



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**EVALUATING THE USE OF AUTOMATED WRITING EVALUATION PROGRAMS  
AS TOOLS FOR FORMATIVE FEEDBACK IN ENGLISH AS A SECOND LANGUAGE  
POSTGRADUATE STUDENTS**

*Ana Isabel Hibert Santana*



**THE UNIVERSITY  
*of* EDINBURGH**

Doctor of Philosophy

Moray House School of Education and Sports

University of Edinburgh

2021

Ana Isabel Hibert Santana

*Evaluating The Use of Automated Writing Evaluation Programs as Tools For Formative Feedback In English As A Second Language Postgraduate Students*

Doctor of Philosophy, 2021

Supervisors:

Professor Dragan Gašević, Ph.D., *Monash University, Australia.*

Dr. Seongsook Choi, Ph.D., *University of Edinburgh, United Kingdom*

EXAMINATION COMMITTEE:

Kenneth Fordyce, Ph.D., *University of Edinburgh, United Kingdom.*

Christotea Herodotou, Ph.D., *Open University, United Kingdom*

# Abstract

Automated Writing Evaluation (AWE) programs use natural language processing techniques to analyse texts and score them according to pre-defined parameters. Most of these programs also offer feedback functions, providing students with feedback on various aspects of writing. These programs have been increasingly used in English as a second language (ESL) and English as a foreign language (EFL) classrooms to provide formative feedback to students, especially in the context of academic writing. However, existing research into the use of AWE programs has focused more on the end product of revision with AWE programs and whether the use of these can provide quantifiable gains in error reduction and holistic scores. Little research has investigated how these programs can be integrated into the writing process or what pedagogical approaches result in these technologies being incorporated into the classroom in ways that help the students develop their writing skills. This results in two major gaps in current literature on the use of AWE programs: 1) there is little information regarding how students engage with the feedback they receive and how they decide which feedback to use, and 2) scores and error rates can only give a superficial, post-hoc understanding of the effects of AWE in revision, but tells little about the depth of the changes made. The research showcased in this thesis seeks to address these gaps by presenting the results of two studies designed to improve our knowledge of how students use AWE programs. In the first study, screen captures and think-aloud protocols were used to record the interactions of 11 ESL postgraduate students with an AWE program during four revision sessions. The recordings were analysed to identify self-regulation strategies and decision-making strategies used by students when engaging with the AWE feedback. In the second study, a web program was created to collect texts before and after the use of an AWE program for revision, collecting a total of 30 texts suitable for analysis. The texts were compared before and after revision with an AWE program to understand the extent of the changes made during revision by analysing uptake rates and linguistic markers of proficiency to quantify the effects AWE feedback had on the revision of these texts. Results from

both studies suggest students are selective in their use of AWE feedback, drawing from a variety of sources of previous knowledge about English grammar and academic writing conventions to decide whether to accept or reject AWE feedback and this selectiveness results in low feedback uptake and little changes to the texts. These sources include feedback from teachers and mentors, previous exposure to academic texts and knowledge gained from previous English and composition classes. Successful integration of AWE programs into ESL/EFL classrooms should therefore take into account how students engage with the AWE feedback and use that knowledge in pedagogical strategies that scaffold student use of AWE programs and help them develop the cognitive and metacognitive skills need to successfully navigate the feedback they receive.

## Lay Summary

Automated Writing Evaluation (AWE) programs are software programs that promise to help students revise their written texts in English by providing feedback on grammar, spelling, style, content, and organisation. Commercially available AWE programs include, among others, Grammarly, ProWritingAid, Criterion, MYAccess! and Pigai. These programs have been considered as especially useful for students whose first language is not English and who are studying at English-medium universities or are studying academic English at their university, as they can provide students with immediate and continuous feedback on their written texts to help them improve their academic writing skills. However, there is very little information about how AWE programs should be introduced into classrooms to help students develop their writing skills. This thesis uses novel data collection and analysis methods to help understand what strategies students adopt when using AWE programs to revise their texts. I use this data to propose research-informed ways in which AWE programs can be integrated into classrooms to foster the development of writing skills in university students whose first language is not English.

## Declaration of Authorship

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. Parts of this work presented in the findings and discussion chapters have been published in the following journals:

- Hibert, A. I. (2019). Systematic Literature Review of Automated Writing Evaluation as a Formative Learning Tool. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming Learning with Meaningful Technologies .EC-TEL 2019. Lecture Notes in Computer Science* (Vol. 11722, pp. 119–212). Springer International Publishing.  
<https://doi.org/10.1007/978-3-030-29736-7>

## Acknowledgments

A PhD is a journey, and like any good journey, cannot be undertaken alone. There are many people I would thank for helping make this PhD thesis a reality.

First of all, I would like to thank my supervisors, Seongsook and Dragan. There were times when I wondered if I would be able to finish my PhD at all, but your constant support and unwavering faith in my abilities helped me make it through the many obstacles and hurdles I faced on this journey. It wasn't easy, but here we are!

I would also like to thank my family, and especially my parents. They have both been a massive source of inspiration and the anchor that keeps me steady regardless of what challenges I might face. I only have the courage to do the crazy things I do because I know that, whatever happens, you always have my back.

Finally, I would like to thank all my friends both in Scotland and Mexico. You listened to my complaints, commiserated with my hardships, helped me with my work, and bought me pints when I was about to lose my mind in a sea of data and rewrites. Without you to keep me sane, this thesis would probably never had happened.



# Contents

Abstract.....	i
Lay Summary.....	iii
Declaration of Authorship.....	iv
Acknowledgments.....	v
Contents.....	vi
Summary of figures.....	xi
Summary of Tables.....	xiii
Chapter 1 - Introduction .....	1
1.1 Background .....	3
1.1.1 Controversies surrounding the use of AWE programs .....	4
1.1.2 The usefulness of AWE programs .....	5
1.2 Thesis outline .....	7
Chapter 2 – Literature Review .....	11
2.1 Introduction .....	11
2.2 Publication: Systematic Literature Review of Automated Writing Evaluation as a Formative Learning Tool.....	12
2.3 New Directions on AWE Research .....	14
2.3.1 Behavioural engagement with AWE feedback.....	14
2.3.2 Cognitive engagement with AWE feedback.....	20
2.3.3 Affective engagement with AWE feedback .....	21

2.3.4 The role of teachers .....	26
2.3.5 Summary .....	30
2.4 Theoretical constructs .....	31
2.4.1 Feedback, learning, and the acquisition of writing skills .....	32
2.4.2 Self-regulation, engagement, and technology adoption .....	39
2.4.3 Linguistic competence and natural language processing .....	46
2.4.4 Process and product approaches to teaching writing .....	55
2.5 Conclusion.....	59
Chapter 3 – Methodology .....	61
3.1 Research questions and research objectives.....	61
3.1.1 Epistemological foundations of this research.....	64
3.2 The AWE program used .....	66
3.3 RQ1 – The effects of AWE feedback on student texts.....	77
3.3.1 Data collection .....	79
3.3.2 Data analysis .....	88
3.4 RQ2 – Student engagement and decision-making.....	93
3.4.1 Data collection .....	95
3.4.2 Data analysis .....	99
3.5 Validity and reliability .....	109
3.5.1 Validity .....	109
3.5.2 Reliability.....	112
3.6 Conclusion.....	112

Chapter 4 – The effects of AWE feedback on student texts.....	114
4.1 Introduction .....	114
4.2 Results.....	115
4.2.1 Student use of the AWE program .....	115
4.2.2 Nature of text changes.....	117
4.2.3 Quality of text changes .....	120
4.3 Discussion.....	122
4.4 Conclusions .....	125
Chapter 5 – Student engagement and decision-making.....	127
Chapter 5 – Student engagement and decision-making.....	<b>Error! Bookmark not defined.</b>
5.1 Introduction .....	127
5.2 Results.....	128
5.2.1 Self-regulatory strategies.....	128
5.2.2 Decision-making processes .....	139
5.2.3 Quality of uptake.....	145
5.2.4 The use of the AWE program.....	150
5.3 Discussion.....	158
5.4 Conclusion.....	162
Chapter 6 - Discussion.....	164
6.1 The process and product of AWE interaction.....	164
6.1.1 The product of revision with AWE feedback .....	165
6.1.2 The process of revision with AWE feedback.....	169

6.2 Potential of AWE as a tool for formative feedback .....	173
6.2.1 Multiple drafting .....	174
6.2.2 Noticing .....	176
6.3 Self-regulation in the use of AWE programs.....	178
6.3.1 Metacognition and cognition in the use of AWE .....	179
6.3.2 Self-regulation and technology adoption .....	182
6.4 Importance of pedagogical approaches in implementing AWE in the classroom .....	185
6.4.1 Using AWE as a tool for providing formative feedback .....	186
6.4.2 Student engagement with AWE feedback .....	189
6.5 Contributions .....	194
6.5.1 Methodological contributions .....	194
6.5.2. Theoretical contributions.....	196
6.6 Conclusion.....	198
Chapter 7 – Conclusion .....	199
7.1 – Overall conclusions .....	199
7.2 – Implementation of AWE as tools for formative feedback .....	201
7.2.1 Using AWE programs alongside other forms of feedback.....	201
7.2.2 Using AWE programs to set writing and revision goals .....	202
7.2.3 Using AWE programs to apply metalinguistic knowledge .....	203
7.2.4 Teacher training .....	205
7.3 - Limitations .....	206
7.3.1 Methodological limitations .....	207

7.3.2 Theoretical limitations .....	208
7.4 – Implications for future research .....	209
7.5 – Final thoughts.....	212
Reference List.....	214
Appendixes.....	241
Appendix 1 – Slides used during the academic writing workshop .....	241
Week 1 .....	241
Week 2 .....	252
Week 3 .....	262
Week 4 .....	270
Week 5 .....	276
Appendix 2 – Info sheet and consent form for Study 1 .....	287
Appendix 3 – Code for cleaning up the text database for Study 1 .....	292
Appendix 4 – Invitation letter, info sheet and consent form for Study 2.....	305
Invitation letter .....	305
Information sheet and consent form.....	306
Appendix 5 – Semi-structured interview questions.....	311
Appendix 6 – Sample transcript for study 2.....	312
Appendix 7 – Code for cleaning up and analysing student actions .....	317

## Summary of figures

Figure 1 – Information about the study in the registration page .....	70
Figure 2 – Registration form .....	71
Figure 3 – Dashboard .....	72
Figure 4 – Grammaraid Text Editor interface .....	73
Figure 5 – Feedback summary page .....	73
Figure 6 – Desktop version interface .....	74
Figure 7 – Summary section, desktop version .....	75
Figure 8 – Desktop version functions.....	75
Figure 10 - Dashboard page .....	79
Figure 11 - New text editor .....	80
Figure 12 - Summary function of ProWritingAid .....	80
Figure 13 - Editor providing dynamic AWE feedback .....	81
Figure 14 - Feedback provision by the AWE program .....	106
Figure 15 - Number of changes in each text .....	117
Figure 16 - Percentage difference between texts.....	118
Figure 17 - Self-regulated macroprocesses.....	129
Figure 18 - Decision-making utterances .....	139
Figure 19 - Participant decision-making process .....	145
Figure 20 - Student uptake of AWE feedback.....	146
Figure 21 - Feedback uptake by quality of feedback .....	147
Figure 22 - Participant uptake of feedback by quality .....	147
Figure 23 - Uptake by feedback type .....	149
Figure 24 - Sentence length visualisation .....	152

Figure 25 - Order of functions in ProWritingAid ..... 156

## Summary of Tables

Table 1 - Non-exhaustive list of currently available AWE programs .....	66
Table 2 - Overview of action data collected by the GrammarAid web application .....	82
Table 3 - Databases and types of data collected by GrammarAid .....	82
Table 4 - Participant details .....	85
Table 5 - Coding used for changes made to participant texts .....	90
Table 6 - Linguistic features which correlate to writing proficiency .....	92
Table 7 - Participant details .....	96
Table 8 - Bannert et al.'s (2014) self-regulation processes .....	99
Table 9 - Coding scheme adapted from Bannert et al. ....	102
Table 10 - First and second iteration of coding for behavioural engagement .....	104
Table 11 - Final coding scheme for behavioural engagement .....	105
Table 12 - Actions taken by participants in response to AWE feedback .....	106
Table 13 - Codes for quality of feedback uptake .....	108
Table 14 - User actions in AWE program .....	116
Table 15 - Breakdown of feedback by types .....	119
Table 16 - Differences in scores for lexical characteristics. ....	120
Table 17 - Paired t-test for lexical characteristics .....	122
Table 18 - Individual use of self-regulated strategies .....	136
Table 19 - Feedback uptake per quality per student .....	148



# Chapter 1 - Introduction

With the rise in the number of students learning English worldwide, and academic writing in English becoming an increasingly important skill for the success of students in universities across the world (Hasan & Akhand, 2010; Kadmiry, 2021), many of these educational institutions have begun looking for ways in which to support their students in the acquisition of English writing skills. This increase in English language learners has also led to a shortage of teachers who specialise in teaching English as a second language (ESL) and English as a foreign language (EFL) (Sheehan, 2013), and therefore universities have begun considering technological solutions to provide support for their learners (Jiang & Yu, 2020).

Automated Writing Evaluation (AWE) programs are among the technologies that have been brought into academic writing ESL and EFL classrooms to support student learning of writing skills. They are a subset of educational software that gives feedback to students on different aspects of writing, including grammar, style, cohesion, organisation, and collocations, among others (Stevenson & Phakiti, 2019). These programs are often used to provide formative feedback to student writing with the aim of helping them to improve their writing (Hockly, 2019), especially where the high number of students would make it hard for teachers to provide constant, timely feedback to all of them (Huang & Renandya, 2020).

However, the use of these programs in ESL and EFL classrooms has not been without controversy (Stevenson & Phakiti, 2019; Woodworth & Barkaoui, 2020). There are many questions regarding the effectiveness of these programs (Dikli, 2010; Ranalli et al., 2017) and the way in which they should be integrated into ESL and EFL classrooms (C.-F. Chen & Cheng, 2008; Jiang et al., 2020; Tang & Rich, 2017), if at all (Cheville, 2004). Technology integration into educational settings is a complex process that involves educational change and necessitates meaningful inclusion (Scherer et al., 2019), and

current research indicates that education technologies are not being utilised effectively in classrooms (Thomas Ryan & Bagley, 2015). Given the mixed nature of research results in the usage of AWE programs in ESL and EFL classrooms, it is safe to assume that this may also be the case in many of the settings in which AWE programs have been introduced.

Research into the use of AWE programs as tools for formative feedback has been fragmented and inconclusive (Hibert, 2019b; Stevenson & Phakiti, 2014). Very little is known about how students engage cognitively, behaviourally and affectively with AWE programs, and most of the studies done on student usage of AWE programs have had less than five participants (Link et al., 2020; Zhang, 2017; Zhang & Hyland, 2018). Research into student engagement with AWE programs is a small but growing part of AWE research, since only through effective engagement can students reap the benefits of receiving AWE feedback (Zhang, 2020), and engagement has been correlated to better retention of form when using AWE feedback (Saricaoglu & Bilki, 2021).

The purpose of this thesis is to add to the growing body of research on student engagement with AWE programs in ESL and EFL settings. By researching how students use and engage with AWE programs, it is possible to support the creation and development of research-informed pedagogical strategies that allow students to effectively engage with the feedback they receive from these programs (Tian & Zhou, 2020).

Therefore, in this thesis I focus on how AWE programs are used by students to identify areas of opportunity that can aid the integration of AWE programs into ESL and EFL classrooms as tools for formative feedback. For this aim, the main objectives of this thesis are:

- Understand how students use AWE programs to revise their texts, focusing on how students use self-regulation strategies to engage with the feedback they receive from these programs and how this engagement affects the final product of their revision

- Propose research-informed strategies for the implementation of AWE programs as tools for formative feedback based on how student engagement with AWE programs can be supported through pedagogical strategies.

## **1.1 Background**

The first AWE program, Project Writing Grade (Page, 1966) was developed in the 60s to use statistical analysis to replicate human ratings of writing quality. The potential of these programs was eventually recognised as a way to help score and evaluate large-scale, high-stakes writing tasks such as the TOEFL tests (Hockly, 2019) in the late 90s, with the three most widely available commercial systems being developed by Pearson Educational Technologies, Vantage Learning and Education Testing Services (Stevenson & Phakiti, 2019). As natural language processing (NLP) advanced in its capacity to analyse texts, AWE programs gained the ability to not only provide summative evaluation but began to be used to provide formative evaluation on the aforementioned aspects of writing such as grammar, spelling, punctuation, style and organisation (Stevenson & Phakiti, 2014).

The use of these programs soon extended to English as a second language (ESL) and English as a foreign language (EFL) classrooms, where the number of students enrolled made giving personalised feedback to all of them unfeasible (Jiang et al., 2020). The main difficulties students face in these settings are usually related to the usage of appropriate vocabulary, correct grammar forms, and development of ideas (Neupane, 2017), all areas which AWE programs claim to address. While most AWE programs are not designed with ESL and EFL students in mind, there has been growing interest in the role of AWE programs as tools for formative feedback in ESL and EFL classrooms, and so some existing programs have started to adapt to these new markets. These adaptations include adding features such as multi-language grammar handbooks and providing feedback in languages other than English (Stevenson & Phakiti, 2019). Developers in other countries have also begun to enter the AWE market by developing their own AWE programs designed for their own EFL populations. China, for example, has developed its own AWE programs, such as Pigai and iWrite, which are geared

toward native Chinese speakers learning how to write in English, emphasising feedback on collocations and giving the students lists of synonyms to expand their vocabularies (Huang & Renandya, 2020).

### **1.1.1 Controversies surrounding the use of AWE programs**

There are an estimated 1.2 billion English learners around the world, and one of the biggest challenges in English language education is the shortage of teachers (Sheehan, 2013), which results in unmanageable class sizes that hinder the development of language abilities (Jiang et al., 2020; Tang & Rich, 2017). Given the importance of English as a language of international communication, and the importance of mastering writing skills in academic success (Kadmiry, 2021), it is no surprise that some universities are turning to technological means for giving their students additional support in this area. While there is some debate as to what emphasis, if any, should be placed on formal linguistic features when teaching second language (L2) writing, it can be argued that the correct usage of language forms is especially important to ESL and EFL students, because L2 writing is inherently dependent on language skills (Ryu, 2020), and having mastery over linguistic forms helps students convey their ideas and meanings more clearly (Casal & Lee, 2019).

Even when we accept that L2 instruction should focus on the formal aspects of language, there is still some debate as to how this can be achieved. While AWE programs offer what is commonly known as written corrective feedback, Truscott (1996) challenged the idea that providing this type of feedback to students would help them develop the language skills they needed to produce academic texts and in fact argued that written corrective feedback might harm rather than help L2 development. This sparked an ongoing debate on the merits of providing ESL and EFL students with written corrective feedback (Bitchener, 2012; Ferris, 1999), although current empirical evidence seems to support the notion that written corrective feedback can be useful to L2 learners (Shao, 2015). Ferris (2019) has also pointed out that regardless of this ongoing debate on written corrective feedback, accuracy in writing is important to the academic audiences of L2 writers, since lack of

accuracy in L2 texts may lead to stigmatisation, and students themselves claim to prefer receiving error feedback from their instructors.

In this regard, AWE programs can afford L2 learners more opportunities to receive error feedback thanks to their ability to provide immediate feedback on several aspects of writing. AWE programs as tools for formative feedback are touted as having a myriad of benefits, such as taking over low-level grammatical and mechanical feedback so teachers can focus on providing higher level feedback on content and organisation (J. Li et al., 2015). These programs can also allow students to receive instantaneous feedback on their drafts to suit their workflow (Stevenson & Phakiti, 2014), motivating them to create several drafts of their writing and edit based on said immediate feedback (Saricaoglu & Bilki, 2021), and helping them develop the metalinguistic knowledge they need to improve their English writing skills (Link et al., 2014). AWE programs have also been promoted due to their consistency of feedback and scoring (Dikli, 2006), compared to teacher feedback which can be inconsistent and inaccurate (Truscott, 1996), although some would call AWE feedback generic and unhelpful rather than consistent (Zhang, 2020).

AWE programs have also been criticised for prioritising a formulaic view of writing (Cheville, 2004) which ignores the individual creativity and expression of students (Vojak et al., 2011) and does not evaluate their ability to write for an audience (Hockly, 2019). Furthermore, AWE programs have also been accused of giving vague and confusing feedback that does not necessarily help students revise their texts (M. J. Wang & Goodman, 2012), and there is not enough evidence that these programs can help students improve their writing skills (Stevenson & Phakiti, 2019).

### **1.1.2 The usefulness of AWE programs**

The question of whether AWE programs can serve as tools for formative feedback has led to a proliferation of research that seeks to understand whether the use of these programs can help students improve their writing skills and the place these programs might have in ESL and EFL classrooms. Current research into the use of AWE programs can be divided into three main strands.

The *first* is research into the reliability of AWE tools when providing feedback and scores for evaluating student writing. These are measured by precision rates (how many of the errors identified by the program are true errors) and recall rates (how many errors are detected by the program) for AWE programs and how well their scoring aligns with scores assigned by human raters (Xi, 2010). The *second* strand of research seeks to understand whether the use of AWE programs can lead to improvements in student writing. These are usually operationalized as increases in writing scores provided by either the AWE programs or human raters, or reductions in error rates (Hibert, 2019b). The *third* strand of research, and one that has become increasingly common in the last few years, seeks to understand how students engage with AWE programs as part of academic writing classes and techniques through which these programs can be integrated into process-oriented approaches to teaching English writing skills (Jiang et al., 2020; Link et al., 2020; Tang & Rich, 2017).

Research on whether AWE programs can emulate human scores has been mostly positive, with the rate of agreement between the AWE scoring engine and human raters being very similar to the rate of agreement between human raters (Dikli, 2006). However, research on whether AWE programs can be used to reduce error rates or increase writing scores is fragmented, both theoretically and methodologically (Hibert, 2019b). Thus, there is little consensus as to whether there are quantifiable benefits to the use of AWE programs in ESL and EFL settings, either in the reduction of errors in texts or increases in scores (Stevenson & Phakiti, 2014, 2019). Furthermore, research into how students use AWE programs to engage with their own writing and how these systems can be integrated into the classroom is on its first stages, although there are some encouraging findings that show students engage with AWE feedback cognitively, behaviourally and affectively (Jiang & Yu, 2020; Koltovskaia, 2020; Zhang, 2020), and are selective in their uptake of AWE feedback (Bai & Hu, 2017; Chodorow et al., 2010). However, there remain many questions not only about the effectiveness of AWE programs, but also about the nature of student engagement with AWE feedback and how it can be supported in the classroom through pedagogical interventions and strategies.

Given that these programs are already being used in thousands of ESL and EFL classrooms around the world (Qian et al., 2020), the most pressing question at the moment is not whether these programs should be introduced into classrooms, but rather how these programs can support the teaching and learning of academic writing skills in universities. That is not to say that the question of the usefulness and pertinence of introducing AWE programs into classrooms should be ignored or that their presence in ESL and EFL classrooms deserves no scrutiny, however, both lines of inquiry are outside the scope of this thesis. My purpose in undertaking this research is to understand how the current usage of AWE programs can be supported through theory-supported and research-informed practices in light of the fact that these programs are having an increasing presence in ESL and EFL classrooms.

## **1.2 Thesis outline**

In this thesis, I argue that understanding how AWE programs can be used in ESL and EFL classrooms requires a deeper understanding of how students use AWE feedback and their thought processes while engaging with it. Therefore, I address research oriented around the product of AWE usage and research oriented into the process of AWE usage, roughly mapping them to the text-oriented and writer-oriented approaches to teaching writing that have been commonly used to teach, research and understand English academic writing (Hyland, 2015). I will argue that both approaches to researching the use of AWE programs are integral to understanding the role these programs could play as tools for formative feedback in ESL and EFL classrooms. By using novel research methodologies to examine both the process of students using AWE to revise and the finished product of their revisions, I bring both product and process perspectives together to provide a more holistic understanding of how these programs are used by postgraduate students.

The main aim of this thesis is to contribute to existing research into the use of AWE programs by using novel data collection and analysis techniques to provide a more in-depth understanding of how students use AWE programs. Having a better understanding of how students engage with the

feedback they receive from AWE programs can provide a stronger foundation for research-informed implementation of these programs as tools for formative feedback in ESL and EFL classrooms. To do so, the organisation of this thesis is as follows:

Chapter 2 gives a general overview of the relevant literature around the use of AWE programs and the research that has been carried out to understand in the proficiency level of texts written by ESL and EFL students. This chapter is divided into two parts. The first is a systematic literature review of research into the use of AWE programs in higher education, which was presented at the 2019 EC-TEL conference in the Netherlands and subsequently published as part of the conference proceedings. The second part of the chapter first gives an overview of the directions AWE research has taken since the paper was published in 2019, then picks up from the recommendations in the paper and delves into topics related to theoretical and methodological considerations around AWE research which can help provide a better understanding of the usage of AWE programs as tools for formative feedback, and is therefore used as a theoretical base for the rest of this thesis.

Chapter 3 lays out the general methodology for the research carried out as part of this thesis. It explains the two main stages of data collection and the justification for conducting both: namely the necessity of studying the use of AWE programs from both a product-oriented and a process-oriented approach to attain a more holistic understanding of how students use AWE feedback to revise their texts. Two different studies were conducted to examine each of these aspects separately, and this chapter explains how each study was designed and conducted, including data collection, participation, selection, and development of tools for collecting and analysing data, and ethical considerations.

Chapter 4 details the results of the first study carried out for this thesis. A text-centred product-oriented approach was selected for this study, using a web application I programmed to collect texts from students and analysing them to understand how the texts themselves change as a response to AWE feedback. The main purpose of this study was to understand the effect that AWE feedback has



on texts and the nature of student uptake of the given feedback. This study introduces the analysis of linguistic markers of writing proficiency as a novel way to quantify the effect of AWE feedback on students' texts, something that to my knowledge has not been done before in AWE research but could provide a more in-depth understanding of how students revise their texts through the use of AWE feedback.

Chapter 5 details the results of the second study carried out for this thesis. This study took a writer-centred, process-oriented approach to AWE research, looking into how students use an AWE program to revise their texts. The main purpose of this study was to provide depth and context to the findings of the previous study by exploring students' engagement with the feedback provided by the AWE program, and from there to form an understanding about the decision-making processes students use to engage with it. This study used think-aloud protocols to understand student mental processes while they are using the AWE program, something that has been suggested for researching engagement with AWE programs (Link et al., 2020; Zhang, 2020) but to my knowledge has not been used so far.

Chapter 6 brings together the results from Chapters 4 and 5 into a broader discussion about the use of AWE programs as tools for formative feedback in higher education settings. By making extensive use of the literature review carried out in Chapter 2, the findings of both studies are put into context within the wider research around the use of AWE programs to help understand how students engage with and use the feedback they receive from AWE programs. I also discuss the implications these findings have for understanding how these programs can be implemented in ESL and EFL classrooms as tools for formative feedback that support student-centred approaches to learning writing skills.

