGPT To Complete Assignments Is Cheating. *Forbes*. https://www.forbes.com/sites/michaeltnietzel/2023/03/20/more-than-half-of-college-students-believe-using-chatgpt-to-complete-assignments-is-cheating/

O'Neill, R. & Russell, A. (2018). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, *35*. 10.14742/ajet.3795.

Polio, C. & Park, J (2016). Language development in 2nd language writing. In R.M. Manchón & P.K. Matsuda (Eds.), Handbook of second and foreign language writing (pp. 287-306). De Gruyter. https://doi.org/10.1515/9781614511335

Semke, H. D. (1984). "Effects of the red pen", *Foreign Language Annals* 17(3), 195-202.

Stevenson, M. (2016). A Critical Interpretative Synthesis: The Integration of Automated Writing Evaluation into Classroom Writing Instruction. *Computers and Composition*, *42*, 1–16. https://doi.org/10.1016/j.compcom.2016.05.001

Torrance, H. & Pryor, J. (1998). Investigating Formative Assessment: Teaching, learning and assessment in the classroom. Open University Press.

Universitat Oberta de Catalunya. (n.d.). *Differences between British and American English - UOC*. Universitat Oberta De Catalunya - UOC Language and Style. Retrieved May 24, 2023, from https://www.uoc.edu/portal/en/servei-linguistic/convencions/british-american-english/punctuation/index.html

Wali, F. and Huijser, H. (2018), "*Write to improve*: exploring the impact of an automated feedback tool on Bahraini learners of English", *Learning and Teaching in Higher Education: Gulf Perspectives*, Vol. 15 No. 1, pp. 14-34. https://doi.org/10.18538/lthe.v15.n1.293

What Is the Oxford Comma (or Serial Comma)? (2023, April 19). Grammarly. https://www.grammarly.com/blog/what-is-the-oxford-comma-and-why-do-people-care-so-much-about-it/

Wiliam, D. (2011). *Embedded Formative Assessment*. Solution Tree.

Wilson, J., & Andrada, G. N. (2016). Using automated feedback to improve writing quality: opportunities and challenges. In Y. Rosen, S. Ferrara & M. Mosharraf (Eds), *Handbook of research on technology tools for real-world skill development* (pp. 678-703). IGI Global: US. doi: 10.4018/978-1-4666-9441-5.ch026

APPENDICES

Appendix A

Seven-question survey on students' perceptions of the AWE tools used in class and learner autonomy

1 Which online feedback tool do you consider the most effective?

Write & Improve

Grammarly

Chat GPT

2 What do you consider the strengths and weaknesses of Write & Improve?

(Open text box)

3 What do you consider the strengths and weaknesses of Grammarly?

(Open text box)

4 What do you consider the strengths and weaknesses of Chat GPT?

(Open text box)

5 Which of the tools do you think you're most likely to use in the future to help you improve your writing?

Write & Improve

Grammarly

Chat GPT

6 To what extent did using these tools in class make you more aware of the grammatical and vocabulary errors you make?

I'm much more aware of my mistakes now than before.

I'm a little more aware of my mistakes now than before.

I'm about just as aware as before.

7 How likely do you think these tools could make you a more independent learner?

Very likely

Likely

Not so likely

Very unlikely

I already consider myself an independent learner.

Appendix 2

Informal, qualitative Canvas discussion on Chat GPT

Give me your opinion on the feedback given to you by Chat GPT.

To what extent can you learn from it?

How useful was the feedback provided?

Have you got any misgivings about using it?