Finally, Chapter 7 includes the conclusions for the thesis. In this chapter, I discuss the specific pedagogical practices that can be understood to help with the implementation of AWE based on existing literature and the findings of this thesis, with some recommendations for the use of these

programs. I then expand on the limitations encountered during this research and what this means for the overall interpretation of the results, as well as recommendations that future research may follow or take into account to address these gaps or build upon the results of this research.

Suggestions for future research into AWE programs include methodological, theoretical and practical dimensions.

## Chapter 2 – Literature Review

### 2.1 Introduction

Although AWE programs have been in development since the 1960s (Page, 1966), these programs began gaining a steady rise in popularity in the 2000s partially thanks to growing numbers of ESL and EFL students enrolling in English-speaking universities or universities where learning English is mandatory, as well as an increased emphasis on high-stakes standardised testing (Hockly, 2019; Stevenson, 2016). Increases in computing capacity, natural language processing (NLP) and latent semantic analysis (LSA) have also led to an increased interest in developing these systems to further support language learning (Xi, 2010).

This increase in the usage of AWE programs for high-stakes testing has also come hand in hand with an increase in its use for the provision of formative feedback (Hockly, 2019). A growing body of research has begun to explore whether AWE programs can help students improve their writing proficiency and which role these technologies might have in the ESL/EFL classroom (Koltovskaia, 2020; Stevenson, 2016). The purpose of this chapter is to explore existing literature on the use of AWE programs as tools for formative feedback, as well as certain theoretical constructs that might inform further research into these technologies.

This chapter comprises three sections. The first section contains a verbatim copy of a paper I submitted to the European Conference on Technology Enhanced Learning (EC-TEL) in 2019 and subsequently published as part of the proceedings for that conference as consistent with the University of Edinburgh policy for “Including Publications in Postgraduate Research Theses”. This paper consisted of a systematic literature review which explored existing research into the use of AWE programs from the point of view of the theoretical constructs that informed this research and

the different methodologies used to investigate the role of these programs in ESL and EFL classrooms.

The second section of this chapter explores the research that has been carried out into the use of AWE programs since the publication of the EC-TEL paper, including research on cognitive, behavioural and affective engagement with AWE programs, as well as a growing interest in the role teachers play in the implementation of AWE programs in ESL and EFL classrooms.

The third section looks at relevant theoretical frameworks that can be used to address some of the identified gaps in the literature. In this section, I examine existing research on written corrective feedback and how this research can inform our understanding of the use of AWE programs as tools for formative feedback. Then, I discuss how an understanding of self-regulation strategies can help understand how students engage with AWE programs, both from a theoretical and a methodological point of view. I also explore how the use of natural language processing technologies could help us evaluate writing proficiency, defined in this thesis as “a measurement of one’s capacity for written expression that utilises both syntactic and lexical aspects of language” (Ha, 2019, p. 8). Finally, I discuss both frameworks within the framework of process and product approaches to writing, and how this can help inform research into the use of AWE programs. I also present a brief discussion of how these frameworks might be used to advance our understanding of AWE programs as tools for formative feedback, as well as their place within the context of this thesis.

## **2.2 Publication: Systematic Literature Review of Automated Writing Evaluation as a Formative Learning Tool**

The following section contains the verbatim copy of the following publication:

Hibert, A. I. (2019). Systematic Literature Review of Automated Writing Evaluation as a Formative Learning Tool. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming Learning with Meaningful Technologies .EC-TEL 2019. Lecture Notes in Computer*

*Science* (Vol. 11722, pp. 119–212). Springer International Publishing.

<https://doi.org/10.1007/978-3-030-29736-7>

## **2.3 New Directions on AWE Research**

The systematic literature review presented in the previous section concluded that research into the use of AWE was fragmented. Too much focus has been placed on the written text as the end product of revision with AWE tools, and not enough on the processes that lead to this product. This makes it hard to use existing evidence to inform the implementation of AWE programs in the classroom as we know little about how students use these programs beyond research done on post-hoc student perceptions and attitudes toward the use of AWE (C.-F. Chen & Cheng, 2008; Fang, 2010; Huang & Renandya, 2020; Lai, 2010).

In the last few years, however, research into AWE programs has begun to take on a new direction that broadly aligns with the conclusions presented in the paper reproduced in the previous section and focuses more on student usage of AWE programs (Jiang & Yu, 2020). There is growing interest in how students engage with AWE programs and how they use the feedback they receive from them, and the literature published in the last three years reflects this renewed interest (Koltovskaia, 2020; Li et al., 2019; Link et al., 2020; Tian & Zhou, 2020; Wang et al., 2020; Zhang, 2020; Zhang & Hyland, 2018, among others). Furthermore, there is a slowly growing research interest on the role of teachers in mediating the use of AWE feedback and how their attitudes, pedagogical strategies, and agency affect student usage of these programs, although research in this topic is still quite sparse (Jiang et al., 2020; Z. Li, 2021; Link et al., 2020).

### **2.3.1 Behavioural engagement with AWE feedback**

Although some previous research into the use of AWE programs has focused on constructs that could be roughly considered related to engagement, such as motivation (Cotos, 2012; Fang, 2010) and metacognition (Liao, 2016a), Zhang and Hyland (2018) were one of the first to integrate existing theory on engagement into the study of how students use AWE programs. They adapted Fredricks, Blumenfeld, and Paris's (2004) tripartite division of engagement as emotional, cognitive and

behavioural to the use of AWE programs. Fredricks et al. (2004) define behavioural engagement as a construct which encompasses positive conduct, involvement in learning and academic tasks, and participation in school-related activities. Emotional engagement is defined as the interests, values, and emotions that underpin students' affective reactions in learning environments. Finally, their definition of cognitive engagement includes motivation, effort, and strategic learning.

In Zhang and Hyland's (2018) adapted framework, cognitive engagement refers to the extent to which students attend to the feedback they receive, behavioural engagement refers to how learners revise their texts and their uptake of feedback, and their perceptions and attitudes toward the use of AWE programs would fall under emotional engagement (Zhang & Hyland, 2018). Using this framework as a starting point, studies which have looked at student uptake of feedback can tell us about how students engage behaviourally, cognitively, and affectively with AWE feedback.

**2.3.1.1 Student uptake of feedback.** Chodorow (2010), Bai & Hu (2017), Koltovskaia (2020) and Tian & Zhou (2020) looked at student uptake of feedback and found that students ignore most of the feedback they receive, but all three reached the conclusion that this might be because not all the feedback given by the AWE programs was useful and participants were actively selective about which feedback they incorporated into their revisions, although Koltovskaia found that most of the feedback students did use was superficial. However, the notion that feedback uptake is a binary between blind acceptance and selective behaviour has been called into question (Jiang & Yu, 2020).

For example, Jiang & Yu used Activity Theory (AT) to contextualise student uptake of AWE feedback and found three different patterns of what they call appropriation behaviour, which they define as a process through which a person adopts tools to use in particular environments. This process allows students to internalise ways of thinking characteristic to specific practises. Their findings showed that students' appropriation behaviours differed in number of submissions, revision operations, time spent, and goals of appropriation (Jiang & Yu, 2020). That is, students' usage of AWE feedback depended on their revision goals and the objects of their appropriation, since teacher and learner

goals for a writing task might not be the same (Bitchener, 2012), and these goals might even be different for each individual learner (Jiang & Yu, 2020). These learner goals identified by Jiang & Yu ranged from wanting to improve accuracy for the sake of obtaining a higher score rather than achieving mastery of writing forms, to wanting to write for meaning rather than focusing on form, to merely wanting to finish writing tasks.

This stands in contrast to previous literature which suggested that student uptake of AWE feedback was dependent on students' proficiency level. Some studies have suggested that higher proficiency students would benefit more from AWE feedback because their linguistic knowledge helped them make better sense of the feedback they received and incorporate it into their work (Tian & Zhou, 2020; Zhang & Hyland, 2018), although linguistic knowledge by itself did not mean that students were able to accurately judge the adequacy the feedback they receive from AWE programs (Koltovskaia, 2020). Lower proficiency students have been found to have more limited cognitive and behavioural engagement with AWE feedback because they do not have the knowledge needed to understand it (Zheng & Yu, 2018).

In contrast to this perspective, Chen & Cheng (2008) found that high-proficiency students were dissatisfied with the feedback they received from the AWE program, as they considered it to be too mechanical and generic, and suggested that lower proficiency students might make better use of it. This may be because lower proficiency students placed more emphasis on language form than meaning (Huang & Renandya, 2020) and the multiple drafting capabilities offered by AWE programs provide more opportunities for noticing of errors in form (Liao, 2016a). To further complicate the issue, Li et al. (2017) found very little difference in error reduction rates between intermediate-high and advanced-low proficiency levels.

The fact that research results have been mixed with regard to the relationship between proficiency level and behavioural engagement with AWE feedback strengthens Jiang & Yu's (2020) argument that there are other mediating factors that need to be taken into account to understand how



students appropriate or integrate AWE feedback into their revision practises. As a response to this, both the Technology Acceptance Model (TAM) and the use of sociocultural perspectives have been proposed as theoretical frameworks that can help explain some facets of student engagement with feedback.

**2.3.1.2 Technology Acceptance Model (TAM).** Some recent research has attempted to bring another dimension of understanding to student behavioural engagement with AWE feedback by extending the TAM to capture the complex network of factors that affect student adoption of AWE programs. The TAM was originally developed to understand adoption of technology, based on the idea that perceived usefulness and perceived ease of use are the main factors that determine a person's behavioural intention to use technology, a model which has since been extended to encompass other factors such as social influence and cognitive influence processes (Venkatesh & Davis, 2000).

Zhai and Ma (2021) analysed the driving factors behind student acceptance of AWE technologies and used their research findings to propose a model which sought to explain the factors that mediate how students accept and use AWE programs. Perceived usefulness of AWE feedback was found to be one of the main factors for student acceptance (Lai, 2010; R. Li et al., 2019). However, other environmental factors were also found to have an effect. Subjective norm, the perception that people important to the subject such as teachers and peers feel they should use the program, was found to significantly affect perceived usefulness. This finding is consistent with previous literature which suggests that teacher perceptions of AWE programs have a significant impact on student perceptions and willingness to use AWE feedback (C.-F. Chen & Cheng, 2008; J. Li et al., 2015; Tang & Rich, 2017).

The impact of teachers in the adoption of AWE programs is further evidenced by the fact that the infrastructure supporting use of the program had a significant relationship with perceived ease of use in this model. That is, students who were given facilitating resources, technical knowledge, and other forms of support were more likely to adopt AWE feedback (Zhai & Ma, 2021). Again, this is

consistent with previous literature, which has identified teacher scaffolding as one of the key factors in the successful adoption of AWE feedback by ESL/EFL students (Liao, 2016a; Tang & Rich, 2017; M. J. Wang & Goodman, 2012; Woodworth & Barkaoui, 2020; Zheng & Yu, 2018).

One of the main takeaways of research into the use of the TAM to contextualise student adoption of AWE feedback is the importance of contextual and social factors in students' ability to make sense of the feedback and integrate it into their revision process. Writing is a socially embedded process (Hockly, 2019) in which meaning is not necessarily the product of using correct form but of engagement with active readers (Cheville, 2004) and the negotiation of meaning (C.-F. Chen & Cheng, 2008). AWE programs have often been criticised for ignoring these social aspects of writing (Stevenson & Phakiti, 2019; Vojak et al., 2011) and missing the dialogic aspects of the writing and learning process (Mehrabi-Yazdi, 2018), although some attempts have been made to address these shortcomings by combining AWE feedback with teacher and peer feedback (Koh, 2017; Lai, 2010; Tang & Rich, 2017). In order to understand the role of revising AWE feedback as a socially situated activity, some researchers have turned to sociocultural theory to research how students engage with AWE programs (Jiang & Yu, 2020).

**2.3.1.3 Sociocultural theory.** Sociocultural theory has often been used in general written corrective feedback (WCF) research, as WCF has been considered an important tool that can be used to scaffold L2 development (Woodworth & Barkaoui, 2020) through socially mediated activities that help explain learner engagement and feedback uptake (Tian & Zhou, 2020). From this perspective, teachers can be considered as subjects motivated towards a specific object, providing feedback to students, mediated by the use of artefacts or tools such as AWE programs (Jiang et al., 2020), and AWE adoption is a result of the interactions between teachers, students, and system working together to achieve a stated goal (Zhai & Ma, 2021).

This theory puts the concept of scaffolding at the centre of the interaction between teachers, students, and AWE programs. Scaffolding is usually considered to be the provision of help from a

knowledgeable other to a learner, gradually releasing more responsibility to the learner (Palermo & Thomson, 2018). This process eventually allows the learner to gain independence with regard to a particular domain or area of learning (Kuiper et al., 2017). This 'knowledgeable other' does not necessarily have to be a teacher or instructor but can also include more knowledgeable peers. When talking about learning writing skills, WCF can scaffold students' development by highlighting their weaknesses and strengths, and helping them to develop their consciousness of form and conventions (Mahfoodh, 2017).

Some current research into the use of AWE programs has considered the role of these technologies as tools that can help teachers scaffold writing skills (Cotos et al., 2020; Hibert, 2019a; Hockly, 2019; Z. Li, 2021; Link et al., 2020; Tang & Rich, 2017; Waer, 2021), although most of the previous research has focused on how teachers can scaffold the use of AWE programs in the classroom (C.-F. Chen & Cheng, 2008; Huang & Renandya, 2020; Jiang et al., 2020; Lavolette et al., 2015; Liao, 2016a, 2016b; M. J. Wang & Goodman, 2012; Zhang, 2017; Zheng & Yu, 2018) or has focused on middle-school students (Grimes et al., 2010; Palermo & Thomson, 2018) rather than university and postgraduate students, which are the focus of this thesis.

Both perspectives, the role of teachers in scaffolding the use of AWE programs and the role of AWE programs in helping teachers scaffold writing skills, are valuable for understanding how to integrate AWE technologies in the classroom. From the point of view of behavioural engagement, I have already discussed how the TAM considers the importance of teacher support and their attitudes toward AWE programs in the adoption of AWE as a tool for revision. Research has shown that students are less satisfied and less willing to engage with AWE feedback when teachers show mistrust toward AWE feedback, or when they use AWE feedback as a replacement for their own (C.-F. Chen & Cheng, 2008; Jiang et al., 2020). When teachers take on a mediating role between the student and the AWE feedback, students tend to report satisfaction with the amount and quality of

the feedback they receive (J. Li et al., 2015; O'Neill & Russell, 2019). Still, the link between student attitudes toward the use of AWE feedback and their actual behavioural engagement is still unclear.

What is clear from current research into the use of AWE programs is that they cannot be used as a substitute for teacher feedback or teacher interaction with students (C.-F. Chen & Cheng, 2008) given that the feedback provided by AWE programs is often unreliable (Bai & Hu, 2017) and too generic to be helpful (Dikli, 2010), which results in mostly surface-level revisions by students (Koltovskaia, 2020). The role of teachers in using AWE tools to scaffold writing skills development is discussed in section 2.3.4, but another construct that can help us understand how students interact with AWE feedback is that of cognitive engagement.

### **2.3.2 Cognitive engagement with AWE feedback**

While behavioural engagement, or constructs that can give insight into how students engage behaviourally with AWE feedback, has been a focus of research into the use of AWE, cognitive engagement has not been as widely studied. To the best of my knowledge, only five studies have been carried out to date which specifically investigate cognitive engagement with AWE feedback (Koltovskaia, 2020; Tian & Zhou, 2020; Zhang, 2017, 2020; Zhang & Hyland, 2018), although some studies do consider metacognition as a factor in AWE usage without explicitly using engagement as a theoretical construct (Cotos, 2012; Liao, 2016a).

Cognitive engagement has been conceptualised as both a psychological investment in learning and the cognition processes that allow for strategic learning (Fredricks et al., 2004). In the domain of research into AWE feedback, cognitive engagement refers to how students process AWE feedback, either by noticing or understanding (Koltovskaia, 2020), and includes both revision strategies and metacognitive strategies (Zhang & Hyland, 2018). Part of the reason cognitive engagement has not been studied as extensively as behavioural engagement might derive from the fact that cognitive engagement is less directly observable as it focuses on efforts that happen within the mind of the learner (Henrie et al., 2015).

The lack of research into cognitive engagement presents a large gap in our understanding of how students use AWE feedback, as cognitive and behavioural engagement are closely linked to each other and influence each other (Fredricks et al., 2004). Koltovskaia (2020) has pointed out that, without cognitively engaging with the feedback they receive, students can still improve the accuracy of the texts by randomly accepting feedback in the same way one can pass a multiple-choice exam by guessing, but this does not lead to true learning.

Cognitive engagement with AWE feedback, like behavioural engagement, is also mediated by a range of factors, including previous linguistic knowledge and experience (Tian & Zhou, 2020) and the students' ability to use cognitive and metacognitive strategies (Zhang & Hyland, 2018). For example, goal setting was identified by Jiang & Yu (2020) as one of the most important mediators for appropriating AWE feedback. Zhang (2020) also found that monitoring skills helped students use AWE feedback to create awareness of the gap between their current and desired understanding of form, as well as monitoring their revision process. The use of metacognitive strategies has also been linked to higher levels of grammatical accuracy in writing classes that use AWE programs to provide formative feedback (Liao, 2016a).

Most of the research on cognitive engagement with AWE feedback, however, has focused on post-hoc interviews and recalls. One issue with post-hoc self-reports is that human memory is imperfect and therefore there is the possibility that students might misremember (Winne, 2013), something that could be alleviated by using think-aloud protocols to examine how students use cognitive and metacognitive skills to navigate AWE feedback (Link et al., 2020; Zhang, 2020), as these methods have been shown to produce reliable data on cognition (J. A. Greene et al., 2011). However, to my knowledge, no such study regarding AWE feedback has been carried out.

### **2.3.3 Affective engagement with AWE feedback**

Affective engagement with AWE feedback is another dimension of engagement that has been scarcely studied as such, but there is plenty of research into student perceptions of using AWE

feedback to revise that can give us a broad understanding of how they engage affectively with it. Affective engagement, which Fredricks et al. (2004) called ‘emotional engagement’, refers to affective reactions to learning, including emotions such as interest, anxiety, and boredom, among others. In terms of interaction with AWE programs, affective engagement can be defined as students’ emotional and attitudinal reactions to feedback (Zhang & Hyland, 2018). In the field of AWE studies, there is a significant subset of research that has delved into how students perceive AWE programs and their emotions when interacting with them (e.g. Chen & Cheng, 2008; Fang, 2010; Lai, 2010; O’Neill & Russell, 2019, among others). Studies into the application of the TAM into the use of AWE feedback also consider the dimension of computer anxiety, which would also fall under the affective engagement category (R. Li et al., 2019; Zhai & Ma, 2021).

**2.3.3.1 Emotional reactions to AWE.** Student perceptions of the usefulness of AWE feedback are important because they are predictors of whether students will adopt AWE programs for revision or not (Zhai & Ma, 2021), and negative attitudes toward AWE programs is correlated with students and teachers being unwilling to integrate AWE feedback into their revision process (C.-F. Chen & Cheng, 2008).

Regulating emotions has also been found to be an important mediating factor in the use of AWE feedback and the improvement of their writing skills, because students who are better able to regulate their emotions also persevere in the face of frustration and are less discouraged by the programs’ limitations (Jiang & Yu, 2020; Zhang & Hyland, 2018). However, when it comes to student attitudes and emotional reactions to AWE feedback, current research is mixed.

For example, frustration seems to be an emotion that is commonly connected to the use of AWE feedback in the literature. Li et al. (2015) found that students were often frustrated with the inconsistency between scores assigned by the AWE programs and scores assigned by the teacher, something that negatively affected their perceptions of the program’s usefulness. Writing scores can also be a source of frustration for students when they do not get the desired scores, their scores are

too low, or they do not correspond to scores given by teachers (J. Li et al., 2015; Zhang, 2017; Zhang & Hyland, 2018). Another common source of frustration when using AWE programs is the generic nature of the feedback they give (C.-F. Chen & Cheng, 2008; Dikli, 2010; Lai, 2010; M. J. Wang & Goodman, 2012), or when students receive the same generic feedback even after making revisions (S. Wang & Li, 2019).

Student perceptions of AWE feedback also varied depending on the type of feedback they received. For example, their perceptions of grammar and mechanic feedback tended to be positive, leading to a greater uptake of those types of feedback (Bai & Hu, 2017; Koltovskaia, 2020), while the reverse is true for content and organisational feedback, which tends to be more generic and less focused, and thus less helpful (Fang, 2010; J. Li et al., 2015). That is not to say that the emotions generally associated with the use of AWE programs to revise texts are all negative. Some studies have found that students in general seemed to hold positive attitudes toward the use of AWE feedback in their classes in spite of its technical limitations (Fang, 2010; Huang & Renandya, 2020; J. Li et al., 2015). Some have even found that making students explicitly aware of these limitations can help them develop the ability to critically evaluate feedback (Z. Li, 2021). Furthermore, studies suggest that using AWE programs to can encourage student autonomy and increase their confidence in their writing (M. J. Wang & Goodman, 2012; Xia & Zhong, 2017; X. Yang & Dai, 2015).

However, while previous research has focused on the factors that affect student perceptions and emotions when using AWE feedback, current research has begun to shift focus toward how affective engagement impacts revision behaviours and the use of cognitive and metacognitive strategies in revision. That is, research is now starting to focus on the link between affective, cognitive, and behavioural engagement and what this means for the adoption of AWE programs in the classroom. Findings so far suggest that the relationship between affective engagement and other types of engagement is quite complex.

Different types of affective engagement have different effects on cognitive and behavioural engagement, and how these relate to each other. However, this relationship is not as simple as positive affective engagement leading to higher cognitive and behavioural engagement. Affective engagement can lead to increased cognitive engagement in several ways. For example, negative affective engagement manifesting itself as questioning the AWE feedback can lead to higher levels of cognitive engagement (Koltovskaia, 2020). However, when affective engagement manifests itself as positive emotions such as 'motivation', it can also lead to better cognitive engagement (Zhang & Hyland, 2018) when students focus on the advantages of AWE feedback to increase their awareness of the importance of revision (Zhang, 2020). On the other hand, positive affective engagement with the AWE programs which resulted in students trusting the feedback given to them by the program led to lower levels of cognitive engagement with it (Koltovskaia, 2020).

Negative emotions brought about by the use of AWE, like anxiety, also led to lower levels of cognitive and behavioural engagement when students do not have the linguistic skills to make use of feedback (Zhang & Hyland, 2018). Anxiety in particular is an important emotion to analyse with regard to affective engagement, as it affects whether users are willing to adopt specific technological tools (Venkatesh & Davis, 2000) and has been found to negatively affect engagement (Perski et al., 2017).

**2.3.3.2 Computer anxiety.** In the previous discussion on how the TAM can be adapted to inform behavioural engagement with AWE programs, the constructs of computer anxiety and AWE anxiety were glossed over in favour of discussing other factors that affected student adoption of these programs. Computer anxiety can be defined as the anxious reactions negatively related to performance that a user experiences when using a certain program (Zhai & Ma, 2021).

Students in some studies have expressed that the complexity of the many functions offered by AWE programs, as well as some technical problems, caused them anxiety and led to a lower uptake of AWE feedback (Lai, 2010; Zhang & Hyland, 2018). The inconsistency of the scores provided by AWE



programs, especially when they were used as part of the students' grades, were also a common source of anxiety (C.-F. Chen & Cheng, 2008; J. Li et al., 2015).

However, other research on the factors that affect student willingness to use AWE feedback has found that students using AWE programs did not experience anxiety as a strong emotion related to their usage of the tool (M. J. Wang & Goodman, 2012), and others have found no evidence to suggest that computer anxiety has any significant relationship to behavioural intention to use (R. Li et al., 2019) or perceived ease of use (Zhai & Ma, 2021). Zhai & Ma attributed the lack of anxiety in their participants to the interactions between teachers and students and the technical support that both received while using these programs.

To summarise, current research into the use of AWE programs as tools for formative feedback has begun to move away from trying to determine whether these programs result in gains to certain measures of writing proficiency and into examining how these technologies can be used effectively in various learning contexts (Jiang & Yu, 2020). This new focus of research into the use of AWE programs has found that, in general, when students engage with the feedback they receive, they not only make better use of the feedback and tools afforded to them by the AWE program, but the feedback also leads to a more deliberate examination of the texts students produce and their own writing processes (Cotos et al., 2020; Zhang, 2020).

However, engagement alone cannot guarantee that students will make substantial revisions to their texts, especially if they lack the requisite linguistic knowledge to make sense of the feedback they receive (Koltovskaia, 2020). Even when students have a higher linguistic proficiency, this alone does not mean that they will activate previous knowledge and use available resources to engage with the AWE feedback (Jiang & Yu, 2020). This is why the role of teachers is of the utmost importance in the integration of AWE programs into ESL and EFL classrooms.

### **2.3.4 The role of teachers**

One of the main takeaways from this literature review so far is the importance of teachers in the process of introducing AWE programs into classrooms as tools for formative feedback, especially given that “effective implementation of AWE is strongly influenced by how instructors overcome the drawbacks of technology to maximise its full potential in the classroom” (Link et al., 2014, p. 339).

There are two main points that need to be discussed regarding the role of teachers in implementing AWE feedback in the classroom. First, it is important to discuss the effect these technologies have on teacher workload and the amount of feedback they provide to the students. Second, it is important to review current research on the pedagogical practices that can be used by teachers to integrate AWE programs into the classroom in a way that helps support student learning and the improvement of students’ writing skills.

**2.3.4.1 Amount of feedback.** AWE programs have long been touted as a way to liberate teachers from the obligation of continuously providing feedback, especially low-level feedback, to students in settings where the amount of students would make this constant provision unfeasible (Vojak et al., 2011). However, merely using AWE feedback as a replacement for teacher feedback tends to lead to frustration among students (C.-F. Chen & Cheng, 2008) and to a lack of student engagement with the feedback (Jiang et al., 2020).

Therefore, there needs to be a balance between the role of the teacher and the use of the AWE program to support students in their writing process. The general consensus seems to be that, with AWE taking care of low-level mechanical feedback on form, teachers can instead focus on feedback relating to content and organisation (Stevenson, 2016; Stevenson & Phakiti, 2019). For the most part, it seems as though this might be the case, as some studies have found that using AWE programs significantly reduced teacher workload (Link et al., 2014).

And yet, the fact that the AWE programs take on some of the work of providing low-level feedback on student drafts does not necessarily mean that this will translate into more higher level feedback

from the teachers. In this case, high-level refers to the organisation of ideas, while low-level involves grammatical and lexical factors (Jung et al., 2019). Both Link et al. (2020) and Jiang et al. (2020), for example, found that while teachers who used AWE programs tended to give less feedback than the teachers in the control group, teachers on both groups provided a similar amount of higher-level feedback. Having to provide less low-level feedback did not mean teachers would give more higher-level feedback, as this seems to be dependent on both course objectives and how much help with content and organisation specific students require.

In general, Jiang et al. (2020) identified three different patterns of teacher use of AWE programs: teachers who resisted using the AWE program, teachers who used it as a replacement, and teachers who used it as a supplement. These are roughly similar to the findings reported by Chen and Cheng (2008), where one of the teachers studied mistrusted the program and used it infrequently, while another integrated AWE feedback into a process writing approach which also included peer and teacher feedback, and a third who offloaded most of the work of providing feedback and grading to the AWE program. In both of these studies, the more successful implementations of AWE programs were done by teachers who used it as a supplement to their teaching rather than as substitute teacher.

**2.3.4.2 Pedagogical strategies.** Evidence shows that students make better use of AWE programs when they are integrated as part of a focused pedagogical strategy based on a process approach to teaching writing (Tang & Rich, 2017) and that teachers have enormous influence on student perceptions of these programs and students' ability to use AWE feedback (C.-F. Chen & Cheng, 2008; J. Li et al., 2015). It is therefore surprising that little research has directly addressed the role teachers play within this process, the specific pedagogical strategies that might help integrate AWE programs into the classroom to scaffold writing skills, and teacher's views of using this technology to support students.

Furthermore, very little research carried out into the use of AWE programs in the classroom has focused on or even mentioned equipping teachers with the necessary tools and knowledge to take advantage of the capabilities afforded by AWE programs. Among those studies which focus on classroom implementations, most make no mention of having provided participant teachers with any training related to the use of the AWE program used for that study (Bai & Hu, 2017; Liao, 2016a; Link et al., 2020; Ranalli, 2018), or they focused exclusively on the technological capabilities of the AWE tool rather than pedagogical strategies for its implementation (C.-F. Chen & Cheng, 2008; Jiang et al., 2020). Link et al. (2014), for example, provided no formal training for their teachers in the use of the AWE program, and although the teachers in the study were identified as being proficient users of technology, one of the main conclusions of their study was the need for full training in the use of these tools, both from the teacher and student standpoint.

A study carried out by Tang & Rich (2017) is unusual in this regard, since it used an approach in which teachers that participated were given training on the use of the AWE program, ways of incorporating AWE feedback into their writing instruction, and the challenges of using AWE feedback to provide formative feedback. Teachers were also instructed to require students to achieve a minimum score from the AWE program before submitting assignments, and to combine AWE feedback with peer and teacher feedback. They also had the teachers use AWE-generated reports of student writing to give specific feedback based on the assessment criteria for the class. The teachers reported that this approach led to the students feeling more motivated to rewrite and revise, guided students' writing and revising according to the assessment criteria and helped the students become more autonomous.

Regarding pedagogical approaches, while little literature on the use of AWE programs has focused on specific strategies, some lessons can be gleaned from the different ways in which these programs have been used. Specifically, it seems that using AWE programs at the early stages of the revision process, to be later supplemented by teacher and/or peer feedback, seems to lead to students

having better opinions of the AWE program and the writing process itself (C.-F. Chen & Cheng, 2008), and better outcomes as measured by score gains (Tang & Rich, 2017).

Another common approach to integrate AWE programs into the classroom has been by using the scoring function of the program to encourage the students to submit more drafts of their texts to improve their scores (C.-F. Chen & Cheng, 2008; J. Li et al., 2015; Z. Li, 2021; Tang & Rich, 2017). That is, students are asked to achieve a minimum score in the AWE program before they can submit it for further feedback or for evaluation. Some authors have argued that such an approach might encourage students to revise their texts more often, as students may consider getting a better score as a game (P. Wang, 2013).

However, this pedagogical strategy needs to be implemented with caution, given the fallibility of the scoring engine on most AWE programs. These scoring engines can privilege longer essays, which means scores can be artificially inflated by adding more paragraphs even if they are not related to the rest of the text (Hockly, 2019). They also privilege the use of syntactically complex structures and complex vocabulary, and some researchers have successfully obtained full marks in AWE scoring engines by submitting syntactically and grammatically correct gibberish text (Vojak et al., 2011).

Two other pedagogical strategies have been identified as potentially useful in the literature. AWE programs sometimes have report-generating functions that allow teachers to have an overview of student progress, as well as frequent errors made by the students. Teachers can make use of these capabilities to give students personalised feedback (Jiang et al., 2020) and using the information provided in the reports to help students set revision goals to align themselves with the course objectives and assessment criteria (Tang & Rich, 2017).

Another promising approach to integrating AWE programs into the classroom is to turn the programs' weaknesses into learning opportunities. Some teachers have succeeded in activating students' metalinguistic awareness by teaching them how to critically analyse the AWE feedback

they receive (Link et al., 2014) or having them exercise critical thinking skills by trying to ‘game’ the software to learn how it generates its feedback and scores (Z. Li, 2021).

In summary, even though the role of teachers in the implementation of AWE programs in the classroom is still not very well understood, and research on the subject has been scarce, teacher attitudes and pedagogical approaches seem to be some of the most important factors in students’ adoption and use of AWE programs for revision and development of their writing skills (Jiang et al., 2020; Tang & Rich, 2017). While this section has outlined some of the potentially useful pedagogical strategies that have been identified by current research into AWE, this is an area of research that needs more attention in the future, especially because of the implications it might have on teacher training for the use of AWE programs in the classroom.

### **2.3.5 Summary**

This section has outlined the most recent directions taken by research into the use of AWE programs as tools for formative feedback. The first focus of this review was on engagement and what research can tell us about the ways in which students engage behaviourally, cognitively, and affectively with AWE feedback. Some discussion was presented as to how these three dimensions of engagement affect student uptake of feedback, their use of strategies to engage with the feedback they receive, and how these constructs can be used to inform AWE implementation in classrooms. The second focus of this review has been on the key role of teachers within this process, and the pedagogical strategies that research has identified as being potentially useful when integrating AWE programs as part of specific pedagogical practices.

While some research continues to be carried on to explore whether AWE programs can result in increases in various measures of student writing proficiency (Parra & Calero, 2019; S. Wang & Li, 2019), the focus of this literature review has been on the integration of AWE programs into classrooms and the strategies that help students make use of these tools to improve their writing and revision skills. This is in line with the conclusion presented in the systematic literature review

included at the beginning of this chapter, which argued that research into AWE needs to focus on theoretical frameworks that can justify and inform implementation of AWE programs in classrooms. Given that these programs are already in use in EFL/ESL classrooms around the world (Stevenson & Phakiti, 2019), the question is not necessarily whether these programs should be used, but *how* they should be used to maximise their benefits to students (Jiang & Yu, 2020).

In the next section, I build on these topics to suggest some other theoretical constructs that might be of use when exploring the implementation of AWE programs as tools for formative feedback which have not received much attention in current AWE research.

## **2.4 Theoretical constructs**

In the paper included in the first part of the literature review, I argued that it was important to study how students use self-regulation strategies to engage with the feedback they received and pointed out that little research had been done on this subject up until that point. In another paper, I also argued that students display several metacognitive skills when engaging with AWE feedback, but they needed previous knowledge on which to apply these skills (Hibert, 2019a). As mentioned in a previous section, self-regulation is important when studying cognitive engagement with AWE feedback because cognitive and metacognitive strategies are an important part of cognitive engagement (Fredricks et al., 2004; Wolters & Taylor, 2012). However, while several studies touch on the cognitive and metacognitive skills used by students when engaging with AWE feedback (Cotos et al., 2020; Jiang & Yu, 2020; Koltovskaia, 2020; Zhang & Hyland, 2018; Zhang, 2020, among others), to my knowledge none of them specifically use self-regulation as a framework to explore how students use AWE programs in university and postgraduate settings.

Another conclusion I reached in my systematic literature review was that research into the use of AWE programs should take advantage of both the data-collecting capabilities of AWE programs themselves, and of available natural language processing (NLP) tools which are currently being used to study the development of language skills in ESL and EFL contexts. NLP tools such as Coh-Metrix,

TAALES, TAACO, and others have increasingly been used to understand the development of writing proficiency in ESL and EFL students. In particular, there is a growing body of research on how linguistic markers can be used to understand language development and proficiency (Jung et al., 2019), but these concepts have not yet been used in AWE research.

Therefore, this section explores both self-regulation and linguistic markers of writing proficiency and how they can inform the study of AWE programs as tools for formative feedback. First, I will offer a brief overview of current research on written corrective feedback as applied to L2 language learning to frame further discussion into the use of AWE-generated feedback. Then, I will discuss how self-regulation as a concept can mediate between the ideas of technology adoption and engagement. As discussed in the previous section, both concepts are useful in understanding how students use AWE programs. Then, I briefly discuss what current research says about the ways in which the analysis of linguistic markers can inform our understanding of writing proficiency and language level in postgraduate ESL and EFL students. Finally, I discuss how these two theoretical constructs can be integrated into the study of AWE programs as tools to enable the writing process by mapping them into the process and product approaches to teaching writing and framing them around a discussion of feedback provision.

#### **2.4.1 Feedback, learning, and the acquisition of writing skills**

So far, I have been discussing the use of AWE programs for the provision of formative feedback and the different directions that have been taken by current research into the topic without considering feedback as a theoretical construct. This is because AWE research itself has not often delved into what feedback means or what current literature tells us about how feedback is given and received, assuming students should receive some form of written corrective feedback to improve their writing skills. The question in current AWE research has not been whether feedback should be given or the forms feedback should take to be effective, but a question of whether the feedback specifically given by AWE programs can help students develop their writing skills (Stevenson & Phakiti, 2019).



However, any discussion of automated feedback needs to be framed by an understanding of what feedback is and what its role might be in teaching L2 writing. While current literature seems to agree that feedback is indeed useful for students' learning development (Hattie & Timperley, 2007), some have argued that this is not necessarily the case for helping students develop their writing skills (Truscott, 1996, 2007). I will first examine the provision and reception of feedback in general terms before narrowing the scope down to feedback in the context of second language writing, and finally will include some considerations that should be considered when discussing the automation of feedback.

**2.4.1.1 The role of feedback in education.** Hattie and Timperley (2007) conceptualise feedback as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (p. 81). Feedback is so important in education that several meta-analyses on the impact of feedback have shown that it has the highest effects on learning, even though the intensity of those effects varies depending on several factors such as the type of feedback, how it is delivered, when it is given, etc. (Brookhart, 2008; Hattie & Gan, 2015). The variable effects of feedback might also be related to the fact that, while teachers commonly provide their students with varying forms of feedback, students report that little of that feedback helps with their learning (Brooks et al., 2019), and many of them outright ignore it (Sutton, 2009).

To help understand the varying effects of feedback on student learning and help improve our understanding of the subject, Hattie and Timperley (2007) proposed a model that distinguished between four levels of feedback: task, process, self-regulation and self. The task level refers to feedback that immediately relates to the task at hand, such as pointing out right or wrong answers. The process level refers to feedback that gives the students the information they need to transfer the knowledge to different contexts. The self-regulation level refers to feedback that addresses "the way students monitor, direct and regulate actions toward the learning goal" (p. 93). Finally, the

feedback about the self as a person is not necessarily related to the task, and usually consists of utterances such as “good work!”.

Brooks et al. (2019) suggest that it should clarify expectations and standards, be targeted and ongoing through the learning period, promote self-regulated learning practices, and provide opportunities for implementing the feedback. This can be summed up by the three questions Hattie and Timperley (2007) posit good feedback should answer: Where am I going? How am I going? Where to next? Unless students use the feedback they receive to take some action, also known as ‘closing the feedback loop’ (Nicol & MacFarlane-Dick, 2006), feedback merely becomes ‘dangling data’ that does not necessarily help with student learning (Boud & Molloy, 2013).

For the process of closing the feedback loop to be achieved, researchers suggest that feedback should be scaffolded and provided through a dialogic conversation that involves both students and teachers as active participants (Hyland & Hyland, 2006; Sutton, 2009), and helps students use various sources to make sense of the information they receive and improve the quality of their work (Boud & Molloy, 2013; Pardo, 2018). This feeding forward promoting student involvement falls under the third level of feedback as suggested by Hattie & Timperley, that which promotes self-regulation in students, and is correlated with greater effectiveness of feedback (Hattie & Timperley, 2007). The effects of self-regulation on student learning and revision are discussed in more depth in section 2.4.2.

**2.4.1.2 Written Corrective Feedback and Second Language Acquisition.** So far, I have discussed feedback as pertains to education in general. However, the domain of second language acquisition, and specifically the acquisition of L2 writing skills, brings with itself several unique challenges and theoretical perspectives that need to be discussed in more detail before delving into the theoretical considerations of AWE feedback in particular.

Studies on SLA acquisition usually focus on two types of feedback: oral feedback and written feedback. Both of these kinds of feedback can be provided by teachers and peers using a variety of

in-person and asynchronous methods, or generated automatically by computer programs (Hyland & Hyland, 2019). For the purposes of this thesis, I will be focusing on written feedback on student errors, also known as written corrective feedback (WCF) (Ferris, 2019).

Ellis (2009) proposed a typology of CF based on the strategies used for providing it: direct CF, indirect CF, metalinguistic CF, focused/unfocused CF, electronic feedback, and reformulation. Direct feedback happens when the teacher provides the correct form to the student, while indirect feedback lets students know that an error exists but does not provide explicit corrections.

Metalinguistic feedback, as the name suggests, provides metalinguistic clues to point students to the nature of their errors. Focused and unfocused feedback refer to whether the teacher attempts to correct all errors or focuses on some specific types of errors. With electronic feedback, the teacher not only indicates an error has occurred but provides a hyperlink to examples of correct usage. Finally, in reformulation, a native speaker reworks the students' entire text to make it sound more native-like.

When it comes to SLA, it seems that most students value receiving WCF and, in fact, often express a desire to receive feedback in response to their written texts (Ferris, 2004; Hyland & Hyland, 2019), although research on the topic has been quite controversial. Questions remain about how WCF should be provided, by whom, and what students are expected to do with it once they receive it.

Historically, the main point of contention in the field of learning L2 writing has revolved around the issue of whether written corrective feedback (WCF) can really help students acquire writing skills. In his 1996 paper, "The Case Against Grammar Correction in L2 Writing Classes", Truscott set the stage for a vigorous debate between proponents and detractors of WCF as to the usefulness of this type of feedback in second language acquisition. Truscott (1996) argued that current research did not support the idea that WCF could be useful and, indeed, some research suggested it might actually be harmful to the acquisition of writing skills, and that there were little theoretical or practical reasons to believe WCF might be useful in the first place.

This prompted a series of responses (Bitchener & Knoch, 2009; Ellis, 1998; Ferris, 1999, 2004) and spurred a renewed interest in WCF research that aimed to provide the empirical base needed to continue the debate (Bitchener & Knoch, 2010; Ferris, 2004, 2019). Previous research on the use of WCF had been deemed lacking for several reasons. First of all, most of the existing research up until that point did not include a control group to provide evidence of the effectiveness of CF vs no feedback (Ferris, 2004; Truscott, 1996). Second, a lot of the research had design issues, such as measuring only revision and not transfer into new texts, use of different instruments in pre and post-tests, lack of pre-tests, and short-term studies that did not study the long-term impact of WCF (Bitchener, 2008).

New research into WCF has focused not only addressing these flaws to study the effectiveness of WCF and questions of how and why WCF helps students acquire target linguistic forms (Ferris, 2019), but also the ways in which students engage with this feedback and the role of automated feedback in the acquisition of L2 writing skills, although there are still many questions regarding the usefulness of different types of WCF, how it should be implemented, and how students engage with it (Hyland & Hyland, 2019).

However, this recent research has also shown that WCF is effective in helping L2 students improve their writing skills, although the effects of WCF may vary with context and research on this topic is still inconclusive (Hyland & Hyland, 2019). These contextual factors might include the time students have available to revise their texts, level of confidence in prior language instruction, levels of motivation, and instructional methods, among others (Ferris, 2019). The effectiveness of feedback may also depend on when and how it is delivered, and the type of feedback strategy that is used (Kang & Han, 2015). For example, while focused and unfocused feedback both seem to help students develop writing skills (Bonilla López et al., 2018), focused feedback has sometimes shown to have larger effect sizes (Ellis et al., 2008; Kang & Han, 2015), and some students seem to prefer to receive feedback focused on specific target forms rather than unfocused feedback (Bitchener, 2008).

From this review of the research, it can be concluded that WCF is generally helpful in improving L2 students' writing skills, but this is contingent on several factors, including the type of feedback strategy used. This is important when discussing the effects of AWE feedback on the acquisition of writing skills, since AWE programs provide very specific types of feedback that respond to the algorithms with which they have been programmed. It is therefore important to discuss how this theoretical grounding can help us understand the nature of AWE feedback, and thus its advantages and limitations.

**2.4.1.3 Feedback provision and AWE.** Following Hattie and Timperley's (2007) model, AWE feedback would fall under task-level feedback, as it only relates to the task at hand, revising their texts, and only gives them feedback about their performance in the form of scores and suggestions for revision. A problem with this type of feedback is that it often doesn't generalise to other contexts, as being able to revise one text does not mean that students have developed the metacognitive skills to notice the same errors in other contexts (Hyland & Hyland, 2006).

Even though some AWE programs provide the option for receiving metalinguistic explanations of the error codes given, an important point of contention in AWE research is precisely whether students receiving AWE feedback can learn to transfer the surface-level knowledge they gain about revising writing errors. There is little evidence that receiving AWE feedback by itself can transfer into long-term improvements in writing proficiency (Stevenson & Phakiti, 2014). Li et al. (2017), for example, found that only one of the nine categories of grammar they studied showed a long-term gain, although other research has found no improvement in subsequent drafts that did not use AWE feedback (see Wilson & Roscoe, 2020).

AWE feedback is also unfocused, as AWE programs give feedback on every type of writing feature it has been programmed to, including grammar, spelling, style, content, organisation, etc. (Stevenson & Phakiti, 2019). As mentioned in the previous section, evidence as to the effectiveness of focused vs. unfocused feedback remains mostly inconclusive. Some authors have expressed concern about

the cognitive load of attending to unfocused feedback, but research has shown that this is not necessarily the case (Bonilla López et al., 2018). With regards to AWE feedback in particular, some concerns have been raised about the cognitive load of evaluating the accuracy of feedback, although research on this topic appears inconclusive (Ranalli, 2018; Ranalli et al., 2017). To my knowledge, there has been no research focusing on the intersection of these two concerns on students' ability to attend to the feedback they receive.

Another important concern about the way in which AWE programs deliver feedback is its impersonal nature. In order to be effective, feedback needs to take into account where the student's current status and the distance between it and their goals (Hattie & Gan, 2015). AWE programs, however, are programmed to target errors and therefore make no distinctions in student levels or target specific student needs. This characteristic perhaps can somewhat help to explain conflicting results in AWE research: Barrot (2021), Koltovskaia (2020), for example, suggest that AWE feedback might be more useful to students with high proficiency levels, while Chen and Cheng (2008) and Hockly (2019) argue that AWE programs are more useful to lower proficiency students. Furthermore, Jian (2020) found that, while higher proficiency students had better access to strategies that might help them appropriate AWE feedback, it was not a guarantee that they would actually incorporate the feedback into their revision.

Feedback is only useful insofar as it is used by the students and feeds forward into their learning process to help them develop their skills (Brooks et al., 2019), and should be part of an ongoing dialogical process between students and teachers (Boud & Molloy, 2013; Carless & Winstone, 2020). An important concern about the way AWE programs provide feedback is that it does not allow this dialogic process to happen, as students cannot ask follow-up questions or ask for clarification from the program (Cheville, 2004). Several studies on AWE (C. Chen & Cheng, 2006; Jiang et al., 2020; J. Li et al., 2015; Roscoe et al., 2017) have mentioned this lack of dialogue a source of frustration for the students, who were not sure what to do with feedback they did not understand.

Therefore, from a purely theoretical point of view, there are many issues with AWE feedback that might hinder its ability to effectively help students develop their writing skills. While our understanding of feedback seems to underscore the importance of its role in education (Hattie & Timperley, 2007) and current research in SLA suggests that WCF can help students improve their L2 proficiency (Ferris, 2019), research on the usefulness of AWE remains inconclusive (Hibert, 2019b; Stevenson & Phakiti, 2019).

Feedback should not only help students at the task level, but should help them understand how the feedback can transfer to other tasks and must help them develop the self-regulation skills necessary to monitor their engagement with learning tasks (Hattie & Gan, 2015). However, the surface-level feedback provided by AWE programs by itself might not help students improve their writing skills. Therefore, it seems important to understand the intersection between self-regulation and technology adoption, and how technologies can be used to foster self-regulation in students. The lessons learned from broader research into the subject might help us better understand how AWE technologies could be integrated into the classroom.

#### **2.4.2 Self-regulation, engagement, and technology adoption**

Self-regulation involves forms of learning that are strategic and make use of metacognitive strategies (Winne & Perry, 2000), and has been defined as the “active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behaviour, guided and constrained by their goals and the contextual features in the environment” (Pintrich, 2000, p. 453). Self-regulated learning theories seek to understand how cognitive, motivational, and contextual factors can affect and influence learning (J. A. Greene & Azevedo, 2007). The idea that self-regulation is a key component of learning writing skills is not new. Zimmerman and Kitsantas (1999) mentioned that, given that writing is usually carried out as a solitary activity, writers need to make use of several self-regulatory strategies to enhance their acquisition of said skill.

In broader terms, there are several frameworks that have been put forward to explain how students self-regulate their learning, of which the most notable and widely cited are probably those by Zimmerman, Boekaerts, Winne and Hadwin, and Pintrich (Panadero, 2017).

Of these four frameworks, the one used to inform this thesis is that of Winne and Hadwin (1998), as their model that focuses the most on the role of metacognition, which has been identified as key in the study of AWE implementation, and is a model that has been widely used in research involving educational technology (Panadero, 2017). According to this model, there are four phases of self-regulation: task definition, goal setting and planning, engagement, and large-scale adaptation. (Winne, 2011). Task definition includes generating an understanding of the task to be performed. In the case of AWE research, this would include understanding the revision task and the operations that need to be done in order to successfully revise a text. During goal setting and planning, students generate goals and form a plan to achieve them. Student goals can be aligned with the assessment criteria teachers have put forth to evaluate the texts students are revising, although not necessarily (Jiang & Yu, 2020) For example, student goals might be simply to get a good grade or pass the class rather than improve their writing skills. In the third phase, students enact their study tactics and strategies to reach the goals they have set for themselves, checking the quality of the products of the study session against the standards established in the second phase. It is in this phase that the idea of engagement comes in, as students can use the tools and feedback provided by the AWE program to revise their texts and align them with their goals. Finally, once the main processes are completed, the student evaluates their performance and makes long-term changes in their motivations and strategies. This process is not linear, as metacognitive monitoring by the students in later processes might be used to inform and revise previous processes, such as the conditions of the task, the goals set for completing it, etc. (J. A. Greene & Azevedo, 2007).

Mediating these tasks are what Winne and Hadwin called the COPES processes (Winne & Hadwin, 1998): conditions, operations, products, evaluations, and standards. These processes arise from



information that students use or generate during the learning process, and help on the completion of the tasks at each phase of self-regulation (J. A. Greene & Azevedo, 2007). These processes can be found not only in learning, but also in revision activities. In this case, the conditions, or resources available, include the feedback given by the AWE program, tools for acquiring metalinguistic awareness, as well as previous knowledge of language, among other things. The operations are the cognitive processes which are carried out by the students as they engage with the text they have written and the feedback they've received. The products are the revised texts, which are then evaluated by the students against the standards set by the teachers or the AWE program itself in the assessment criteria.

This general overview can allow us to understand the different processes that are involved in the self-regulation of learning and how they can be applied to research into AWE. The concept of engagement has been mentioned as something that happens in the third phase of self-regulation according to Winne and Hadwin's model (Winne, 2011), but is affected by the conditions and products of the other phases through the use of students' metacognitive monitoring (J. A. Greene & Azevedo, 2007). However, Winne and Hadwin's model does not provide a comprehensive definition of what engagement means and does not take into account existing theoretical research on student engagement. Given that both of self-regulation and engagement are used to understand student performance in learning tasks (Wolters & Taylor, 2012), it is important to examine how these two theoretical frameworks intersect and how they can inform our understanding of technology adoption, especially when it comes to the adoption of AWE feedback.

**2.4.1.1 Engagement and self-regulated learning.** The concepts of self-regulation and engagement are quite closely linked and, as we have seen, some models of self-regulation include engagement as one of its key phases (Winne, 2011), while some models of student engagement identify self-regulation as one of the components of engagement (Fredricks et al., 2004; Lichtinger & Kaplan, 2011). The main difference between these two concepts is that, while the concept of student

engagement highlights the importance of metacognition, it does not place emphasis on the purposeful agency of students, as one can be engaged with a task without self-regulating (Wolters & Taylor, 2012). This student agency is key when talking about the use of educational technologies, as it is one of the determining factors in whether students will adopt and use a specific technology or not (Winne, 2006).

The most commonly mentioned concepts related to how self-regulation manifests itself as engagement in current AWE research are metacognition and cognition, as the employment of these strategies are a key factor of cognitive engagement that leads to behavioural engagement (Fredricks et al., 2004). However, AWE research tends to be vague as to what these strategies actually are, with the exception of Zhang & Hyland (2018) and Koltovskaia (2020). Both of these authors identify goal setting and planning as especially relevant, because these activities are considered to be part of what defines students who are cognitively engaged (Wolters & Taylor, 2012), and research has shown that setting clear goals is positively correlated to student engagement and academic achievement (Collaço, 2017; Lichtinger & Kaplan, 2011). Furthermore, students who set themselves clear goals to improve writing accuracy and proficiency tended to appropriate AWE feedback with more regularity than students who didn't (Jiang & Yu, 2020).

In recent years, research has focused on self-regulation not as a series of traits but as events, fuelled by the technological advances which make it possible to record student behaviour while performing certain activities (Bannert et al., 2014). This is especially relevant to the study of AWE programs, as many of them provide some trace data, in the form of student reports, number of submissions, before and after comparisons of drafts, among others, that can be used to study self-regulatory strategies in students. These same trace data have also been identified as important to capture engagement, given that traditional research involving self-reports often display a gap between what students report they do and what they actually do (Azevedo, 2015) since self-reports do not allow us

to measure how students employ study tactics or how they adapt them to specific contexts (Hadwin et al., 2007).

Bannert et al. (2014), for example, used process mining techniques to analyse individual learning sessions by using think-aloud data. Their coding scheme included seven metacognitive processes (orientation, goal specification, planning, searching, judgements of relevance, evaluation, and monitoring), four cognitive processes (reading, repeating, elaborating, and organising) and the element of motivation. Their findings showed that more successful students showed more learning and regulation events, and that the patterns of self-regulation examined by their model corresponded well with theoretical models of self-regulation, demonstrating that using trace data is a viable way to analyse cognitive engagement.

The field of learning analytics has long used trace data about student interaction with educational technologies to research how students self-regulate their learning (Gašević et al., 2015; Saint et al., 2020; Winne, 2017). However, to my knowledge, no current studies into the use of AWE programs have used trace data to analyse student interactions with these programs. Although some of the current research into AWE has looked at number of submissions as a proxy for behavioural engagement (Cotos et al., 2020; Jiang & Yu, 2020; Z. Li, 2021; Waer, 2021; Zhang, 2017; Zhang & Hyland, 2018), and others have looked at how much AWE feedback students use in their revision (Bai & Hu, 2017; Chodorow et al., 2010; Koltovskaia, 2020; Link et al., 2020; Tian & Zhou, 2020), these measures only give a very partial and incomplete view of student behavioural and cognitive engagement. The use of trace data can be used to investigate how students engage with learning or, in this case, revising tasks (Hadwin et al., 2007; Jovanović et al., 2017) in a more fine-grained manner.

**2.4.1.2 SRL and the use of technology.** The role of self-regulation in technology adoption has also been a focus of research in several areas, as the quality of tool adoption seems to be related to the use of metacognitive strategies (Clarebout et al., 2013). Winne (2006) argued for the importance of

self-regulation in the use of technology by pointing out that using tools in learning contexts required at least four considerations. First, students must recognise a situation in which a tool can be used. Second, then they must be able to decide which tool will 'work' best for that specific situation. Third, they must be able to use the tool skilfully, and finally they must have the motivation not only to use it, but to self-regulate their use. It is evident how these considerations are relevant in the context of AWE usage, as not only do students need to recognise how AWE programs can help in the revision process, but they also need to be able to use the feedback skilfully and monitor their progress in order to carry out their revisions. Previous sections of this literature review which explored research on cognitive and behavioural engagement with AWE programs discussed some of the 'skills' needed to successfully use AWE programs.

Furthermore, metacognitive skills greatly influence learner use of technological tools, as the awareness of their own problems determines whether they will be inclined to use tools to help solve their problems (Clarebout et al., 2013). Some skills, such as problem solving, decision making and reflection have been identified as important for the voluntary adoption of technology outside the classroom (Gurbin, 2015), which is relevant to the use of AWE programs because they are usually used in the students' own time and often outside the classroom. Beyond the use of metacognitive strategies, cognitive strategies have also been found to be an important part of technology adoption, and have been specifically linked to adoption of AWE programs (Zhai & Ma, 2021).

The relationship between self-regulated learning and technology works both ways. While self-regulation is an important component in the adoption of new technologies, as was mentioned in the previous section, some of these technologies also generate trace data that has been increasingly used to study how students self-regulate their learning and the actions that they take while engaged in studying activities (Fan et al., 2021; Saint et al., 2020). Therefore, while the concept of self-regulation can help us understand how students use technology, technology can also give us the tools to study how students self-regulate.

**2.4.1.3 Lessons for AWE research.** In the paper presented at the EC-TEL conference, which was reproduced in full at the start of this chapter, I argued for the use of self-regulation as a framework to understand student interactions with AWE programs. In this section, I have expanded on this idea by discussing how self-regulation can be used to help study student engagement, as well as student use of software technologies for educational purposes.

In summary, the theoretical framework of self-regulated learning is intrinsically linked to the theoretical framework of engagement, since students who self-regulate use a series of cognitive and metacognitive strategies to engage with learning tasks. The use of think-aloud data can help us understand the cognitive processes used by students (Bannert et al., 2014), while trace data recorded by programs used by students can give us an insight into their behaviour through the analysis of observable interactions (Winne, 2017). By combining both of these approaches, research into the use of AWE programs can go beyond proxy measures such as error reductions, uptake rates, and holistic scores, and gain a fine-grained understanding of student's cognitive processes as they interact with AWE programs. This, in turn, helps us understand how and why students decide which AWE feedback they integrate into their texts, as well as the process of revision they undertake when using AWE programs.

The research on how students self-regulate their use of technology and the implications for the acceptance and uptake of these technologies can also provide us with a framework to understand which skills need to be developed in students so they can use AWE feedback to effectively revise their texts. The use of the Technology Acceptance Model has given some insights on the factors that affect student use of AWE feedback in their revision process (R. Li et al., 2019; Zhai & Ma, 2021), and the literature on how self-regulation plays a part in technology adoption and usage can give further insights into how to integrate the use of AWE programs into classrooms to maximise their effectiveness.

### **2.4.3 Linguistic competence and natural language processing**

The term 'communicative competence' has been used for the past fifty years as a theoretical framework to explain and understand how language is taught and learned. The most commonly used model, that of Canale and Swain (1980), comprises of three main competences: grammatical or linguistic competence, sociolinguistic competence and strategic competence. Linguistic competence includes knowledge of the rules of morphology, syntax, semantics and phonology. Sociolinguistic competence includes the knowledge of sociocultural rules that allow speakers to produce utterances appropriate to the context. Strategic competence refers to the strategies that are used by speakers when communication breaks down, such as paraphrasing and coping strategies. Celce-Murcia et al. (1995) updated Canale and Swain's (1980) and then again in Celce-Murcia (2007) by proposing six components of communicative competence: socio-cultural competence, discourse competence, linguistic competence, formulaic competence, interactional competence and strategic competence.

Linguistic competence refers to "the mastery of the language code itself" (Pillar, 2011, p. 6) and includes phonological, lexical, morphological and syntactic elements of language (Celce-Murcia, 2007). For the purposes of this thesis, our focus will be on linguistic competence, given that AWE programs mostly provide feedback related to this domain, although some programs also provide feedback related to formulaic competence, especially programs like Pigai that focus on L2 learners and therefore offer feedback on collocations (X. Yang & Dai, 2015).

While communicative approaches to teaching languages focus more on getting meaning across (Celce-Murcia, 2007), knowledge of grammar is important to provide learners "with the knowledge of how to determine and express accurately the literal meaning of utterances" (Canale & Swain, 1980, p. 30), and therefore teaching for communication should also include grammatical elements (Savignon, 2017). Therefore, it makes sense to examine this dimension independently as it relates to the feedback provided by AWE programs and how we can evaluate the results of receiving this feedback.

One of the main issues derived from our understanding of communicative competence, especially when it comes to teaching a second language for academic purposes, is a set of objective criteria that can be observed and tested to understand the level of proficiency which the student has attained (Bachman, 1991; Lehmann, 2007). While “native-like” proficiency is often considered to be the end goal of second language learning, the term itself is problematic to define (Canale & Swain, 1980) given that the term itself is arbitrary and often involves issues of power between different groups of speakers (Savignon, 2005).

There have been several attempts to measure communicative competence and language proficiency in general, with several frameworks and criteria for testing ESL and EFL being proposed over the years (Bachman, 1991; Canale & Swain, 1980; Savignon, 2005). In recent years, there has been an increased interest in how advances in natural language processing can help us understand language development in ESL and EFL students and how we can measure language proficiency using these technologies (Jung et al., 2019). In the last ten years, research has focused on the features of linguistic competence such as lexical and the syntactical features that can help understand L2 development and act as predictors of writing proficiency (Casal & Lee, 2019; Jung et al., 2019; Ryu, 2020). Whereas, in the past, research into these features would be done manually, recent advances in NLP have allowed us to compute them more quickly and reliably (M. Kim & Crossley, 2018). Most of the research on this topic consists of using certain linguistic features and creating regression models with the features that correlate significantly with measures of writing proficiency, most commonly expert ratings, either through holistic scores or analytic rubrics (Matthews & Wijeyewardene, 2018; Yoon, 2018).

In this context, writing proficiency can be understood as the capacity to produce writing that uses both syntactic and lexical aspects of language (Ha, 2019). Some definitions also include text cohesion under the umbrella of writing proficiency (M. Kim & Crossley, 2018), however research on whether automated indices of cohesion can help us understand the development of writing proficiency in L2

settings is mixed at best, and no consensus has been reached on the matter (M. Kim & Crossley, 2018; Matthews & Wijeyewardene, 2018; Ryu, 2020). Furthermore, there is evidence to suggest that more proficient students produce texts that are more linguistically sophisticated, rather than more cohesive (Crossley & McNamara, 2012).

There are other aspects of writing proficiency, such as discourse knowledge and sociolinguistic knowledge (Yoon, 2018), which are important to consider but, since these are not addressed by AWE programs, they are beyond the scope of this thesis. Linguistic features are also being used to generate scores in Automatic Essay Scoring engines (Jung et al., 2019), which power some of the scoring and evaluation features behind AWE programs (Hockly, 2019). Therefore, it is worthwhile to consider the syntactic and linguistic aspects of writing proficiency separately and explore what recent research tells us about how lexical and syntactic features can help us understand the development of language and writing proficiency.

First, I describe the several tools that have been developed for the automated analysis of lexical, syntactic and coherence analysis of L2 texts. Then, I briefly discuss the lexical and syntactical features that have been found to correlate to human judgments of writing quality and proficiency, followed by a summary and discussion of how these can be applied to the study of AWE programs as tools for formative feedback.

**2.4.2.1 NLP tools for linguistic analysis.** Several tools have been developed in recent years by researchers to analyse different linguistic aspects of texts. These tools are based on recent advances in computational linguistics, discourse processes, cognitive science, and corpus linguistics, among others (McNamara et al., 2014). Corpus linguistics in particular have long been used in the domain of ESL and EFL teaching, especially through data-driven learning that uses language corpora as tools to help students develop language knowledge and increase their awareness of language patterns (Luo & Zhou, 2017). The same principle is applied to the design of these new NLP tools, which use



linguistic databases and language corpora to train algorithms that analyse texts for specific linguistic traits.

Of these, the ones that have the greatest theoretical backing and are most widely used for research include Coh-Metrix, developed by Graesser, McNamara, Louwerse & Cai (2004), the Tool for the Automatic Analysis of Lexical Sophistication (TAALES) developed by Kyle et al. (2018), the Tool for the Automatic Analysis of Cohesion (TAACO), developed by Crossley, Kyle & Dascalu (2018), and the Syntactic Complexity Analyser (SCA) developed by Lu (2010).

Coh-Metrix was originally developed to analyse 50 types of cohesion relations and over 200 linguistic measures, based on the premise that cohesion is an objective property of texts (Graesser et al., 2004). Ever since its inception, Coh-Metrix has been one of the most widely used tools for analysing the linguistic factors that correlate to writing quality and proficiency (Aryadoust & Liu, 2015; Crossley et al., 2016a; Crossley & McNamara, 2012; Guo et al., 2013; Jung et al., 2019). TAACO was designed as a spiritual successor to Coh-Metrix, designed to be easier to use and address more indices of local, global and overall text cohesion than its predecessor (Crossley et al., 2019).

The Syntactic Complexity Analyser (SCA) incorporated fourteen syntactic measures derived from previous research done into the field of syntactic complexity (Lu, 2010). These fourteen measures were divided into four groups: length of production unit, sentence complexity, subordination, coordination, and particular structures. The purpose of this program was to provide a tool for the analysis of syntactic complexity for research, and also as a way for teachers to compare the syntactic complexity of writing samples between students or of particular students for assessment purposes.

Finally, TAALES was designed to provide a freely available tool that would help researchers to calculate a variety of indices of lexical sophistication (Kyle & Crossley, 2015), and was later updated with more indices related to word recognition norms, contextual distinctiveness, word neighbourhood, among others (Kyle et al., 2018). The current version includes more than 400 indices of lexical sophistication, as well as some index diagnostics.

These four programs have been extensively used in research into linguistic features that predict writing proficiency and quality, both for lexical and syntactical elements of language. In the next two sections, I discuss each element separately and present a short summary of existing research in this area.

**2.4.2.2 Lexical measures of L2 writing proficiency.** Analysing how the lexical features of a text can help us understand the development of language proficiency has become an increasingly growing field in the last few years. Current research has shown that there is a close correlation between having a broad vocabulary and being able to use it as a strong indicator of language proficiency (Ha, 2019; Ryu, 2020), and lexical characteristics have shown to be good predictors of assessment of writing quality (Jung et al., 2019; Matthews & Wijeyewardene, 2018; Ryu, 2020).

Lexical sophistication can be defined as “the production of advanced and difficult words” (Crossley & Kyle, 2018, p. 48), and is considered to be a multidimensional construct which encompasses several measures related to the depth and breadth of vocabulary knowledge (Yoon, 2018). The most traditional method of evaluating lexical sophistication has been word frequency, which calculates how often words occur in general usage, although in recent years other lexical features have been found to better explain lexical knowledge and development (Kyle et al., 2018). For example, lexical sophistication is also related to words that are less contextually diverse, less concrete, with less imageability and familiarity, as well as words that have fewer orthographical and phonological neighbours (Crossley & Kyle, 2018).

There are many lexical indices that have been used for analysing L2 writing proficiency and quality. A popular tool for analysing lexical sophistication, TAALES, measures over 400 indices of lexical sophistication (Kyle et al., 2018). Crossley et al. (2011) and Kyle & Crossley (2015) provide a detailed explanation of many of the lexical sophistication indices that have been used in research. Describing and analysing each of these indices is beyond the scope of this thesis, although only a small number of indices have been found to correlate to writing proficiency and quality. In this thesis, I consider

those indices which have been found to correlate to writing proficiency and quality in more than four separate studies. These indices include lexical diversity, word range, word familiarity, word frequency, word meaningfulness, word hypernymy, and word imageability.

Lexical diversity refers to the number of different words that appear in a text (Bestgen, 2017), and is one of the main measurements used in automated scoring engines (Choi & Jeong, 2016) that can be indicative of writing quality and human judgments of lexical proficiency (Crossley et al., 2012).

Several studies have found links between employing a wide range of vocabulary and human ratings of proficiency (Aryadoust & Liu, 2015; Bulté & Housen, 2014; Crossley et al., 2011; Crossley & McNamara, 2012), and writing quality (Crossley et al., 2014).

Word range is a measure of how widely a word occurs across documents within a given corpus, and has been used as a measure of lexical sophistication as L2 speakers who are more proficient use words that occur in fewer contexts (M. Kim et al., 2018). This is another measure that has been found to correlate to human ratings of quality (Kyle & Crossley, 2015, 2016), proficiency (Kyle et al., 2018), and argumentative vocabulary scores (Yoon, 2018).

Word familiarity measures how frequently words tend to occur in text and discourse (Crossley et al., 2011) and are therefore recognised more quickly (Crossley et al., 2012). This construct has a negative correlation with proficiency, given that more proficient L2 writers tend to use less familiar words (M. Kim et al., 2018). Perhaps because of this, word familiarity has also been found to be a predictor of writing quality (Aryadoust & Liu, 2015; Guo et al., 2013; Kyle et al., 2018) and proficiency level (Crossley & McNamara, 2012; Jung et al., 2019; Ryu, 2020).

Word frequency is a measure of how frequently certain words occur in the English language, and is often considered an important aspect of lexical richness (Crossley et al., 2013). Word frequency is usually calculated from databases like the Centre for Lexical Information (CELEX), a 17.9 million word corpus (Crossley & McNamara, 2009). This index is considered to be one of the best predictors of growth of language proficiency over time, for the same reasons as word familiarity: more proficient

writers tend to use words that are less frequent in everyday usage (Berger et al., 2017). Word frequency has also been found to correlate to proficiency level (Aryadoust & Liu, 2015; Crossley et al., 2011; Hongwei & Liqin, 2013; Kyle et al., 2018; Ryu, 2020), writing quality (Crossley & McNamara, 2012; Guo et al., 2013; Jung et al., 2019; McNamara et al., 2010), and lexical sophistication (M. Kim et al., 2018).

Word meaningfulness relates to the associations a word has with other words; the higher the word meaningfulness, the more associations a word has, and therefore the earlier those words are acquired by L2 learners (Crossley et al., 2011). This measure has also been found to be negatively correlated with language proficiency (Crossley et al., 2012; Crossley & McNamara, 2012; Kyle & Crossley, 2015), and writing scores (Crossley & McNamara, 2011) in certain studies, as more proficient writers tend to use less common words with fewer associations.

Word hypernymy is based on hierarchical associations between superordinate words (hypernyms), like food, as opposed to subordinate words (hyponyms) like apple (Crossley et al., 2011). It has been correlated to language proficiency, since more advanced writers tend to use more specific words (M. Kim et al., 2018) and hypernymic relations increase as students spend time learning English (Crossley et al., 2012). This measure has also been found to be a predictor of writing quality (Crossley et al., 2011; Guo et al., 2013; Kyle & Crossley, 2016) and language proficiency (Kyle et al., 2018; Kyle & Crossley, 2016; Ryu, 2020).

Finally, word imageability measures the ease of constructing a mental image of a word in one's mind, as measured by human ratings (Graesser et al., 2004). This measure has been linked to language acquisition as more concrete words, like tree, are usually learned earlier than more abstract words like liberty (M. Kim et al., 2018), and therefore essays that use words with lower imageability tend to be produced by more proficient writers (Ryu, 2020). Like the other measures of lexical proficiency discussed so far, word imageability has also been found to correlate with writing

quality (Crossley et al., 2014; Guo et al., 2013; Kyle & Crossley, 2016; Ryu, 2020) and linguistic development (M. Kim et al., 2018).

While the list of indices presented here is not exhaustive, these indices are the ones that have most consistently been presented in the literature as being indicative of writing proficiency and correlating better with human judgments of writing quality. That being said, there is still much research to be done to understand how automated indices of lexical proficiency and sophistication can help us understand the development of writing proficiency and the quality of student writing. As a first approach, however, current research suggests these indices can be used as proxy measures for writing quality and proficiency.

**2.4.2.3 Syntactic measures of L2 writing proficiency.** Another important focus in the use of NLP to understand how language proficiency develops is that of syntactic complexity, since having the capacity to produce complex syntactic structures can help students and learners to attain their communicative and rhetorical goals (Casal & Lee, 2019). In this context, syntactic complexity can be defined as “the variety and level of sophistication of the syntactic forms that are evident within a learners’ language output” (Matthews & Wijeyewardene, 2018, p. 145). Syntactic complexity is also considered a multidimensional construct, encompassing global complexity, complexity by subordination, by subclausal or phrasal elaboration, and coordination, all of which require different measures (Lu, 2017).

Research into this topic is quite complex, as it has been found that the syntactic features developed by ESL and EFL learners are not necessarily the same features that allow them to be better evaluated with regard to their writing quality (Crossley & McNamara, 2014). Furthermore, the use of some of these syntactic measures to predict holistic scores of writing quality may depend on the genre which is being studied (Casal & Lee, 2019; Lu, 2017).

Adding to this, findings in this domain have been, at times, contradictory, and little consensus exists on which syntactic measures can be used to predict language development and writing quality (Casal

& Lee, 2019). While some studies have found that, for example, texts that have a higher mean length of clauses can correlate to human judgments of writing quality (Crossley & Kyle, 2018; W. Yang et al., 2015), longer clauses that also use sophisticated lexical items are better predictors of higher scores (M. Kim & Crossley, 2018). This may be, in part, because of the ways in which different NLP tools conceptualise syntactic complexity (Lu, 2017).

Again, making an exhaustive list of all syntactical factors that have been used to analyse L2 writing is beyond the scope of this thesis, and has been addressed by others. Kyle (2016), for example, has an extensive description of many of these factors. Lu (2017) did a review of several studies on syntactic complexity using three different programs for analysing and came up with a list of syntactic complexity measures that were found to predict or correlate to human ratings of writing quality. However, unlike the lexical measures discussed in the previous section, which have several studies backing their relationship to writing proficiency and quality, the syntactic complexity measures identified by Lu only had one or two studies which found significant correlations between these measures and writing quality. Furthermore, there is still some debate as to whether syntactic complexity and grammatical accuracy can be considered as appropriate measures of good writing (Hyland, 2003). Therefore, it was considered that, while syntactic measures might be useful in understanding how texts evolve through the use of AWE feedback, research into the topic of syntactic complexity has, to my knowledge, not achieved sufficient consensus to warrant using these measures at this time.

**2.4.2.4 Summary.** By using the aforementioned tools, research has found that some of these lexical and syntactical features correlate to human judgments of writing quality. However, while the evidence for the syntactic structures is still inconclusive, most researchers agree that lexical features are good predictors of writing quality and language proficiency (M. Kim & Crossley, 2018).

While current research has found that the correlation between these automated indices and expert ratings of quality are not strong enough for high-stakes assessment, there is value in the use of these

indices for low-stakes, formative assessment (Matthews & Wijeyewardene, 2018), which is what AWE programs are currently used for in classroom integrations. One of the main concerns about using AWE programs is that the revisions made by using their feedback tends to be quite superficial (Jiang & Yu, 2020; Stevenson & Phakiti, 2014), so most of the analysis that seeks to understand the effects of AWE programs in texts tends to focus on scores (El Ebyary & Windeatt, 2010; Huang & Renandya, 2020; X. Yang & Dai, 2015) or error reduction rates (Dikli, 2010; Ranalli, 2018; Saricaoglu & Bilki, 2021), both of which can only give a superficial understanding of how the composition of texts changes as a response to AWE feedback. Using some of the indicators mentioned in this section to evaluate the output of revision using AWE programs might help us achieve a more fine-grained understanding of the quality and magnitude of the revisions carried out by students using these programs.

#### **2.4.4 Process and product approaches to teaching writing**

So far, I have examined self-regulation as a theoretical framework that can help us understand student engagement with AWE programs and the use of NLP to examine linguistic markers of writing quality and proficiency. The first theoretical framework can help us understand the process of student interaction with AWE programs and how these programs are used both in classrooms and in the students' own time. The second theoretical framework can be used to examine the end product of revision with AWE feedback, examining how the linguistic composition of texts is affected by the revision process and therefore adding another dimension to research into how helpful AWE programs are in improving the quality of student texts. These two frameworks can be roughly mapped to the process and product approaches to teaching writing as a way to understand the role of AWE programs in the writing process.

The process approach to writing is one of the more traditional approaches of teaching writing, which focuses on the use of proper syntax and grammar through the emulation of teacher-supplied materials (Neupane, 2017). However, this perspective drew criticism as it defined writing only by its

product, ignoring the fact that texts are more than their formal characteristics and that such an emphasis gives no insights as to how to develop this final product (Kadmiry, 2021). This led educators and researchers to adopt a more process-focused view of teaching writing, which emphasised writing as a cyclical process that emphasises student freedom and the ways in which a text can be developed (Hasan & Akhand, 2010). The emphasis here is on the relevance, organisation, coherence, and originality of the texts produced (Bogolepova, 2016). The goal here is not to offer an exhaustive description of the merits of either model for teaching writing skills, but rather to provide a brief overview of what these two approaches are, and how understanding them might help contextualise research into the use of AWE programs as tools for formative feedback. While genre approaches to writing are also widely used today (Hyland, 2003), this dissertation focuses exclusively on the genre of academic writing and therefore this theory was not deemed applicable in this particular context.

**2.4.3.1 Product approach.** The product approach to teaching writing has as its main goal the production of an error-free, coherent text (Neupane, 2017). While product-centred approaches to teaching writing are not as widely used nowadays as process-oriented approaches (Hyland, 2003), most research into the use of AWE programs has implicitly or explicitly taken a product-centred stance. That is, as mentioned previously, most of AWE research in the past has focused on the finished product of the writing process, whether it is in the increasing of scores or the reduction of error rates.

Product-oriented approaches to writing focuses on the grammatical features of texts, emphasising linguistic knowledge, cohesive devices and vocabulary choices, especially in ESL and EFL writing (Hyland, 2003). This approach to teaching writing is usually based on four stages (Steele, 1992): studying model texts to learn the features of the genre, controlled practice of highlighted features, organisation of ideas, and the production of the final product. Models are primarily used as a tool to teach different formal aspects of writing such as the structure of well-written paragraphs, modes of



argument, types of discourse, etc. (Kadmiry, 2021), and the main goal of this is accuracy, clear exposition, and the mastery of lexical and syntactic forms (Hyland, 2003).

Despite widespread criticisms, product-oriented approaches to writing are still widely used in ESL & EFL instruction (Kadmiry, 2021), especially for novice writers (Hyland, 2003). There is also merit in using a product-centred approach to study the effects of AWE programs in ESL & EFL contexts. First, it is true that formal correctness is a goal for L2 learners (Hyland, 2003) and some of their biggest challenges when learning how to produce texts in their L2 language have to do with the use of vocabulary and grammar to effectively convey their ideas (Neupane, 2017). Second, one of the promises of AWE programs is that their feedback can be used to help students correct low-level and mechanical errors in their writing, thus freeing up time for teachers and peers to focus on higher-level organisational and content feedback. Therefore, it seems relevant to research whether these programs can deliver on these promises and whether they can indeed be trusted to provide students with low-level and mechanical feedback for revision that focuses on correctness and mastery of lexical and syntactical forms. As was mentioned previously in section 2.3, most of the research on AWE has taken this path, looking at either reductions in error rates or increases in scores as metrics of success for the implementation of AWE programs from a product perspective.

However, it has been pointed out that formal and linguistic aspects are not the only important parts of the writing process, and emphasis has to be made on what writers want to say, the importance of writing for an audience, and the understanding of writing as a creative process (White, 1988).

**2.4.3.2 Process approach.** The process-oriented approach to writing is heavily based on Flower and Hayes' (1981) cognitive writing model, in which writing is the product of interaction between the task environment, the writers' long-term memory, and the writing process. This approach to teaching writing emphasises the need to develop a metacognitive awareness of the process of writing to reflect on the strategies that are used by the students to write their texts (Hyland, 2003).

Theoretical approaches to process-oriented writing draw on many of the cognitive and metacognitive strategies of self-regulation that have been previously discussed in section 2.4.1. In its most basic form, this approach includes planning, translating and reviewing as three key stages of the writing process (Flower & Hayes, 1981) that occur non-linearly and are part of a recursive, interactive process (Hyland, 2003). The planning process involves setting goals, planning and organising ideas, and monitoring one's progress (Kadmiry, 2021), which are also components of self-regulation in students. Motivation, another important component of self-regulation, is also an essential component of the process writing model, as willingness to engage in writing is considered to be a requisite to writing development (Hayes, 2012). A typical writing task using the process approach consists of several steps, which include different brainstorming techniques, both teacher and peer feedback, as well as some rounds of editing (Steele, 1992).

Process approaches to writing have also drawn criticism, as it is not considered to be appropriate for all students and the emphasis on the process sometimes leaves students without explicit instruction on the structure of the target texts they are expected to write, which is an issue because students are rarely graded on the process they undertake, but rather the final product they turn in (Horowitz, 1986; Kadmiry, 2021). As a response to this criticism, Hayes (2012) developed an updated version of the process writing model which included explicit instruction of form as an important element of the process, especially for EFL and ESL writers. Another important addition of Hayes' updated model was the acknowledgment of the importance of collaboration and feedback in the writing process (2012) to help transform content and expression in this recursive process of writing (Hyland, 2003).

As mentioned in section 2.3, research into the use of AWE has begun to emphasise the importance of studying the use of AWE programs, not only for their effects on the finished product of student writing but also as an integral part of the writing process itself. The process-oriented approach to writing is based on the idea that the use of cognitive processes and strategies during the writing process is what separates good writers from poor writers (Flower & Hayes, 1981), something that

echoes the arguments I made previously for self-regulated learning, where the use of cognitive and metacognitive strategies is what separates successful learners (J. A. Greene & Azevedo, 2009).

Current AWE research which focuses on the role of these programs in the process of writing also emphasises the role of cognition and metacognition in the use of AWE programs for revision (Cotos et al., 2020; Koltovskaia, 2020; E. L. Wang et al., 2020; Zhang, 2020). The question then becomes how the use of AWE feedback can help the students in the process of writing and redrafting, rather than whether it can provide measurable improvements in the finished product. Of course, these perspectives are not opposed to each other and, in fact, can complement each other in AWE research, just as few teachers actually use only one approach to teaching writing in ESL & EFL contexts (Hyland, 2003).

Especially in the context of AWE research, it is important both to understand how the feedback provided by these technologies contributes to the process of writing and redrafting and the effect the use of these technologies have on the finished text produced by students. Therefore, rather than limiting ourselves to a single perspective when researching AWE technologies as tools for formative feedback, it might be worthwhile to broaden our scope. By using the lens of both product and process-oriented approaches to teaching writing, we can achieve a more holistic understanding of the effect AWE programs have on developing writing skills in ESL and EFL students.

## **2.5 Conclusion**

While current research into the use of AWE programs has begun taking on a more student-based approach to understand how these programs can be used as part of a process-oriented writing strategy, there is still scarce understanding on the mental processes and engagement strategies used by students when interacting with these programs. Another identified gap in the literature is the lack of concrete, research-informed pedagogical practices that can be used to integrate AWE programs into classrooms as tools for formative feedback, as discussed in section 2.3.4. This is, in part, because

there is little understanding of how students actually use AWE programs in their revision process and the strategies they employ when engaging with AWE feedback.

## Chapter 3 – Methodology

The previous chapter explored the research on the use of AWE technologies for formative feedback in ESL and EFL classrooms. This chapter begins by summarising the most important conclusions that can be drawn from the literature review, and how these conclusions have informed the research questions addressed in this thesis. Then, I explain the epistemological foundations and methodology underpinning my research, justifying the chosen methods, and the approach I used to research AWE programs.

### **3.1 Research questions and research objectives**

The systematic literature review presented in the 2019 EC-TEL conference summarised existing research and concluded that research into the use of AWE should move beyond trying to prove quantitative gains in scores or other arbitrary measures of proficiency and should instead focus on how AWE programs can help develop autonomous writers. In that paper, I also argued that the AWE research should focus on how AWE programs are adopted by the students by drawing on existing literature on technology adoption, such as the role of self-regulated learning, the social aspects of technology adoption and engagement, and using NLP to examine the effects of AWE feedback. The rest of the literature review then focused on these models and how they can inform the use of AWE programs.

As was mentioned in the previous chapter, a lot of work has been done on the effects of AWE on scores, error correction rates and student perceptions, but little research has focused on understanding how students use AWE programs (Hibert, 2019b). I also discussed how both a process and a product approach can be used to assess and teach writing skills, and I argued that there are two main gaps in our understanding of student usage of AWE programs. First, most of the research into the use of AWE has looked at the product of revision by analysing scores and reductions in error

rates, but if we are to have a full understanding of the place AWE programs might have in the classroom, then we also need to look at it from a process perspective. Second, scores and error reductions metrics tend to be superficial measures of the effect of AWE on revised texts, as they ignore other aspects of revision such as vocabulary, linguistic characteristics, and text complexity, while being more susceptible to avoidance strategies (Ranalli, 2018).

The goal of this thesis is to address these gaps in understanding, looking at both the process and the product of student interaction with AWE. From a product perspective, the focus was on exploring different measures that would allow an understanding of how the texts themselves change as a response to AWE feedback. Another conclusion of my systematic literature review was that current research into AWE has failed to explore the modes of data collection and analysis afforded by the technological aspect of AWE programs. This includes analysing student behaviours through data collected from their interactions with the AWE program and collecting data on the texts before and after the usage of the AWE program for revision. The theoretical construct used to explore this perspective and compare the texts before and after revision was that of linguistic features of texts which can be used as predictors of writing proficiency (Ryu, 2020). AWE programs themselves use these linguistic features as part of the algorithms they use to provide feedback and score texts (Dikli, 2006; Kumar & Boulanger, 2020), and research has found that some of these features tend to correlate with human judgments of writing proficiency (Jung et al., 2019). By using data from both student interaction with the AWE program and comparing the linguistic characteristics of the texts before and after receiving AWE feedback, it can be possible to have a more holistic understanding of the product of student interaction with these programs than simply looking at scores or error reduction rates.

Engagement and self-regulation were the constructs chosen to study the process of writing with AWE programs, given that effective engagement with feedback leads to improvements in writing (Zhang, 2020), and engagement allows students to exercise their agency in deciding how and what

they learn from feedback (Koltovskaia, 2020). The goal is to use these theoretical constructs to understand how students interact with the feedback they receive from AWE programs, either accepting or rejecting the feedback, and how these decision-making processes affect the quality of their feedback uptake. Understanding the strategies students use when engaging with AWE programs can help identify areas of opportunity where students can be supported in using AWE feedback to improve their writing skills.

Therefore, the research questions that inform this thesis are as follows:

1. What changes do students make to their texts as a response to AWE feedback?
  1. What revision behaviours do the students exhibit when interacting with an AWE tool?
  2. How substantial are the revisions made to texts as a response to AWE feedback?
  3. Do the changes made by the students improve the quality of the texts as measured through linguistic factors that correspond to human judgments of writing quality?
2. How do students engage with an AWE program?
  1. What, if any, self-regulatory strategies do students employ when engaging with AWE feedback?
  2. What decision-making process do students use to decide whether to accept or reject feedback given by an AWE program?
  3. Do students' decision-making processes help them make good decisions when accepting or rejecting AWE feedback?

Given that each of the research questions posited above deals with a different aspect of writing, product and process, I decided to conduct two separate studies, one to address each research question. This allowed a greater amount of flexibility in the amount and types of data that could be collected and analysed for this dissertation. Given the different approaches used for each one of

these studies, their methods, and the rationale behind the research design of each is treated separately in this chapter.

One point that needs to be mentioned is that neither study used a control group. It has often been suggested that lack of control groups means that an argument cannot be made for or against the provision of feedback (Truscott & Hsu, 2008) or the usefulness of AWE programs (Link et al., 2020). However, the reason why control groups were not used for either of the studies presented in this thesis was because the purpose of this research is not to make an argument for or against the usefulness of these programs, but rather contribute to an understanding of how students use and engage with these programs to revise their texts.

With this in mind, the next section will deal with the ontological assumptions and epistemological foundations of this research to justify the two-study research design adopted for the purposes of this thesis.

### **3.1.1 Epistemological foundations of this research**

The main objective of this research is to help understand how students use AWE programs to revise and improve their writing. To achieve this goal, I decided to use two different sources of data. First, there are certain elements that should be quantified to help understand the usage of AWE programs, such as the amount of feedback that was used by the students and some quantifiable changes in the quality of their texts which will be discussed in a future section. At the same time, however, if we wish to understand how students use AWE feedback, we also need to look at their perspectives and experiences with using the AWE programs to revise their texts to understand the reasoning behind their uptake and revision.

Given this, it was decided that a mixed methods approach would be best suited to answer the research questions and fulfil the research objectives. The usage of mixed methods research has to be understood within the context of the so-called paradigm wars of the 70s and 80s (Hall, 2013), which focused on whether mixed-methods research had a sufficient epistemological foundations or



whether quantitative and qualitative methods were fundamentally incompatible (Thierbach et al., 2020).

To simplify the debate, as a full account of it would be beyond the scope of this thesis, quantitative research is informed by a positivist paradigm that emphasised the objective reality of nature and uses mostly deductive methods of research to examine relationships between data (Ansari et al., 2016). On the other hand, qualitative research consists of a collection of approaches mostly informed by a constructivist paradigm that presupposes that there is no single objective reality and collects textual data to document human experiences (Feilzer, 2010; Saldaña et al., n.d.). Since both epistemological perspectives have been claimed to have radically different conceptions of the world, they were seen as at odds with each other, and so the notion of mixed methods research was not considered feasible on epistemological grounds (Samuel Barkin, 2015).

However, starting in the late 80s and into the 90s, mixed methods research began acquiring more prestige as a valid research methodology as a valid research approach in its own right and not just as a 'middle ground' or a 'mishmash' of quantitative and qualitative research (Bryman, 2006). Mixed methods research usually prioritises the need of any particular study rather than a strict epistemological stance from the part of the researchers, and it is now acknowledged that epistemological underpinnings may vary from project to project depending on the specific needs of that piece of research (Thierbach et al., 2020).

Several epistemological foundations have been proposed for mixed methods research, including pragmatism constructionism, critical theory, and critical realism (Pluye & Hong, 2014). The research design of both studies involved in this thesis was informed by a pragmatist epistemology, which focuses on the research questions and pairing it with the most appropriate research methods and thus has been used to underpin mixed-methods research (Hall, 2013; Harrits, 2011). The pragmatist philosophy has been characterised as an alternative paradigm that "accepts, philosophically, that there are singular and multiple realities that are open to empirical inquiry" (Feilzer, 2010, p. 8).

In the case of this thesis, I acknowledge that there is an objective reality to AWE use that can be seen from feedback uptake and other quantitative metrics of textual changes, but that this uptake is motivated by each student’s personal experiences, level of skill, goals, and other factors that can only be understood through a more qualitative approach. Therefore, both approaches will be used to explore the research questions posited above, but have been separated into two studies for clarity.

Issues of validity and reliability within this pragmatist paradigm will be further discussed in section 3.4. However, to maintain the consistency of the feedback received by the students and, therefore, help improve the validity of this research, both studies used the same AWE program. Therefore, it is important to first discuss the AWE program chosen for this study and the rationale behind this choice, before delving into the specific methodology for each of the studies and broader issues of validity and reliability.

### 3.2 The AWE program used

**3.2.3.1 Tool selection.** There are many AWE programs available commercially that have been used to research the use of AWE as tools for formative feedback. A breakdown of currently available commercial programs can be found in Table 1. The list is not exhaustive, but it represents some of the most widely used AWE programs, or programs that have been used in research into AWE feedback.

**Table 1 - Non-exhaustive list of currently available AWE programs**

Software name	Company	Cost	Feedback	Additional Tools
CorrectEnglish	CorrectEnglish	Commercial	Spelling Subject/verb agreement Context Word suggestions Punctuation Common L2 mistakes	Dictionary Thesaurus Writing Templates Translator
Criterion	ETS	Institutional licence only	Grammar Usage Mechanics	Planning tools Writer’s handbook Teacher comments

			Style and organisation Development	Peer review tools
Ginger	Ginger	Free	Grammar and spelling Rephrasing	Dictionary Text reader Translator
Legal Analysis and Writing Skills	N/A	Bespoke	Collocation	N/A
IADE	N/A	Bespoke	Rhetorical moves	N/A
MSWord <sup>1</sup>	Microsoft	Commercial	Spelling and grammar	Dictionary Thesaurus
MYAccess	Vantage Learning	Commercial	Focus and meaning Content and development Organisation Language use, voice, and style Mechanics and conventions	Goals and examples Multilingual feedback Teacher comments
Pigai	Beijing Cikuu Science and Technology Co.,Ltd	Commercial	Grammar Accuracy Coherency Collocations	Thesaurus Chinglish examples
Research Writing Tutor	N/A	Bespoke	Communicative effectiveness Writing norms	Examples of target forms
ProWritingAid	ProWritingAid	Free and paid versions	Spelling and grammar Style Coherency Overused words Sentence structure Pacing and readability	Explanations and examples of grammatical terms used in the feedback
Grammarly	Grammarly	Free and paid versions	Spelling and grammar Style Tone	Explanations and examples of grammatical terms used in the feedback

The question of which AWE program to use for the research in this thesis was an important one. The research was carried out at the University of Edinburgh, which does not use any AWE programs as part of its writing classes or at an institutional level. Therefore, of the available AWE programs listed in Table 1, some of the most commonly used ones in AWE research like Criterion, MYAccess and Pigai could not be used for this study because they require an institutional subscription to access.

---

<sup>1</sup> While Microsoft Word is not necessarily a dedicated AWE program, it has been used in research into automated feedback (Abuseileek, 2013), and so was included in this list.

Other programs, like IADE, Research Writing Tutor, were created specifically for research projects and are therefore not widely available for use.

The program chosen was ProWritingAid, as it is available to anyone and it provides a similar array of features as other AWE programs designed specifically with academic writing in mind, as can be seen in Table 1. While Grammarly is another program that is freely available and more widely used than ProWritingAid<sup>2</sup>, the latter was more useful to this research because it provides an API which allows integrating their AWE engine into other applications.

This API allowed me to create a web application that used ProWritingAid's AWE engine to give students feedback while collecting data on their usage of the program and collected copies of their texts at key moments for comparison purposes. A further explanation of how this application was programmed and how data was collected is given in the data collection section for the RQ1 study.

**3.2.3.2 Tool features.** ProWritingAid offers the same features as most AWE programs commercially available. Besides checking for grammar and spelling, the programs include several other features, which include, but are not limited to:

- **Style.** Highlights areas where style might be improved, including passive verb usage, adverb usage, usage of hidden verbs, using simpler and clearer words, and removing wordiness, among other things.
- **Thesaurus.** Shows possible replacements for nouns, verbs, adverbs, and adjectives.
- **Overused.** Looks at commonly overused words from the program's database of training texts and checks if the currently revised text contains any of them.
- **All repeats.** Checks the text for commonly repeated words and phrases.
- **Echoes.** Checks for words and phrases that are repeated in close proximity within the text.

---

<sup>2</sup> As per their reported number of users in their respective websites

- **Length.** Shows a visual representation of the lengths of the sentences used in the text.
- **Readability.** Provides a series of readability scores for the text, calculated from the Flesch Reading ease scale, which calculates readability based on the number of words per sentence and the number of syllables per word.
- **Sentence length.** The program provides a visual bar chart representing the length of every sentence in the text, to allow users to identify too-long sentences and pacing issues.

The free version of ProWritingAid allows full access to the revision features for free for texts up to 500 words in length. The paid version of the application allows to use the program's features on texts of any length. The paid version of the desktop app was used for the study conducted to answer RQ2 as described below.

The ProWritingAid API was used for the study designed to answer RQ 1 as described below. The ProWritingAid API allowed access to most of the same features as the desktop version. Grammar, spelling and style feedback would be dynamically provided in a textbox, while the rest of the feedback was only provided under the 'Summary' section, which gave participants an overview of the points outlined above.

Both the desktop version and the API used the same AWE engine, and therefore it was possible to maintain consistency of feedback between the two studies described in this thesis. The next two sections describe the design of each of the studies and how the chosen AWE program was integrated into them.

**3.2.3.3 The bespoke version of ProWritingAid.** As I mentioned in the previous section, in order to collect the texts produced by the students and answer RQ1, I designed a web application that worked on the ProWritingAid engine to collect data. The web application was programmed using Python which would use the ProWritingAid API to give students feedback while collecting data on their usage of the app. The Python programming language was chosen because the Flask framework was a straightforward way of developing a web application with limited resources and technical

knowledge. The web application was hosted at <http://grammaraid.ed.ac.uk/> and the code used to develop it can be found at [https://github.com/annie-primera/datacollection\\_v2](https://github.com/annie-primera/datacollection_v2).

The API allowed the developer to configure whether the engine would use their databases for academic writing, creative writing, general writing, business writing, technical writing or casual writing for generating the feedback. Since the focus of this research was on academic writing, that was the setting that was programmed into the app.

The sign-up screen, which can be seen in Figure 1 and Figure 2, contained information about the purpose of the study and data handling, as well as including a link to a full information sheet and the researcher's e-mail in case the students had any questions or wanted to withdraw. The sign-up sheet also asked for basic demographic information such as programme of study, IELTS score, first language, and time spent learning English. Participants needed to tick a checkbox stating that they understood the information and consented to being part of the study.

## Register a new account

### Before you register..

This web application is part of my PhD research. The full information sheet can be found [in this link](#), but the data that will be collected from you will be:

- Demographic information
- Time spent using the platform
- Functions used
- Feedback uptake
- Submitted text

All of this data will be stored anonymously.

If you want to withdraw from this study at any moment, for any reason, you can contact me at [ana.hibert@ed.ac.uk](mailto:ana.hibert@ed.ac.uk) and I will immediately withdraw you and delete any data collected on you.

By creating this account, you confirm that you are aware that your data will be collected, and that you understand the purposes for which the data will be used, as outlined [here](#).

In order to register for this software, it is important that you fill out all the fields in the registration form. Your information will always remain anonymous.

*Figure 1 – Information about the study in the registration page*

## Register

Gender

- male
- female
- non-binary
- other

Programme of study

TOEFL/IELTS score

Approximately how many years have you studied English? (type only a number)

What is your first language?

Email

Password

Confirm password

By ticking the box below and creating an account, you acknowledge that you have read the information provided and you agree to participate in this research. You also acknowledge that you understand which data will be collected, for what purposes, and that you are aware of the mechanisms for withdrawing from the study.

I accept

Register

Figure 2 – Registration form

The app consisted of a dashboard, shown in Figure 3, which students could use to submit new texts and edit existing ones. This dashboard allowed for the creation of new texts via the “New Text” button, as well as the editing of texts by clicking on their title. It also included a link to a tutorial on how to use the program and a page with frequently asked questions about the feedback given by the AWE program.

## Your dashboard

Here, you can create a new text, or continue working on previous texts. If you are not sure where to start, you can consult the [tutorial](#). You can also consult the [frequently asked questions](#).

New text

## Sample text

*Figure 3 – Dashboard*

Upon clicking on “new text”, students were taken to a simple text editor that did not offer any feedback for them to write their first draft. Upon clicking “submit”, a copy of that first draft was saved on the database, and students were taken to another editor that included the ProWritingAid feedback on their writing. This editor highlighted errors in grammar, spelling, composition and style in different colours. Grammar and spelling errors were highlighted in red, style suggestions were highlighted in yellow and composition issues such as sentence fragments or incorrect use of certain types of punctuation were highlighted in blue. A screenshot of this interface can be found in Figure 4.



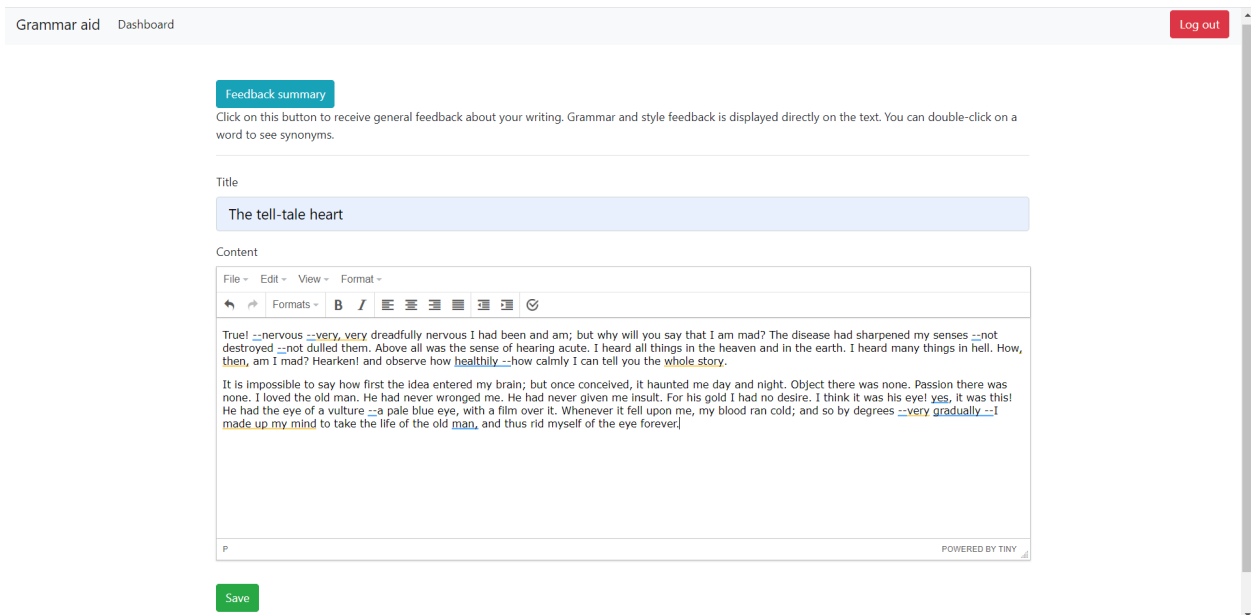


Figure 4 – Grammaraid Text Editor interface

The editor also included a “summary” function that gave the participant further analysis of their text composition provided by ProWritingAid, including length of sentences, most commonly repeated words, repeated phrases and consistency issues, among others. A screenshot of the “summary” section can be seen in Figure 5.

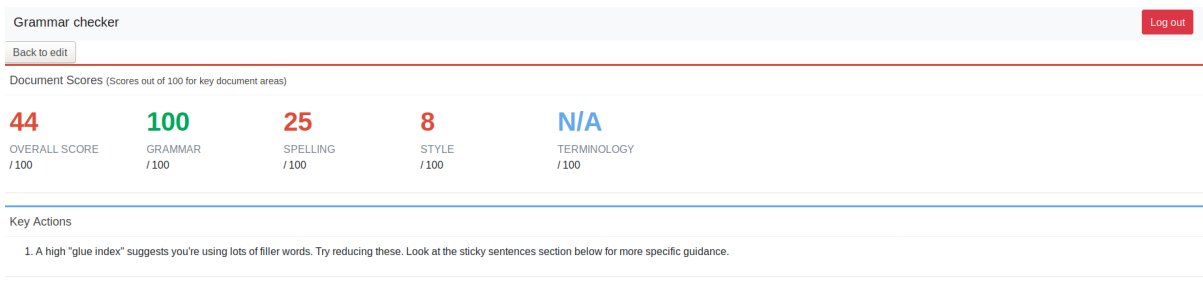


Figure 5 – Feedback summary page

The application saved copies of texts at different moments of the writing process. Besides saving the first draft, it would save a copy every time students used the “save” or “summary” functions, therefore obtaining snapshots of the text at different moments of the editing process, especially before and after receiving AWE feedback, which will be useful for comparison purposes. Other data

collected through the app included number of logins per user, number of individual submissions per user, and how many times the user accessed the “summary” function.

The specific tools that were integrated into the bespoke program for data collection will be discussed in section 3.3.1, which delves more deeply into how data was collected for the study designed to answer RQ1.

**3.2.3.3 The desktop version of ProWritingAid.** For the second study carried out to answer RQ2, students used the desktop version of ProWritingAid from the researchers’ computer, using a guest account created for the purpose to preserve anonymity and secure the student’s data.

The desktop version has the same functionalities as the bespoke version that was used for the first study, as well as some additional functionalities because it is the “full” version of the program. Like the bespoke program, it can be configured for academic writing, creative writing, general writing, business writing, technical writing and casual writing. As can be seen in Figure 6, the basic interface is the same as the bespoke version, with the program using different colours to highlight issues it finds with the text.

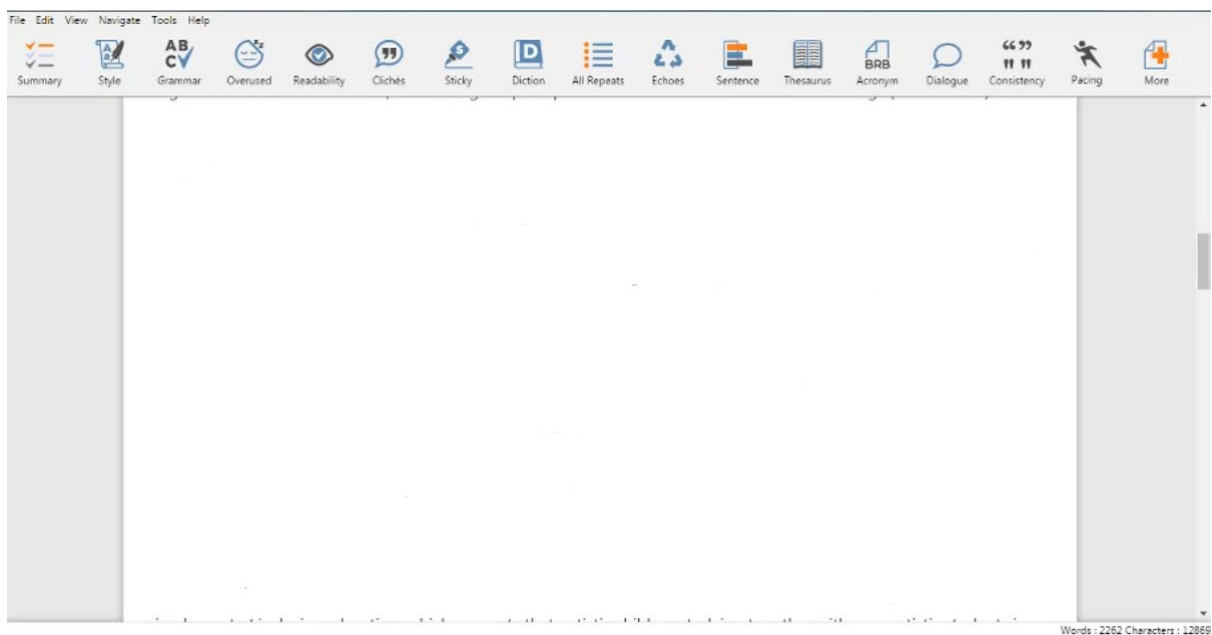


Figure 6 – Desktop version interface

The desktop version also includes a “summary” button, which gives the same information as the summary function in the bespoke version used for the first study. A screenshot of this “summary” section can be found in Figure 7.

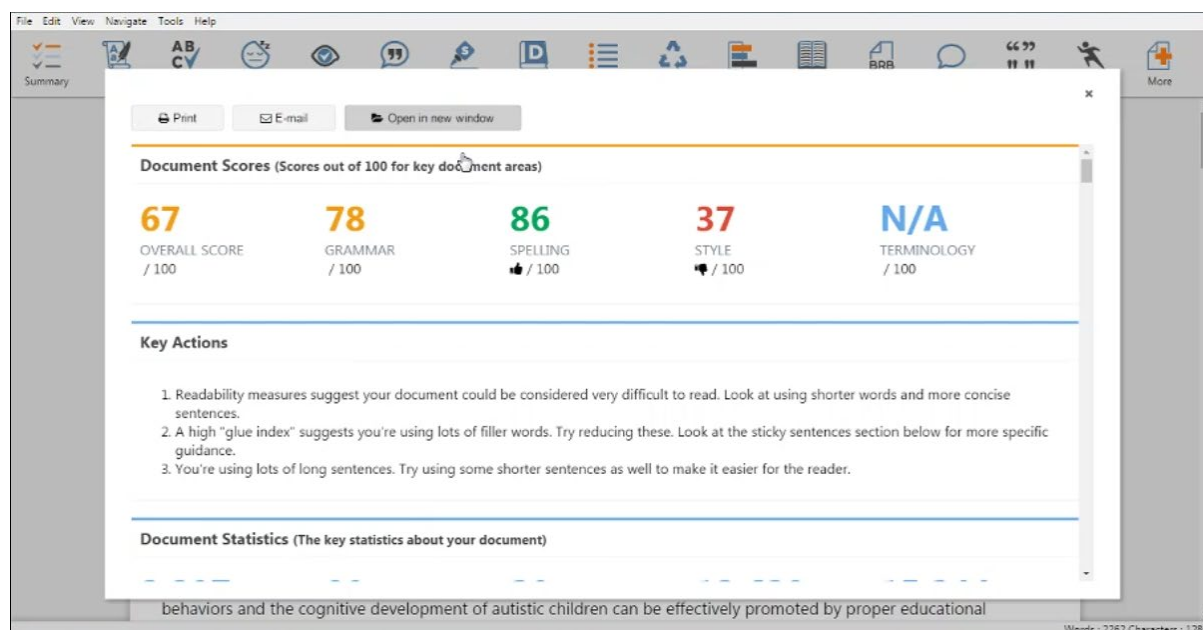


Figure 7 – Summary section, desktop version

Since the desktop version is the “full” version of the program, it also included some extra functions that were not present in the bespoke version due to limitations with the ProWritingAid API. The menu bar with all the functions can be seen in Figure 8.

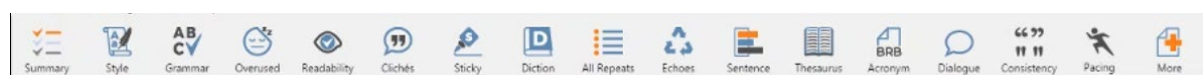


Figure 8 – Desktop version functions

These additional functions included the following:

- Overused – A section highlighting commonly overused phrases and words by comparing the frequency of these words and phrases with published writing. These include words with

vague or indefinite meaning such as “could”, “might” or “maybe”, words that tell rather than show, and intensifiers that could be replaced by stronger words.

- Readability – Uses the Flesch Reading Ease scale to score the readability level of the text and highlights what it considers to be “hard to read” paragraphs for revision
- Cliches – Finds and highlights common writing cliches in the text
- Sticky – Highlights “sticky sentences”, which are sentences that contain too many “glue words”, which are words that don’t carry much meaning in and of themselves such as “the”, “from”, “and”, etc.
- Diction – Highlights awkward sentences, which use unnecessarily long collocations such as “due to the fact that” instead of “because”.
- All repeats – Shows phrases that you have repeated many times in your text
- Echoes – Shows phrases and words that you have repeated in close proximity to each other
- Sentence – Shows you the length of every sentence in your text to avoid overly long or run-on sentences
- Thesaurus – Gives synonym suggestion on all nouns and verbs in the text
- Acronym – Points out acronyms that have not been explained in the text
- Dialogue – Highlights dialogue tags such as “said”, “asked”, “shouted”, etc.
- Consistency – Looks for consistency in the use of punctuation such as the oxford comma, hyphens, quotes and double quotes
- Pacing – Identifies “slow-paced” parts of creative writing manuscripts

Each of these functions includes a short explanation about what the feedback means, as well as links to longer documents that explain how the program comes to the conclusions that it does and gives tips on using that feedback to revise texts. A cheat sheet including all these functions and a short explanation of them was given to participants to ensure that they were informed as to the functions of ProWritingAid even if they did not click on the links for further information.

### **3.3 RQ1 – The effects of AWE feedback on student texts**

The first research question focused on the product of revision with AWE programs, namely, the texts themselves. As has been mentioned previously, most research into the use of these programs in ESL and EFL settings has focused on how holistic scores, machine-generated scores, accuracy, or error rates change as a result of introducing an AWE program to the revision process in writing classes. Some of these studies are quite specific, focusing on accuracy of a limited number of grammatical categories (Z. Li et al., 2017; Long, 2013; Saricaoglu & Bilki, 2021).

However, reductions in error rates or increases in scores and accuracy do not paint the whole picture of revision with AWE programs, for two main reasons. First, several researchers have pointed out that, when faced with AWE feedback, some students tend to adopt avoidance strategies rather than address the issues raised by the AWE program (El Ebyary & Windeatt, 2010; Ranalli, 2018). These avoidance strategies are not exclusive to AWE feedback and have been pointed out as an issue with WCF in general (Truscott, 2007). These strategies include deleting parts of the text that have been flagged as problematic or avoiding using grammatical constructions or structures that have been marked as errors in the past.

A second issue, specific to the use of scores as a metric of success within the context of using AWE programs as tools for formative feedback, is that the scoring algorithms used by these programs can be gamed by writing longer texts, or using more complex vocabulary or syntax (C.-F. Chen & Cheng, 2008; Y.-J. Wang et al., 2013; Wilson & Czik, 2016). While some researchers consider these score-gaming behaviours as an exercise in critical thinking by the students and therefore not a main cause of concern in the implementation of these programs unless abused (Bai & Hu, 2017), it does have an impact on the validity of scores as measures of success for AWE program implementation.

While this does not mean that there is no merit to using error rates, accuracy rates, and scores to determine the utility of AWE as tools for providing feedback, it does mean that AWE research should complement these metrics with other forms of analysis that can give a more well-rounded picture of

the effects of AWE feedback on student texts. Additionally, the purpose of this thesis is not to evaluate the merits of the feedback provided by AWE programs, but rather to contribute to a deeper understanding of how these technologies are being used. Therefore, it was deemed relevant to find other metrics by which the texts produced by students could be analysed.

Outside of the domain of research into the use of AWE programs as tools for formative learning, there has been much research carried out into how certain linguistic features of texts can be used to gauge the writers' proficiency by finding correlations between different linguistic features and human ratings of writing quality (M. Kim et al., 2018; Ryu, 2020). These linguistic features include lexical characteristics such as lexical diversity (Bestgen, 2017; Kato, 2019), syntactic characteristics such as 'that' verb complements (Crossley et al., 2014), and markers of coherence and cohesion (Crossley et al., 2019). An overview of the current state of research into the linguistic features that correspond to writing proficiency has been presented in the literature review in Chapter 2 but, in summary, there is evidence to support the assertion that some linguistic features can be predictors of human ratings of writing quality.

It was believed that, by using these linguistic features to analyse the texts collected as part of this study, it would be possible to have a more comprehensive overview of how students have changed their texts. To my knowledge, no research into the use of AWE has used these types of measures to analyse the end product of revision with AWE programs. However, it is important to begin considering these factors, as diverse linguistic features are usually considered when calculating scores in both AWE and AES systems (Jung et al., 2019). Since most software companies consider which features they use to calculate scores to be a trade secret (Kumar & Boulanger, 2020), in this study I consider the features that research has shown to be relevant in understanding proficiency and quality. As a first approach to integrating the study of linguistic features into AWE research, said features were used in this thesis as a proof of concept.

### 3.3.1 Data collection

#### 3.3.1.1 The AWE program.

As mentioned in section 3.1.3.2, a bespoke web application called GrammarAid was programmed to use the ProWritingAid API to provide students with feedback using the ProWritingAid engine and collect data on how students used this program. GrammarAid consisted of a login page, a registration page, a dashboard page, a summary page and two separate editor pages.

Once the students had created an account, using the process detailed in section 3.1.3.2, they would be sent to the 'Dashboard' page (see Figure 9), which allowed them to create new texts and access texts they had uploaded into the application in the past. This page also included a link to a tutorial and a page with frequently asked questions.

#### Your dashboard

Here, you can create a new text, or continue working on previous texts. If you are not sure where to start, you can consult the [tutorial](#). You can also consult the [frequently asked questions](#).

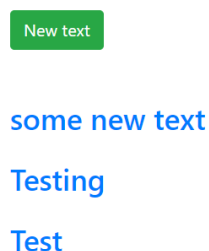


Figure 9 - Dashboard page

Whenever the user clicked on the 'new text' button, a time-stamped entry would be saved into a database called "Actions" that collected students' actions. They would then be taken to a page that included a blank editor. This version of the editor merely asked for a title and a body of text but offered no feedback or otherwise any other information (see Figure 10). The purpose of this featureless editor was to capture a first version of the text uninfluenced by the AWE feedback. When the user clicked on the button 'save and check grammar', two copies of the text would be

saved: one would go to a database called “Student Texts” which kept track of the latest version of the text and displayed it to the user, and the other would go to a separate database called “Text versions” that stored all versions of each text submitted by the user.

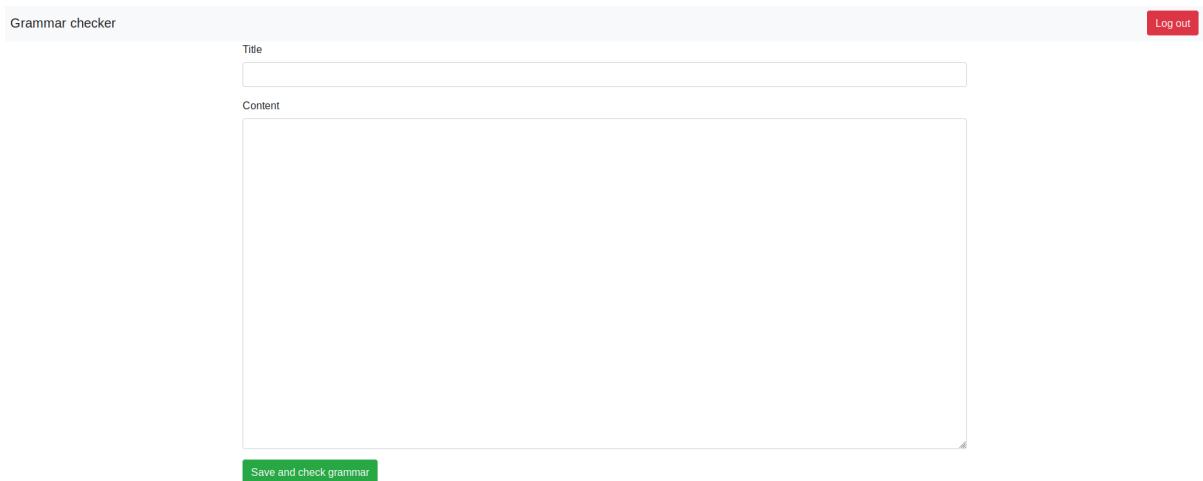


Figure 10 - New text editor

The ‘save and check grammar’ button would send the students to a ‘Summary’ screen that used the ProWritingAid API engine to give students an overview of the feedback given to them by the AWE program. An example of this summary can be seen in Figure 11.

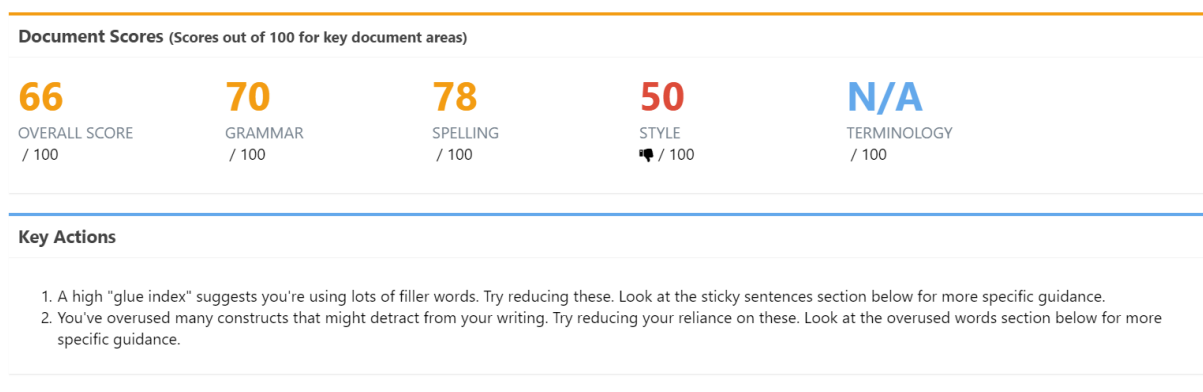


Figure 11 - Summary function of ProWritingAid



After this, the students were taken to a second ‘editor’ screen that allowed them to continue working on their text and used the ProWritingAid engine to give them dynamic feedback on the text they were typing (see Figure 12). This worked similarly to the desktop version, with potentially problematic areas being highlighted by the program, which then provided suggestions on how to address these potential issues. They also had a ‘Summary’ button that led them back to the Summary page that gave an overview of all applicable feedback. Every time the ‘Summary’ button was clicked, a time-stamped entry would be added into the “Actions” database indicating the participant had consulted the Summary.

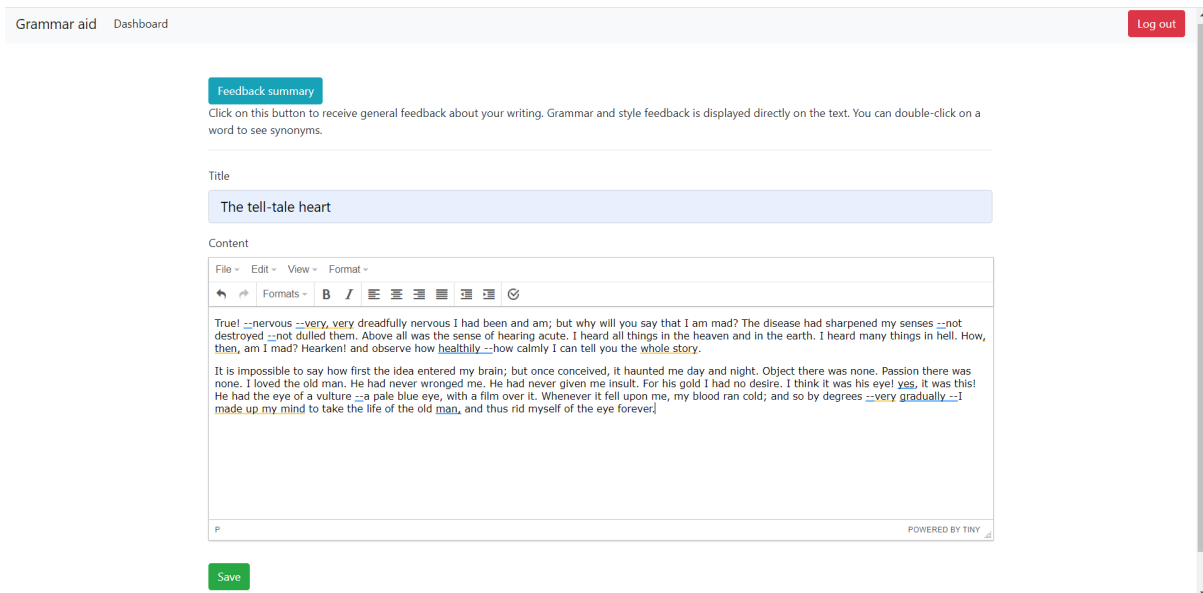


Figure 12 - Editor providing dynamic AWE feedback

Every time the students clicked on the ‘save and exit’ button, the program would add a time-stamped entry into the “Actions” database, in addition to saving the text in two places. First, the text would be updated in the “Student texts” database so the user could be able to access it via the Dashboard. Then, a separate copy of the text would be added to the “Text versions” database. This allowed the program to make several ‘snapshots’ of the texts as the participants worked on them: at least one before they received the AWE feedback and one after they received the AWE feedback.

Every time the students clicked either the 'Save and exit' or the 'Summary' button, the program would add a copy of the text to the "Text versions" database.

If the students returned to a previous text using the Dashboard, a time-stamped entry would be added into the "Actions" database indicated that the students had accessed a text they had previously worked on. A breakdown of the data collected on student actions can be found in Table 2.

**Table 2 - Overview of action data collected by the GrammarAid web application**

<b>Actions</b>	<b>Definition</b>
Log in	User logs into the application
New text	User creates a new text from the Dashboard
Summary	User clicks on the 'summary' button in the AWE editor
Save	User saves the text after receiving AWE feedback
New click	User opens a previously created text
Logout	User logs out from the application

In summary, the GrammarAid application collected data on the users' actions as they navigated through the program and used its different functions, as well as collecting snapshots of the texts submitted by the students at different moments of the revision process. The total data collected by the GrammarAid program was distributed into four databases, as shown in Table 3.

**Table 3 - Databases and types of data collected by GrammarAid**

<b>Database</b>	<b>Description</b>
Users	Stored demographic data on the users, as well as encrypted login details
Actions	Stored the different actions that users could perform while interacting with the application (see Table 2)
Texts	Stored the latest version of the text saved by the students
Text versions	Stored a snapshot of the text every time they created a new text, and every time they saved each text after making changes to it.

**3.3.1.2 The workshop.** A 5-week academic writing workshop was designed for the purposes of recruiting participants for this study. The workshop focused on crafting the introduction, body, and

conclusion of an academic essay, along with the correct way to format and cite papers according to the APA format, which is widely used in the field of social sciences that participants were studying.

The workshop was divided into 5 weeks, with a formative activity assigned to the students at the end of each session.

- The first week introduced the concept of academic writing, how it is different from other forms of writing, and some conventions of academic writing. It also explained the main parts of the introduction with examples. As homework, students were asked to use GrammarAid to write an introduction.
- The second week covered how to revise an introduction, and the students were divided into pair to practice revising each other's texts. The second part of the session explained how to write the body of an essay, focusing on the basic structures of arguments, and referencing and citation conventions. As homework, students were asked to use GrammarAid to write two or three paragraphs of the body of their essay.
- The third week taught students how to revise the structure of their arguments and the organisation of the body of their essay, including brining attention to common mistakes in academic essays. The students were then asked to divide into pairs and practice revising each other's texts. The second part of the session covered the structure of a conclusion, providing examples. As homework, students were asked to use GrammarAid to write a concluding paragraph for their essay.
- The fourth week focused on how to revise the conclusion, and students were asked to divide into pair to practice revising each other's conclusions. On the second part of the session, I covered the topic of abstracts and how to write an effective abstract. As homework, students were asked to use GrammarAid to write an abstract for their essay.
- The fifth week focused on how to revise an abstract. Students were once again divided into pairs to practice the concepts learned on each other's texts. Students were also given

general revising tips for their essays, as well as advice on referencing and citations. The ending of the session was dedicated to specific questions the students might have.

The slides used for this workshop can be found in Appendix 1.

**3.3.1.3 Participants.** The purpose of this study was to understand how students use AWE programs to revise academic texts. Therefore, the participants of this study needed to be students who had several opportunities to write academic texts and the motivation to revise them. Taught postgraduate students were selected for this study as their programs students usually include academic essays as part of their course evaluations, and they are required to write a dissertation at the end of their period of study in order to receive sufficient credits for graduation. Participants for this study were therefore selected among postgraduate students whose first language was not English.

Participants were contacted through the English Language Education (ELE) centre at the Moray House School of Education and sports at the University of Edinburgh. The ELE sent an invitation to the writing workshop to students who had enrolled into the ELE's official academic writing workshops but had been unable to secure a place due to capacity issues. This improved the probability that I recruited participants who had expressed interest in obtaining help to revise their work and improve their English language skills.

Originally, I had planned to offer the workshop in two groups of around 20 students each. However, the response from potential participants was so overwhelming I had to expand to four groups to accommodate the total 79 participants who signed into the workshops. It should be noted that not all the participants completed all 5 weeks of the workshop or signed in to use the AWE program, as participation in the workshop was voluntary and did not count for university credits.

Of the 79 participants that signed in to the workshop, 69 created an account on the GrammarAid application and used the program to submit at least one text. Six participants identified as male and

sixty-three identified as female. First languages for participants included Chinese, German, Greek and Spanish, although the overwhelming majority of the participants were Chinese speakers, which mirrors the makeup of the PGT student population at Moray House. They had studied an average of 12 years of formal English education and had an average of 7 points in the IELTS test. A summary of participant characteristics can be found in **Table 4**.

**Table 4 - Participant details**

ID	Gender	Programme of education	IELTS score	Years of studying English	First language
1	Female	TESOL	7	15	Chinese
2	Female	Education	6.5	3	Chinese
3	Female	Sport Management	6.5	16	Chinese
4	Female	TESOL	7	2	Chinese
5	Female	Dance science and education	6.5	5	Chinese
6	Female	Landscape and Wellbeing	7.5	16	German
7	Female	Comparative education and international development	7.5	16	Chinese
8	Male	Education	7	5	Chinese
9	Female	Education	7	16	Chinese
10	Female	Education	7	12	Chinese
11	Male	Education	6.5	9	Chinese
12	Female	Education	7	6	Chinese
13	Female	TESOL	7.5	10	Chinese
14	Female	Education	6.5	8	Chinese
15	Female	Education	7.5	15	Chinese
16	Female	Education	6.5	15	Chinese
17	Female	Academic writing	7	9	Chinese
18	Female	Education	6.5	10	Chinese
19	Male	Language Education	7.5	3	Spanish
20	Female	TESOL	7	10	Chinese
21	Female	Education (CEID)	7.5	14	Chinese
22	Female	Education	7	10	Chinese
23	Male	Education	7.5	15	Chinese
24	Female	Education	6	10	Chinese
25	Female	Education	7	15	Chinese
26	Female	Child and adolescent psychology	7	13	Chinese
27	Female	Education	7.5	15	Chinese
28	Female	Education	7.5	15	Chinese
29	Female	TESOL	7.5	13	Chinese
30	Female	Education	6.5	8	Chinese
31	Female	TESOL	7	10	Chinese
32	Female	TESOL	7	15	Chinese

33	Male	Sport management policy	6	5	Chinese
34	Female	TESOL	7	13	Chinese
35	Female	Environmental protection and management	6.5	8	Chinese
36	Female	Education	7	15	Chinese
37	Female	Education	7	8	Chinese
38	Female	Sport policy and management	6.5	13	Chinese
39	Female	Education	7	13	Chinese
40	Female	TESOL	7	14	Chinese
41	Male	Education	7.5	14	Chinese
42	Female	Education research	6.5	10	Chinese
43	Female	Education	7	17	Chinese
44	Female	TESOL	7	10	Chinese
45	Female	Education	7.5	13	Chinese
46	Female	TESOL	7.5	17	Chinese
47	Female	Education	7	10	Chinese
48	Female	Education	7	15	Chinese
49	Female	Environmental protection and management	7	12	Chinese
50	Female	Education	6	12	Chinese
51	Female	Education	6.5	10	Chinese
52	Female	Education	7	12	Chinese
53	Female	Education	7	13	Chinese
54	Female	TESOL	7	7	Chinese
55	Female	Comparative education & international development	7.5	12	Chinese
56	Female	TESOL	7.5	18	Chinese
57	Female	Education	7	10	Chinese
58	Female	Education	7.5	15	Chinese
59	Female	Education		15	Chinese
60	Female	TESOL	7.5	18	Chinese
61	Female	Language education	6.5	12	Chinese
62	Female	Education	7	18	Chinese
63	Female	TESOL	7	15	Chinese
64	Female	TESOL	7.5	1	Chinese
65	Female	TESOL	7.5	20	Chinese
66	Female	Comparative Education and International Development	7.5	10	Greek
67	Female	TESOL	7.5	16	Chinese
68	Female	Cultural Landscapes	7	15	Chinese
69	Female	Education	7	10	Chinese

As part of the workshop, participants were given weekly assignments to practise writing the different parts of the essay that were being discussed during the sessions. They were asked to use GrammarAid to revise and submit the texts for the workshop, but also encouraged to use the program to revise texts related to other academic courses they were taking at the time. While most

of the texts collected by GrammarAid were submitted during the period in which the workshop was run, some participants continued to use the program to revise new texts for as long as 10 months after the workshops were finished.

**3.3.1.4 Ethics.** The ethical aspects of this study were all done in accordance with the guidelines given by the University of Edinburgh and the British Educational Research Association, and a summary of the project was submitted to the Ethics in Research Committee of the University for approval. Since this research project involved postgraduate students enrolled at the University of Edinburgh, it was deemed that they had a sufficient level of English to understand the implications of the study and consent to take part in it and were therefore not considered a vulnerable population. This section details the steps that were taken to ensure consent was obtained, and confidentiality and data protection were ensured.

When participants contacted me for the study, I gave them a full information sheet explaining the purposes of the study and the information that would be collected from them if they decided to participate, as well as the way in which the information would be handled. The information sheet can be found in Appendix 2. Prospective participants were asked to indicate that they had read and understood their participation in the study before being signed into the workshop. In the first session of the workshop, I explained in person what the research entailed, how their data would be stored and handled, and answered any questions they might have at that stage. Furthermore, the login page for the GrammarAid application included a brief summary of the same information. In order to register to use the GrammarAid program, students had to click on a box that indicated their consent to participate in this study.

All participant data was kept on a secure server managed by the University of Edinburgh. Participant identities were stored in an encrypted file on my computer, and only I had access to the password to unlock said file. Since the analysis focused on the texts themselves, the reporting of the results and consequent analysis do not make reference to individual students, and so their anonymity during the

study is maintained. Furthermore, no part of the texts analysed appears in the results section or the subsequent discussion to make sure that student identity is not compromised.

### **3.3.2 Data analysis**

**3.3.2.1 Text selection.** A total of 183 unique texts were submitted to the GrammarAid application over a span of 10 months. Given that GrammarAid collected snapshots of each text that was created, as well as a different snapshot every time the text was saved by the students, there were a total of 401 texts in the “Text Versions” database.

One of the main purposes of this study was to understand how texts change in response to AWE feedback. Therefore, a series of filters were applied to the original database in order to select only the texts that had the characteristics needed for this study. The complete code used to clean up the database can be found in Appendix 3.

First, a filter was applied to select those texts which had more than one version recorded by the program. Because of the way GrammarAid collected data, this meant that students created a new text but did not subsequently save it after being provided with the AWE feedback, and therefore there were no records of the participant having made any changes to their texts. Second, all texts under 100 words were discarded, since many of the analysis on linguistic features that would be applied on the dataset are not reliable in texts under 100 words (J. Kim, 2014). Once both of these criteria had been applied, the number of unique texts that could be used for analysis were 70.

A third filter had to be applied when it was discovered that, even though GrammarAid had saved several snapshots of a text, there were no changes between the first and the last version of the text. The SequenceMatcher ratio method from the difflib Python library that can compare pairs of input sequences, returning a measure of similarity between two texts as a number between 0 and 1. (Python Software Foundation, 2020). This measure is calculated by finding the longest subsequence present in both texts and calculates the matches using the following formula:  $2.0 * M / T$ , where M is the number of matches, and T is the total number of elements in both sequences (Jaiswal, 2019).



Texts that returned a similarity measure of 1 between the first and last version were discarded, as this means the two texts are identical and therefore the participants made no changes in response to the AWE feedback.

Once all three filters had been applied, 30 unique texts, each with two or more versions, were found suitable for analysis. The number of snapshots per text ranged from 2 to 15. The texts had a mean of 1,122 words, with the shortest text having only 101 words and the longest text having 4,083 words.

***Interactions with GrammarAid.*** Two sources of data were used to understand student interaction with the AWE program. First, the number of submissions per user was obtained to gain an understanding of how often individual users accessed and edited their texts. This analysis, since it was made at the level of the user and not the text, considered users who submitted texts to GrammarAid but did not make any changes to them, as knowing how many participants did not find the AWE feedback useful enough to edit their texts was deemed important for the purposes of this study. The code used to obtain this data can be found in Appendix 3.

Second, user actions recorded in the “Actions” database were analysed using a process mining methodology, which considers individual events as part of a sequence generated by a process which is part of a narrative (Bannert et al., 2014). This was deemed a suitable method for analysis that would allow me to bring user actions together to form a picture of how each user typically used GrammarAid to access their texts and supplementary AWE feedback. The program chosen for this analysis was Disco (<https://fluxicon.com/disco/>). Disco uses a process mining algorithm based on Fuzzy Miner, which bases its process discovery on a map metaphor, and adds highlighting of frequent activities and paths (Gunther & Rozinat, 2012), which helps in discovering the most common processes used by participants when interacting with GrammarAid.

An analysis of the process mining logs revealed that some actions were logged more than once in the database, as evidenced by the fact that some actions seem to happen twice within milliseconds of each other. This was due to recurring server errors with the application that could not be solved

by the University's IT department in time for this study. To account for this, all actions recorded in the database which were duplicated and recorded as performed within milliseconds of each other were consolidated into a single action for the purposes of analysis.

**Text changes.** As mentioned previously, the SequenceMatcher method was used to calculate the difference between two texts on a scale of 0-1, where 0 meant that the two texts were completely different, and 1 meant that the two texts were completely identical.

However, merely quantifying the differences does not give insight into the nature of the changes performed by the students. Therefore, the 'compare' function of Microsoft Word was used to compare the first and last version of each text to identify specific changes made between versions. The first text was run through the AWE program and the suggestions offered by it were compared to the changes made by the students on subsequent versions of the texts in order to understand which of the changes made to the texts were a result of the AWE feedback and which changes were made by the students out of their own initiative. Instances where the AWE program had offered feedback, but the participant had made no changes to their text, were also recorded.

The texts were then uploaded to NVivo and each change was coded depending on whether it had been made as a response to AWE feedback, or if it had been initiated by the student without prompting from the program. The coding scheme can be found in Table 5.

**Table 5 - Coding used for changes made to participant texts**

<b>Code</b>	<b>Description and examples</b>
Accepted feedback	Participant accepted feedback as suggested by the AWE program without making any further changes. <b>Original sentence:</b> In the article, the shifts <u>will be considered</u> from three perspectives [...]. <b>AWE feedback:</b> Passive voice. Consider changing to: "In the article, we will consider the shifts from three perspectives". <b>Revision:</b> In the article, <i>we will consider</i> the shifts from three perspectives [...].

Own correction in response to feedback	<p>Participant made a change to a part of the text highlighted by the AWE program, but the change did not correspond to the suggestion made by the AWE program.</p> <p><b>Original sentence:</b> [...] global citizenship requires <u>learner</u> to reconsider their identity of belonging to one country or one region [...]</p> <p><b>AWE feedback:</b> Possible missing determiner. Consider using “a learner” or “the learner”</p> <p><b>Revision:</b> [...] global citizenship requires <i>learners</i> to reconsider their identity of belonging to one country or one region [...]</p>
Own correction	<p>Participant made a change to the text without being prompted by the AWE program.</p> <p><b>Original sentence:</b> [...] the ability to read and write in and out of school setting.</p> <p><b>AWE feedback:</b> None</p> <p><b>Revision:</b> [...] the ability to read and write in and out of <i>the</i> school setting.</p>
Ignored feedback	<p>The AWE program suggested a change or revision, but the participant made no changes in response to this prompt.</p> <p><b>Original sentence:</b> [...] language learning <u>focus</u> more on communication competence [...]</p> <p><b>AWE feedback:</b> Possible agreement error. Consider using “focuses”.</p> <p><b>Revision:</b> None.</p>

These codes were used to understand the nature of the changes made to the texts in response to the AWE feedback. Frequency analyses were carried out to map out the distribution of these changes and how much of the feedback the participants incorporated into their writing, as well as the number of revisions participants made without help from the AWE program.

To gain a further degree of granularity in the analysis, each instance of feedback provided by the AWE program was classified according to the type of feedback given by the AWE program: agreement, capitalisation, missing determiner, overuse of passive verbs, prepositions, punctuation, readability improvements, redundancies, repeated words, spacing, spelling, split infinitives, suggestions for stronger wording and general style suggestions. Frequency analyses were also

carried out to understand how each of these different types of feedback was either accepted or rejected by the participants.

**Linguistic markers of proficiency.** The last level of analysis was focused on whether revising the texts with the use of the AWE program led to any changes in linguistic markers of writing proficiency.

Following the review of existing literature on the topic carried out in Chapter 2, six lexical features have been found to correlate with human judgments of writing quality. While research into which linguistic features can be predictors of writing proficiency has produced often contradicting results (Ryu, 2020), for the purposes of this research, only the features that had been independently found to correlate to writing proficiency were chosen. These features are reported in Table 6.

**Table 6 - Linguistic features which correlate to writing proficiency**

<b>Linguistic features</b>	<b>Definition</b>
Lexical diversity	Refers to the diversity of the vocabulary used in a text, usually calculated by looking at type-token ratios. Research suggests that more proficient writers tend to have a broader vocabulary and therefore employ higher lexical diversity in their writing (Bestgen, 2017; Crossley et al., 2010, 2014; Ha, 2019; among others)
Word familiarity	A measure of how commonly words are experienced. I.e. ‘run’ would have a higher word familiarity score than ‘extemporaneous’. Word familiarity has been found to correlate to proficiency, as more proficient students tend to use less familiar words (Jung et al., 2019; M. Kim et al., 2018; Kyle & Crossley, 2015; among others).
Word frequency	A text-based measure of occurrences of words. This measurement is closely related to word familiarity, and has shown to correlate to writing proficiency for the same reasons (Crossley & McNamara, 2012; Jung et al., 2019; Ryu, 2020; among others)
Word meaningfulness	A measure of how likely words are to activate other words. I.e. a word like ‘fish’ will have a higher meaningfulness score than ‘carousing’, and research has showed that words with low meaningfulness can function as predictors of L2 language proficiency (Crossley & McNamara, 2011, 2012; Kyle & Crossley, 2015)
Word hypernymy	Hypernymy relates to hierarchical associations between more general words (hypernyms) and more specific words (hyponyms). An example of this would be the association between building and house, and corresponds to specificity of language. Research suggests more

proficient writers tend to use more specific words (Crossley et al., 2010; Guo et al., 2013; Kyle & Crossley, 2016; Ryu, 2020)

Word imageability Relates to how easily words can evoke mental and sensory images. For example, 'desk' is a highly imageable word, while 'liberty' would have a much lower imageability score. Since more concrete words are learned earlier than more abstract ones, imageability is believed to correlate to language proficiency. (Crossley et al., 2010; M. Kim et al., 2018; Ryu, 2020; among others).

In order to examine these lexical characteristics in the texts collected for this study, TAALES version 2.0 (Kyle & Crossley, 2015) and Python's Natural Language Toolkit (NLTK) were used. Both the first and the last version of the texts were analysed through these programs for the six lexical characteristics mentioned above, and a paired t-test was performed on the measurements to determine whether any of the measures had significantly changed after revision.

To summarise, two types of data were collected to understand the impact of AWE feedback on the revision of academic texts. The first type of data included the students' actions while interacting with the program, which were used to understand how often students used GrammarAid to revise their texts. The second type of data consisted of snapshots of the texts produced by the students, both before and after receiving AWE feedback. These snapshots were used to delve into how texts changed as a response to this feedback according to different metrics. The results and a brief discussion of this analysis can be found in Chapter 5.

### **3.4 RQ2 – Student engagement and decision-making**

The second research question sought to understand student engagement with the chosen AWE program. As mentioned previously, little research into AWE has looked specifically at how students engage with the feedback they receive from these programs. There have been some studies that focus on student perceptions (C.-F. Chen & Cheng, 2008; Fang, 2010; Lai, 2010), but these elicit student opinions on AWE programs after the treatment has ended and do not specifically look at how students interact with the programs. Some studies (Koltovskaia, 2020; Zhang, 2017, 2020;

Zhang & Hyland, 2018) do focus on student engagement with AWE programs, although none of them focus on how self-regulation strategies are used by students to engage with these programs. This is an important point to consider, as AWE programs are usually meant to be used in the students' own time and therefore their use can be considered to be mediated by self-regulation (Hibert, 2019b).

Therefore, the main purpose of this study was not only to look at whether students engaged with the feedback they received from the AWE program, but to understand the self-regulation mechanisms and decision-making processes that underpinned that engagement. Since self-regulation and engagement are two concepts that are closely linked with each other, with self-regulation being considered a core element of engagement (Wolters & Taylor, 2012), the framework of self-regulated learning was used to study participant engagement with the AWE feedback.

In order to fulfil this purpose, the study focused on RQ2 and its three sub research questions to understand engagement. Namely, I identify specific self-regulatory strategies to capture cognitive engagement with AWE feedback, understanding how this cognitive engagement leads to making decisions regarding feedback uptake that manifests itself in behavioural engagement, and whether these engagement processes result in students making good uptake decisions given that not all AWE feedback is reliable (Bai & Hu, 2017).

This study was carried out by recruiting postgraduate students from the University of Edinburgh and following the data collection procedures outlined in the next section. The summer period was chosen because it is the period during which postgraduate students work on their dissertations, and therefore would have an incentive to participate in this study and use the AWE program to revise their work.

Two main sources of data were collected for this study: think-aloud protocols and screen recordings of participants interacting with the AWE program. Traditionally, research into self-regulation has focused on self-reports given by participants after a task, although some have argued that this data

collection method might give an inaccurate picture of actual self-regulation strategies, since what students report they do when they are studying does not necessarily correspond to what they actually do (Hadwin et al., 2007). Think-aloud methods, on the other hand, help identify metacognitive and cognitive processes as they are happening (Bannert & Mengelkamp, 2008), and therefore allow researchers to make inferences about these processes (Veenman, 2013). While there are some concerns that the process of being conscious of one's thoughts and verbalising them might produce cognitive load or affect metacognitive processes, there is no evidence that is the case (Bannert & Mengelkamp, 2008). Furthermore, there have been some calls in recent times for the necessity of using think-aloud methods to help students verbalise their cognitive engagement (Zhang, 2020) and study the underlying reasons as to why students accept or dismiss AWE feedback (Link et al., 2020)

However, any self-reporting instrument has its flaws, and therefore it is recommended to use more than one data collection method to increase the validity of data (Schellings & van Hout-Wolters, 2011), and traces of student actions are considered to provide the fine-grained data necessary to augment self-reports of self-regulation strategies (Hadwin et al., 2007). Furthermore, think-aloud protocols allow some insight into cognitive engagement (Bannert & Mengelkamp, 2008), but do not necessarily offer information on behavioural engagement, which can be defined as uptaking the AWE feedback or revising their texts (Zhang & Hyland, 2018). Therefore, the second main source of data collection were screen recordings, as they allowed me to see which specific actions the students were undertaking and were therefore a rich source of behavioural data.

### **3.4.1 Data collection**

**3.4.1.1 Participants.** It was decided that, to maintain consistency of results, it would be beneficial to recruit participants with similar characteristics as the ones recruited for the previous study. It was also decided to use a different cohort to see if the results from the previous study would be echoed in a different study, therefore adding validity to my research (for a more thorough discussion on

validity, see section 3.4 below). Therefore, the participants in this study were postgraduate students at the University of Edinburgh whose first language was not English. As with the previous study, the rationale for recruiting from this group of students was that they would need a sufficient level of English proficiency to be accepted into a postgraduate program in the first place, and Masters programs usually require students to produce several pieces of academic writing, including a dissertation at the end of their period of studies.

Participants were contacted through the English Language Education (ELE) centre at the University of Edinburgh. The ELE sent an e-mail to ESL students that were studying on a postgraduate program at Moray House School of Education and Sports, and therefore had dissertations to complete during the summer when the study was carried out. A copy of the invitation letter sent to the students through the ELE can be found in Appendix 4. Interested students self-selected and volunteered to participate in this study.

In total, 11 postgraduate students were recruited. Nine of the students were female, and their first languages included Chinese, Spanish, Thai and Turkish. Participants had studied an average of 11 years of English and had an average of 7 in the IELTS test. The participants were studying a postgraduate degree in Teaching English as a Second Language (TESOL), Education, or Sports Policy, although the majority were studying TESOL. A full breakdown of participant characteristics can be found in Table 7.

**Table 7 - Participant details**

Pseudonym	Gender	Program	TOEFL/IELTS	ELE	L1
Melissa	Female	Education	7.5	15	Chinese
Alice	Female	TESOL	6.5	13	Chinese
Jamie	Female	Education	7	12	Chinese
Rachel	Female	TESOL	7	7	Chinese
Daniella	Female	Education	7	12	Chinese
Nancy	Female	TESOL	7	12	Thai
Violet	Female	TESOL	7	14	Chinese
Susan	Female	Language Teaching	7	12	Turkish
George	Male	TESOL	8	10	Spanish



Cristina	Female	Education	6.5	11	Chinese
Iain	Male	Sports Policy	7	8	Chinese

In order to maintain the anonymity of the participants, each participant was assigned an identification number with which all their data was labelled. To make reporting results easier, each participant was also randomly assigned a pseudonym.

**3.4.1.2 Procedure.** The AWE program used for this study was the desktop version of ProWritingAid. Participants were asked to attend 4 revision sessions, lasting 30 minutes each, during which they were given free access to the AWE program on the researcher’s computer. To maintain anonymity and security, an extra Guest account was created on the computer which was wiped after every session so no participant had access to any other participants’ writings or identity. No identifiable data was kept on the computers after the sessions, and all data generated was stored in an encrypted drive.

Before the first session, participants were given an information sheet with details of the study and were asked to sign a consent form. They also received a short tutorial session on how to use the AWE program and were given a ‘cheat sheet’ that briefly explained the main functions of ProWritingAid that were available to them for reference purposes. Copies of all relevant documents can be found in Appendix 4. Finally, participants were taught how think-aloud protocols work and were given an opportunity to briefly practise.

During each of the four sessions, participants were encouraged to bring any part of their dissertation they wished to revise and use the AWE program for such a purpose. They were asked to verbalise their thoughts, actions, and decision-making processes as they were using the program. If the participants stopped talking, they were given a gentle reminder to continue verbalising their thoughts. Their interactions with the AWE program were captured through screen recording by using OBS (<https://obsproject.com/>), which also recorded their voices for the think-aloud protocol.

After the last revision session, a short 15-minute semi-structured interview was conducted with each participant to elicit their overall impressions of the AWE program and their experience with using it for revision. The interview questions can be found in Appendix 5. Only the audio was recorded for these interviews.

**3.4.1.3 Ethics.** Like with the study detailed in section 3.2, the ethical aspects of this study were considered following the guidelines set by the University of Edinburgh and the British Educational Research Association, and a plan for obtaining consent, preserving anonymity and ensuring data protection was submitted to the Ethics in Research Committee at the University. This section outlines the key points of that submission.

The participants of this study were postgraduate students enrolled at the University of Edinburgh. While their first language was not English, it was deemed that their acceptance to the University at postgraduate level meant their level of English was good enough for them to understand the implications of the study and give informed consent to their participation.

After being contacted by students interested in forming part of the study, I sent them an information sheet detailing the purpose of the study and the data that would be collected from them, as well as the steps that would be taken to ensure their anonymity and the protection of their data. A copy of the information sheet can be found on Appendix 4.

In the first session, a written consent form was given to the students, along with an explanation of the information covered in the information sheet, and students were given the opportunity to directly ask any questions they might have about the study. Once I was satisfied that they understood the research, I asked them to sign the consent form. The format for this can be found in Appendix 4.

In order to maintain anonymity of the participants, their personal data was stored in an encrypted file and only I had access to the password. Each participant was assigned a random 4-digit number

for identification, which was used to identify all data collected on them. Furthermore, for the results and discussion section of this thesis, each participant was randomly given a pseudonym for ease of reporting. None of the files they used during this study were stored on my computer, and a fresh user was created for each revision session and deleted after the sessions so no personal information such as login data for their e-mails or copies of their texts would be saved in the machine. All data collected on the students, including screen recordings, think-alouds and interview recordings were stored in an encrypted USB drive and only I had access to the password.

### 3.4.2 Data analysis

**3.4.2.1 Cognitive engagement.** The audio for each of the four sessions and the final interview was transcribed and uploaded to NVivo for analysis, along with the screen recordings. Each of the 11 participants participated in four sessions, although one participant did not feel well enough to think aloud during their last session and so there is no transcript for that particular session. Therefore, there were a total of 43 think-aloud transcripts and 4 short interview transcripts for analysis. A sample transcript can be found in Appendix 6.

These transcriptions were coded for self-regulatory strategies according to a scheme adapted from Bannert et al. (2014). Their original model considers four categories and several sub-categories used to analyse self-regulation on think-aloud protocols, which can be seen in Table 8.

**Table 8 - Bannert et al.'s (2014) self-regulation processes**

Process	Subprocesses
Metacognition	Orientation Goal specification Planning Searching for information Judgement of information relevance Evaluation Monitoring and regulation
Cognition	Reading Repeating

	Elaboration
	Organization
Motivation	Positive and negative motivational utterances on task
Others	Task-irrelevant aspects such as questions, ambiguous utterances, or silence

However, these categories were intended for study sessions rather than revision sessions. While both study and revision sessions rarely include teacher interventions, are solo activities, originate with a general goal, occur in settings where students can control their environment to satisfy their preferences and produce observable traces of cognitive processes (Winne & Hadwin, 1998). One of the main differences between revising and studying is that revising does not necessarily involve learning new material or synthesising information. Some modifications to the original coding scheme were deemed necessary to account for this difference in format.

For example, Bannert et al.'s (2014) classification includes 'searching for information' as a category under metacognition, one which is important during study sessions because part of the goal in these sessions is to acquire new information, although this is not a priority during revision sessions unless the student is experiencing difficulty understanding the feedback received and wants to learn more about that specific point. Other authors consider handling task difficulties, which in this case would encompass not being able to understand the feedback received, as an important part of self-regulation (Azevedo, 2005). Help-seeking behaviours are one of the components of handling task difficulties, something that was deemed relevant for my coding scheme because, when faced with AWE feedback they do not understand, students have the option of using the program's own resources for clarification or looking up relevant information in search engines. Since the sessions were conducted with the researcher present, some students also verbally asked for classification, and so these three help-seeking behaviours were incorporated into the coding scheme coded as "using software tools", "using external tools" and "asking for help". Bannert et al. also considered orientation (overview over material) and planning (planning of proceeding) as two separate categories, but other authors do not consider to be separate actions (Azevedo, 2005). Therefore, it

was assumed that they were similar enough that both could be encompassed into a single 'planning' category.

Some of the cognitive processes considered by Bannert et al. were also deemed to be irrelevant to revising tasks. For example, organisation includes activities such as drawing mental maps or writing down major concepts, two actions that are relevant in studying and in drafting texts, but not in the revision process. Another cognitive element that was not deemed applicable for a revision task was elaboration, as it involves paraphrasing, connecting and inferring, which are important when learning and studying new material but do not necessarily form part of the revision process. Instead, rewriting and self-questioning were added to the cognitive category to better capture student actions during revision. Self-questioning happens when students pose questions and then re-read to find their own answers (J. A. Greene & Azevedo, 2009), something participants did frequently when faced with feedback they were unsure of or did not understand. Rewriting was also added as a code to encompass instances where students did not merely accept the feedback they received but rewrote sections of their text to improve them. Finally, a 'rehearsing' code was added for utterances in which students read their text out loud several times, with and without the feedback, to try to figure out which 'sounded' better.

Finally, Bannert et al.'s 'motivation' classification was reworded to 'student attitudes', keeping its original 'positive' and 'negative' values, to better fit the tripartite model of AWE engagement set out by Zhang and Hyland (2018) which considers student attitudes toward feedback as part of affective engagement.

The final coding scheme used for data analysis can be found in Table 9.

**Table 9 - Coding scheme adapted from Bannert et al.**

<b>Metacognition</b>	
Goal setting	Consists of possible, postponed or intended operations that have are not related to already-existing states.
Planning	Coordinating the selection of operators. Its execution involves making behaviour conditional on the state of the problem and a hierarchy of goals and subgoals. I.e. "I will now do x"
Judging information	Assessing the usefulness or adequacy of the feedback or information provided by the AWE program
Monitoring	Judging the perceived distance between the current product and the desired result
<b>Cognitive strategies</b>	
Reading	Reading out loud from the source material
Rehearsing	Reading a passage and making substitutions based on the feedback to gauge adequacy
Re-reading	Re-reading a passage that has already been read
Self-questioning	Posing a question and re-reading to improve understanding of context
<b>Handling task difficulties</b>	
Asking for help	Asking the instructor/researcher for help and clarification
Using external tools	Using external tools like googling, dictionaries, etc.
Using software tools	Using the tools provided by the software, i.e. thesaurus, explanations, etc.
<b>Student attitudes</b>	
Positive	Expressed positive attitudes toward using the AWE program
Negative	Expressed negative attitudes toward using the AWE program

The think-aloud and interview transcripts were uploaded to NVivo and coded according to the aforementioned scheme. Frequency analyses were carried out to understand the distribution of the self-regulatory processes expressed by the students during the think-aloud protocols.

**3.4.2.2 Behavioural engagement.** Coding the think-aloud transcripts to understand the self-regulation processes that could be gleaned from participant utterances helped give insight into their cognitive engagement with the feedback they received from the AWE program. However, in order to understand their behavioural engagement with AWE feedback, I brought together the data from the think-aloud transcripts and the screen recordings to explore how their cognitive processes expressed through their utterances influenced behavioural uptake of the AWE feedback.

For this level of analysis, a more inductive approach was adopted. In order to understand how students responded to feedback, an iterative coding process was carried out. First, student utterances were coded depending on their response to the feedback, which could be either accepting the feedback, ignoring it, or making their own corrections to the text. This allowed me to put student utterances into the context of their behavioural engagement with the feedback they received. Utterances which were coded as 'accepted' were those which were verbalised in connection to the students accepting the received feedback by clicking on the corrections proposed by the program and integrating them into their text as suggested. Utterances which were coded as 'ignored' corresponded to moments in which the participants rejected or ignored the feedback they received from the AWE program. Finally, utterances coded as 'own corrections' were assigned to utterances where the student either modified the feedback they received from the AWE program, or carried out their own corrections unprompted by the program.

Once all student utterances had been divided into these three categories, a process coding method was used to capture the reasons why the students had accepted or ignored the feedback, or made their own corrections. Process coding methods allow one to obtain a sense of how events develop and evolve (Saldaña, 2016), which was deemed adequate for understanding student's behavioural engagement with AWE feedback, as behavioural engagement can be defined as an event (Wolters & Taylor, 2012). This first iteration of the process coding method generated twenty codes, which were

there further refined into thirteen codes after a second iteration of coding. The codes generated in both iterations can be seen in Table 10.

**Table 10 - First and second iteration of coding for behavioural engagement**

<b>First iteration</b>	<b>Second iteration</b>
<i>Accepting feedback</i>	
Reflecting positively on quality of feedback	Reflecting positively on quality of feedback
Reflecting on how the feedback relates to their own writing	Reflecting on how the feedback relates to their own writing
Accepting unconditionally	Accepting unconditionally
<i>Substitution</i>	
Own changes	Own changes
<i>Rejecting feedback</i>	
Changes meaning	Disagreeing because feedback changes intended meaning
Concept too important to change	
Not aligned with original intentions	
Disagreeing because of intuition	Disagreeing because of intuition
Not really a problem	Not really a problem
Contradicts personal style	Contradicts personal style
Contradicts knowledge of academic writing conventions	
Contradicts collocation knowledge	Using previous knowledge to contextualise feedback
Contradicts previous feedback from others	
Contradicts what the student has read in other papers	
Ambiguous feedback	Complaining about ambiguity of feedback
Unsure, need to look up	Unsure about whether to accept feedback
Unsure about abilities	Expressing uncertainty about their own abilities
Unsure about quality	Expressing uncertainty about quality of feedback
Ignoring unconditionally	Ignoring unconditionally

In order to reduce the number of codes generated during the first and second iterations of coding, a third iteration of coding was applied following a pattern coding methodology, which identifies similarly coded data and brings them together into more meaningful units for analysis (Saldaña, 2016). This third iteration of analysis resulted in seven main themes, under which the thirteen codes obtained from the second iteration were gathered. The final codes used for analysis can be seen in



Table 11 and are organised depending on whether those codes reflect moments in which students accepted the feedback, rejected the feedback, or made their own changes.

**Table 11 - Final coding scheme for behavioural engagement**

<b>Accepting feedback</b>	
Agreeing with feedback	Reflecting positively on the quality of feedback
	Reflecting on how the feedback reflects on their own writing
Accepting automatically	Participant identifies the type of feedback provided as one that has consistently been good and therefore accepts the correction by default
<b>Substitution</b>	
Own changes	Participant disregards the feedback and carries out their own changes to the text
<b>Rejecting feedback</b>	
Disagreeing with feedback	Disagreeing because feedback changes the intended meaning
	Not considering the feedback a real problem
	Considering that the feedback contradicts their personal style
	Using previous knowledge to contextualise feedback
Questioning feedback	Complaining about ambiguity of feedback
	Researching feedback they are unsure of
	Expressing uncertainty about their own abilities
	Expressing uncertainty about the quality of feedback
Rejecting automatically	Participant identifies the type of feedback provided as one that has consistently been bad and therefore ignores the correction by default

**3.4.2.3 Quality of uptake.** Finally, the screen recordings that showed how the participants used the AWE program were used to record every action taken by the students in response to the feedback they were offered by the program. Figure 13 shows what the AWE program looks like. Potential issues identified by the program are highlighted, and when hovering over them ProWritingAid explains what the potential issue might be and offers a suggestion. The student can then either click on the suggestion or click the 'ignore' button so the AWE program ignores that specific instance of feedback. By clicking on the yellow 'i' icon on the upper right corner of the feedback box, students

could also receive an explanation for why the feedback was given and links to further resources that could be used to understand said feedback.

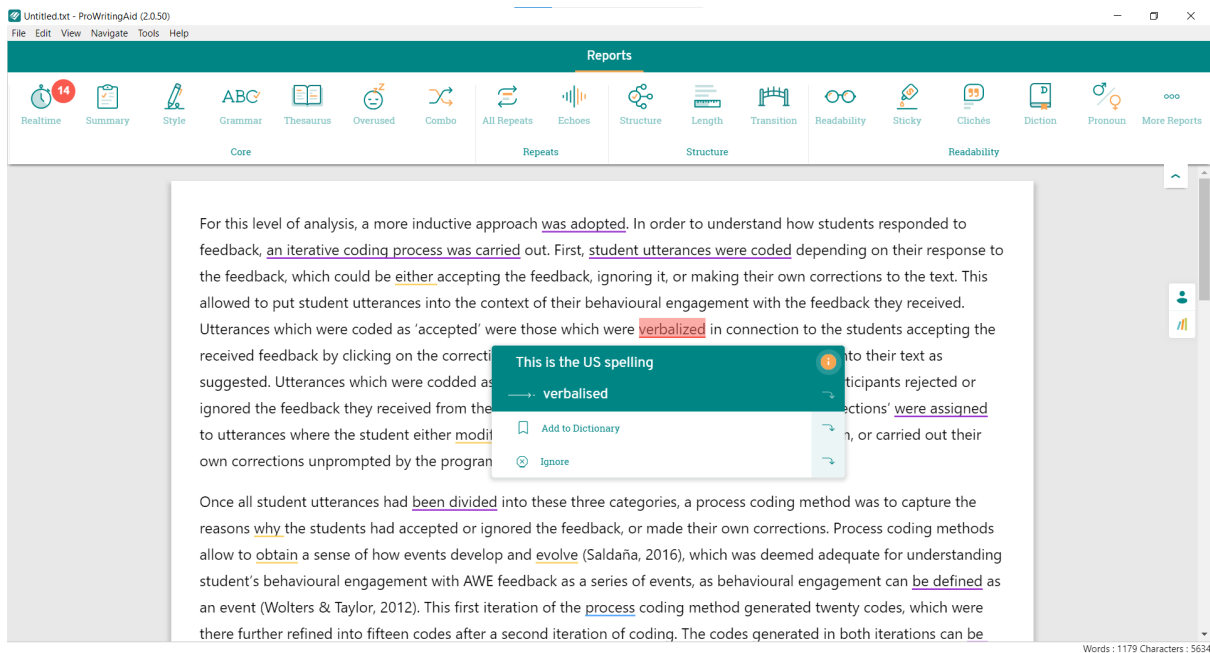


Figure 13 - Feedback provision by the AWE program

Students, however, could also choose to make their own corrections based on the feedback, or simply mark issues so they could look at them later. All of these actions were the basis for the initial coding scheme for these screenshots, which can be seen in Table 12, along with examples of how each code was applied<sup>3</sup>.

Table 12 - Actions taken by participants in response to AWE feedback

Student Action	Definition
Accept	<p>Student clicks on the suggestion given by the AWE program without making any further changes.</p> <p><b>Original sentence:</b> That is the reason why there are <u>a large number of</u> requirements for special education.</p> <p><b>AWE feedback:</b> Specify number, remove phrase or simply use 'many' or 'numerous'</p>

<sup>3</sup> Some of the text has been changed to preserve anonymity, but grammar and feedback were left intact

	<p><b>Revised sentence:</b> That is the reason why there are <i>many</i> requirements for special education.</p>
Reject	<p>Student either clicks the 'ignore' button when presented with feedback, or indicates they are considering the feedback through verbalisations but ultimately performs no actions based on it</p> <p><b>Original sentence:</b> Besides, <u>the sequence of questions can be varied</u> and the answers will be note-taken and tape-recorded.</p> <p><b>AWE feedback:</b> Passive verbs make your writing less direct. Consider: "it can vary the sequence of questions", "they can vary the sequence of questions", or "something can vary the sequence of questions".</p> <p><b>Revised sentence:</b> Besides, <i>the sequence of questions can be varied</i> and the answers will be note-taken and tape-recorded.</p>
Own	<p>Student looks at the AWE feedback but makes a correction different to the one suggested by the AWE program, or the AWE program offers no concrete suggestions to correct the text and therefore the students make their own changes.</p> <p><b>Original sentence:</b> [...] which is played predominantly by <u>female</u>, especially middle-aged [...]</p> <p><b>AWE feedback:</b> Possible missing determiner. Consider "a female" or "the female"</p> <p><b>Revised sentence:</b> [...] which is played predominantly by <i>females</i>, especially middle-aged [...]</p>
Postpone	<p>Student looks at the AWE feedback and marks or highlights it so they can come back to it later, making no immediate changes to the text</p> <p><b>Original sentence:</b> <u>With regard to</u> building a corpus, there are two broad approaches.</p> <p><b>AWE feedback:</b> Readability might be improved by using "regarding".</p> <p><b>Action:</b> Student highlighted the section to come back to later.</p>

Once the students' actions had been classified in this manner, the quality of each action was coded based on a scheme adapted from Bai & Hu (2017) and Chodorow (2010), which divided feedback uptake into "good", "bad" and "neutral". While assigning the codes, it was found that there were several reasons why a certain piece of feedback could be considered "bad", and therefore it was decided that a more fine-grained approach was required. Therefore, the "bad" category was subdivided into three. The resulting codes can be seen in Table 13.

**Table 13 - Codes for quality of feedback uptake**

<b>Feedback type</b>	<b>Code definition</b>
Good	AWE program correctly identifies an issue with the writing and the feedback offered correctly fixes the problem
Bad	
BadID_nofeedback	AWE program incorrectly flags something as a mistake, but offers no further feedback
BadID_badfeedback	AWE program incorrectly flags something as a mistake, and the feedback offered produces a new error that was not originally present in the writing
GoodID_badfeedback	AWE program correctly identifies an issue with the writing but offers feedback that introduces a new error or issue into the writing
Mixed	AWE program correctly identifies an issue with the writing and offers more than one solution, at least one of which introduces a new error
Neutral	AWE program offers feedback that neither improve nor worsen the existing text
None	AWE program flags an issue in the writing as problematic, but offers no specific feedback or guidance to fix it

To increase the reliability of this coding, I waited 6 months to revise the assigned codes, given that intra-rater reliability was deemed a suitable way of ensuring consistency and reliability in the coding process (see section 3.5.2 for a discussion on why). If there were any discrepancies between the initial code and the revised code, these discrepancies were discussed with an expert in the English language to resolve any conflicts and assign the most accurate codes possible. The Cohen's Kappa score for intra-rater reliability was 0.99, meaning that there was an almost perfect agreement between both rounds of coding.

Once these codes were assigned to each of the actions undertaken by the students during the recorded sessions, the data was cleaned up using Python; the code can be found in Appendix 7. Frequency analyses were carried out to understand the frequency and quality student's uptake of feedback. These analyses were carried out both at a global level and a per-student level to capture both trends in the sample and individual differences between students (see Appendix 7). Finally,

boxplots of student uptake were generated to be able to better compare quality of feedback and feedback uptake.

To summarise, data was analysed according to three different dimensions of student engagement. First, self-regulation was used as a framework to understand how students cognitively engaged with the feedback they received from the AWE program. This was achieved by adapting a coding scheme from Bannert et al. (2014) and Green & Azevedo (2009) to analyse the transcripts of think-aloud protocols. Second, the data from both transcriptions and screen recordings were used to analyse how this cognitive engagement influenced behavioural engagement and discover patterns in the decision-making process that resulted from this engagement. An inductive analysis was carried out on the transcript data for this purpose. Third, given that AWE feedback is often unreliable, I examined how this cognitive and behavioural engagement affected the quality of the uptake by using a coding scheme adapted from Bai & Hu (2017) and Chodorow (2010) to examine whether students accepted more “good” than “bad feedback.

### **3.5 Validity and reliability**

In this section, I will discuss some of the considerations that were taken into account to ensure the validity and reliability of the results of this research.

#### **3.5.1 Validity**

One of the reasons for choosing to conduct mixed-methods research is that using different methods can help improve the validity of research, for example by triangulating or helping the development of tools and methods (Thierbach et al., 2020). This was the main rationale for dividing the research in this thesis into two distinct but complementary studies. The study described in section 3.2 was mainly quantitative and was designed to understand how revised texts changed because of AWE feedback. The study described in section 3.3 was mainly qualitative and sought to provide a deeper understanding of how students use AWE feedback.

The main justifications for combining quantitative and qualitative research, as proposed by Greene et al. (1989) are:

- Triangulation – the use of different methods to corroborate results from data or find correspondence or convergence between them.
- Complementarity – using the results of one method to elaborate, clarify or illustrate the results of another method
- Development – informing the development of one method by using the results from another method
- Initiation – discovering paradox or conflicting results to increase the breadth and depth of interpretations
- Expansion – using the methods that are most appropriate for different parts of each study to extend the breadth and range of it

Of these, the most important reasons why I chose to use a mixed-methods approach to separate my research are triangulation, complementarity, and expansion.

Triangulation was the main way in which validity was approached in this thesis. There are several different types of triangulation that can be used in mixed-methods research (for a comprehensive review, see Cohen et al., 2017), but the most relevant for this study are paradigm triangulation , instrument triangulation and sampling triangulation.

The study related in section 3.2 leans more toward a quantitative paradigm, whereas the study in section 3.3 leans more to a qualitative paradigm. The former examines student texts by quantifying the difference between the texts before and after being revised with AWE feedback to understand whether the uptake of feedback is significant. The second study used a qualitative paradigm to understand more in-depth the thought process behind AWE feedback uptake. In the discussion section, we will discuss how tackling this subject from these two completely different perspectives offered complimentary data on AWE feedback usage.

Consequently, different instruments were used for both studies. The first study collected textual data from a bespoke application that used a pre-existing AWE engine, while the other study collected think-aloud data and screen recordings from participants as they engaged with the AWE program. Finally, two different samples were used for each of the studies to ensure a larger number of participants and make sure that, if results converged, that would underscore their validity. When findings obtained from different methods converge, this can enhance our belief that our results have validity (Burke Johnson et al., 2007). As will be discussed in Chapter 6, there were enough similarities in the results obtained from both studies to have a reasonable degree of confidence in their validity through triangulation.

Complementarity was also one of the reasons why I decided to use a mixed methods approach for this research. The quantitative study could help understand the students were using the feedback from the AWE program and how much their texts were changing as a response to it, but the qualitative study could expand on these findings by helping to understand why the students chose to either accept or reject the feedback they received. This way, the second study could help elaborate and clarify the findings of the first study.

In a similar vein, expansion was another one of the reasons for choosing a mixed methods paradigm. It was decided that quantitative methods were more appropriate to analyse the texts produced by the students and the changes that were made as a result of AWE feedback, and that qualitative methods were better suited for understanding student rationale behind their use of AWE feedback. This way, I analysed the same phenomena from different dimensions, and it allowed me to expand the breadth and range of our research. Most research into AWE has focused on either quantitative measures of increases in scores or qualitative analysis of student perceptions, but little existing research has taken both into account to create a more robust understanding of student usage of AWE programs (Hibert, 2019b).

I therefore approached validity in this study from different points of view, beginning from the selection of methodological and epistemological foundations and choosing to do two separate studies with complimentary methodologies to ensure the validity of our results. As will be seen in Chapter 6, the results of both studies seemed to converge and show similar findings, reinforcing the validity of this research.

### **3.5.2 Reliability**

Intra-rater reliability is a widely accepted form of ensuring reliability in medical research, since it is a metric of rater's self-consistency that seeks to ensure the reproducibility of trials (Gwet, 2008). Inter-rater reliability is the method most commonly used in educational research, but there are precedents for intra-rater reliability used to calculate reliability in educational settings, and specifically in the scoring of writing tasks (Cho, 1999; Y. Cohen, 2017; Kayapinar, 2014), prior learning (Stenlund, 2013) and high-stakes testing (Karadenizli-Çilingir, 2019; Kardong-Edgren et al., 2017).

Cho (1999) argued that, for grading ESL writing, intra-rater consistency is as significant as intra-rater reliability, given that the latter cannot be secured without the former because having good intra-rater reliability means that the rater has developed a consistent and coherent application of evaluation criteria. Cohen (2017) also argued that intra-rater reliability was the best method when the emphasis was to be placed in the confidence of ratings and demonstrated that it could be used to estimate the reliability of a specific rater. Therefore, it was decided that calculating intra-rater reliability would be an adequate measure of the reliability of the coding in both studies.

### **3.6 Conclusion**

This chapter outlined the methodology for the two studies conducted as part of this doctoral thesis. The first study dealt with the product of AWE feedback and looked at how each text changes as a response to the feedback. The second study expanded on the findings of the first by delving into why texts changed in the way they did by looking at student engagement with AWE feedback and the



decision-making processes that led to the acceptance or rejection of the feedback they received. The ultimate objective of both of these studies was to deepen our understanding of how students use AWE feedback to revise their texts by looking at the use of AWE programs from both the product and process perspective. The next two Chapters, Chapter 4 and Chapter 5, report on the results of these two studies, along with individual discussions for each of them. Chapter 6 presents an overall discussion that brings together the results of both studies and explores the implications of these results.

## Chapter 4 – The effects of AWE feedback on student texts

### 4.1 Introduction

One of the most popular strands of research on the use of AWE programs focuses on studying whether the use of AWE programs leads to improvement in writing skills or in the texts produced by the students. As mentioned in the literature review, this research has focused mostly on error rates and increases in scores as a measure of how texts change from exposure to AWE feedback (Hibert, 2019b), but research that focuses specifically on student uptake of AWE feedback has found that a lot of the feedback gets ignored (Link et al., 2020; Ranalli, 2018; Tian & Zhou, 2020), and that some students use avoidance strategies rather than revise their texts after receiving AWE feedback (El Ebyary & Windeatt, 2010).

However, these days we have the technology to go beyond error counts and scores in analysing the effects of AWE feedback on written texts. One of the conclusions drawn from my systematic literature review was that few studies done on the use of AWE feedback took advantage of the data-collecting capabilities offered by the programs themselves (Hibert, 2019b). One of the approaches that is the most promising in the study of the impact of AWE programs is the use of natural language processing (NLP) techniques to analyse the texts produced by ESL and EFL students and revised using AWE programs. NLP is already being used by AWE programs to analyse the texts submitted to them and generate feedback and scores (Kumar & Boulanger, 2020). While the exact algorithms used by AWE programs are often proprietary (Dikli, 2006; Wilson & Roscoe, 2020) and therefore it is difficult to know exactly how they generate feedback and scores, it is still useful to use some of these same techniques to analyse texts that have been revised using AWE feedback because they allow us to obtain different metrics associated with the quality of texts (Ryu, 2020) and are therefore useful for making comparisons.

In order to contribute to the understanding of how the use of AWE programs affects the quality of the writing produced by ESL and EFL students, the present study collected usage and text data from the AWE software and applied some NLP analysis to understand how student's texts evolve as a response to AWE feedback.

The research presented in this chapter focuses on answering the first research question in this thesis, which is:

1. What changes do students make to their texts as a response to AWE feedback?
  - What revision behaviours do the students exhibit when interacting with an AWE tool?
  - How substantial are the revisions made to texts as a response to AWE feedback?
  - Do the changes made by the students improve the quality of the texts as measured through linguistic factors that correspond to human judgments of writing quality?

The methodology designed to answer the research questions that were the focus for this study can be found in section 3.2.

## **4.2 Results**

### **4.2.1 Student use of the AWE program**

Out of the 67 students that created an account on GrammarAid, 23 did not actually submit any texts to the platform, so they were eliminated from the dataset. That left 44 users submitting at least one text toward the platform. Of those, 28 users made changes to their texts after receiving AWE feedback. This means that a little under one-half of the participants used the feedback capabilities of the AWE program, while one quarter of the total participants did attempt to use the AWE program but did not find the feedback useful enough to make any changes to their texts.

**Table 14 - User actions in AWE program**

<b>Action</b>	<b>Number of participants</b>	<b>Percentage</b>
Made an account but did not submit texts	23	34%
Submitted at least one text but made no changes	16	24%
Submitted at least one text and made changes	28	42%
Total	67	100%

The students who used the AWE feedback performed various actions, often accessing their texts and viewing the summary report of feedback for them. In fact, most of the times, when students accessed one of their existing texts, they would also use the 'summary' functionality rather than simply making changes based on the feedback and then saving their texts.

The timeframe during which students worked on their texts and came back to revise them tended to vary quite a bit. A text can be considered active in the time between its creation and its last edit. The mean amount of time a text was active was of 65 days, although this measure is skewed by a small number of texts that were worked on for longer periods of time. The median amount of time a text was active was 9.3 days, suggesting that most texts were revised during the course of a week or so. However, some students continued working on their texts for over a month. While the exact nature of the texts submitted by the students cannot be known, it can be assumed that the texts which were worked on for a shorter amount of time were class essays and assignments, and those which were worked on for a longer amount of time corresponded to their dissertations and other longer-term assignments.

Of the students who did not make any changes to their text, however, only 5 students submitted a text and performed no further actions. Several students submitted more than one text to the

platform but made no changes to any of them. Participants 16, 18, 35 and 79, for example, submitted two different texts several days apart, but did not make changes to any of them and made no further attempts to submit texts to the platform. These results appear to show that most participants were willing to engage with the AWE program in the revision of their texts and were willing to give the program at least two opportunities, although some of them ultimately abandoned use of the program.

#### 4.2.2 Nature of text changes

The second research question sought to understand whether texts changed substantially as a result of AWE feedback. Participants made an average of 19 changes to their texts, an on average 83% of the changes were a direct response to the feedback provided by the AWE program, either by directly accepting the feedback provided or by making their own changes in response to given feedback.

Figure 14 shows a boxplot of the number of changes in each in text. Besides a couple of outliers, most of the texts received a moderate number of changes.

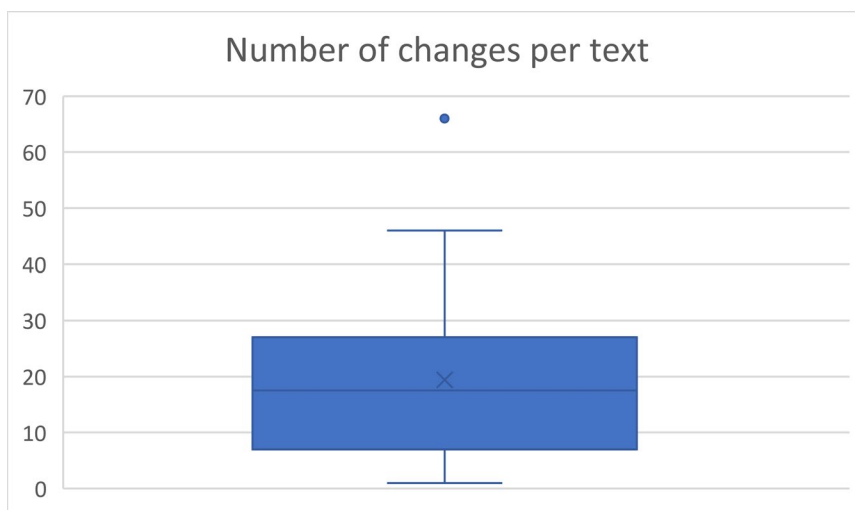


Figure 14 - Number of changes in each text

Besides not being many in number, the changes applied did not substantially alter the texts, as there was an average 10% difference in content between the last and first version of texts, smallest

difference being 3% and largest being 53.1%. Figure 15 shows a boxplot of the percentage of the text that was changed. There are three outliers whose texts changed significantly between the first and last version, but most of the changes were small alterations to the texts.

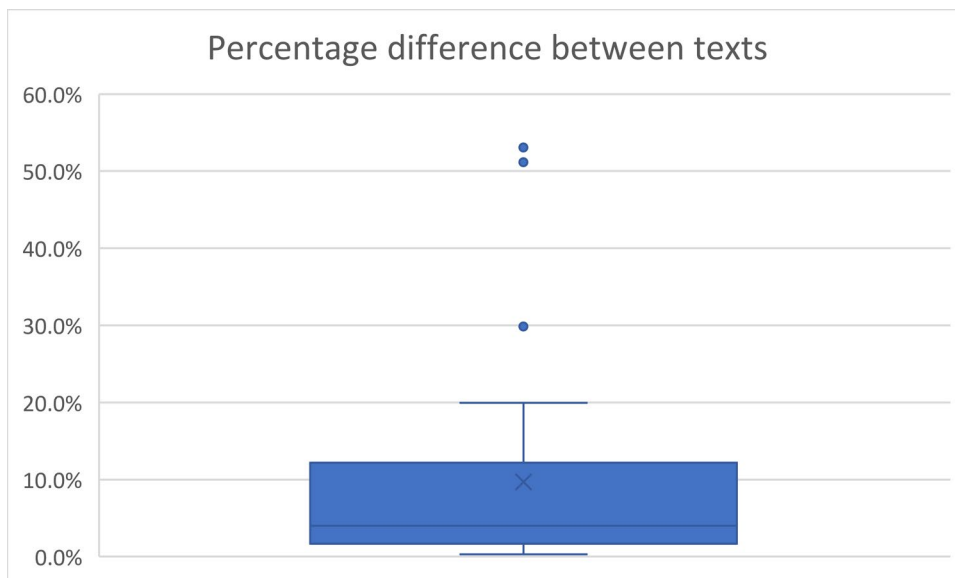


Figure 15 - Percentage difference between texts

This may be explained by the more mechanical nature of the feedback given by the AWE program. Table 15 shows the breakdown of all the feedback provided by the AWE program, presenting the total count of each feedback type as well as whether the participant accepted the feedback, rejected the feedback, or made their own changes in response to it. It is interesting to point out that, while most of the feedback was ignored overall (67% of all feedback was ignored by participants), there were some categories where students were more likely to accept the program's feedback than ignore it, although these corresponded to mechanical fixes such as spacing, repeated words and missing determiners.

**Table 15 - Breakdown of feedback by types**

<b>Feedback Type</b>	<b>Total count</b>	<b>Accepted (%)</b>	<b>Own changes (%)</b>	<b>Ignored (%)</b>
Agreement	24	6 (25%)	0 (0%)	18 (75%)
Capitalisation	3	1 (33%)	0 (0%)	2 (67%)
Determiner	38	19 (50%)	11 (29%)	8 (21%)
Passive verbs	478	22 (5%)	17 (4%)	439 (92%)
Prepositions	12	1 (8%)	0 (0%)	11 (92%)
Punctuation	179	26 (15%)	6 (3%)	147 (82%)
Readability	457	182 (40%)	17 (4%)	258 (56%)
Redundancies	9	0 (0%)	0 (0%)	9 (100%)
Repeated words	2	2 (100%)	0 (0%)	0 (0%)
Spacing	147	121 (82%)	1 (1%)	25 (17%)
Spelling	303	44 (15%)	12 (4%)	247 (82%)
Split infinitive	8	0 (0%)	2 (25%)	6 (75%)
Stronger wording	51	1 (2%)	2 (4%)	48 (94%)

The feedback types that were ignored the most in our sample were redundancies, passive verbs, prepositions, spelling and stronger wording. Most of the spelling feedback which was ignored was because the AWE program failed to recognise names or domain-specific terms (such as “interpretivism”), or because the program was suggesting the participant change spelling from US English to UK English. When it came exclusively to feedback about misspelled words, students accepted all of the program’s suggestions.

Interestingly, however, even when revising the text unprompted by the program, participants tended to make only superficial changes. Out of 85 self-initiated changes in the sample, only 11 either added new text or restructured existing text. Of the remaining changes, 7 corresponded to deletions, and the rest were mechanical in nature, such as adding missing determiners (N=30), modifying wording (N=22), or fixing agreement issues (N=8).

### 4.2.3 Quality of text changes

The third question sought to delve beyond the quantity of the changes made to texts and explore changes in quality. Six different measures of lexical characteristics were chosen because previous research has found correlation between these measures and human ratings of text quality. For each text in our sample, we calculated each score for the first version of the text (pre-revision) and the last version of the test (post-revision), then calculated the difference in scores between both versions to see whether there had been any change in the scores as a result from the revision process with the AWE feedback. Table 16 shows the difference between the scores obtained by the first version of the text and the last version of the text.

**Table 16 - Differences in scores for lexical characteristics.**

Text ID	Lexical diversity	Word familiarity	Word frequency	Word imageability	Word hypernymy	Meaningfulness
28	0.000310	0.25803	0.120788	0.919406	0.031703	1.978842
38	0.001763	1.425266	0.665976	2.113508	0.164866	6.472287
60	0.000181	0.024409	-0.02121	-0.18611	0.047658	-0.87324
61	0.000689	0.175992	0.427751	-1.41445	0.111749	-1.75443
62	-0.000176	-0.29598	-0.12974	-0.0724	0.012904	0.339347
63	-0.002154	-0.07724	0.017189	-1.01801	0.088087	1.366319
92	0.001146	0.167591	-0.21861	3.679602	0.010816	4.64104
93	0.000262	-0.5602	1.052812	-0.83596	-0.144569	-1.55058
94	0.000579	-0.10518	-0.31467	2.059692	-0.007361	1.580439
98	0.000028	-0.18588	-0.08551	-0.0267	-0.010736	0.27469
101	0.000977	-0.12035	-0.24539	0.525574	0.069101	1.360709
103	-0.000248	0.319841	0.357511	-0.37318	-0.050952	-0.96711
105	0.000070	-0.45792	-0.18172	0.267359	0.023288	0.013538



106	0.000057	0.217323	0.011535	0.904559	0.015670	1.428134
108	0.000016	0.168335	0.079245	0.208175	0.019081	0.054333
117	0.002130	1.015496	-0.05446	5.518602	0.214231	5.949709
119	0.000146	-0.16782	-0.12868	1.44484	-0.002951	1.007401
126	0.000094	-0.40038	-0.15451	0.235833	0.006170	0.0982
137	0.001514	-2.64197	0.422147	3.881146	0.473611	-1.78469
140	0.001885	-0.2705	-0.26803	-1.10233	-0.005797	0.139663
141	0.000099	0.369583	0.117088	-0.17712	0.000000	0.869271
149	0.000359	1.398137	0.261038	-2.11912	-0.256748	-0.04589
150	0.000014	-0.00862	0.018562	0.363835	0.022443	0.655106
171	-0.000166	-0.33288	-0.0516	0.600109	0.093501	0.391649
173	0.000026	0.19881	0.053239	-0.61199	-0.008782	-0.2165
178	0.000018	0.17336	-0.03094	-0.51179	-0.020441	-0.60472
179	0.000154	0.162994	0.072355	-0.46003	0.027946	-0.66056
186	-0.000391	-0.68035	0.207995	-0.78294	0.001373	-1.03294
187	0.000156	0.013425	0.094629	-0.63789	0.023695	-1.47973
190	0.000087	-0.20747	-0.03432	0.471616	-0.008130	0.619812

These indices for lexical sophistication have been found to correlate to human judgments of writing quality and proficiency, as it is believed that proficient students tend to use less familiar words, less imageable words, words which are not as frequently encountered, less concrete words and words which are less likely to activate other words (M. Kim et al., 2018). In my study, however, there was no clear indication that any of these factors was uniformly affected by the AWE feedback, with roughly half of the sample texts increasing and the other half decreasing their scores in most categories.

In order to test this further, a paired t-test was carried out for each of the lexical characteristics, comparing the difference between the first and second versions of the texts to see if it was statistically significant. The results are displayed on Table 17.

**Table 17 - Paired t-test for lexical characteristics**

	Pre		Post		t-test	p
	M	SD	M	SD		
Lexical diversity	0.0162	0.01247	0.0165	0.01287	-0.09806	0.9222
Word familiarity	587.317	3.05169	587.302	3.26237	0.01734	0.9862
Word frequency	9.621	1.10667	9.689	1.14538	-0.2362	0.8141
Imageability	307.7152	12.711	308.1440	12.9431	-0.12956	0.8974
Hypernymy	5.183	0.37474	5.215	0.38584	-0.31955	0.7505
Meaningfulness	329.6805	11.2543	330.285	12.0932	-0.20192	0.8407

\*\* p<.01 \*\*\*\*p<.001

According to the results, there is no evidence that any of the lexical characteristics under review were significantly impacted by revision using the AWE-generated feedback. This is not very surprising, considering how little the texts changed as a result of the feedback.

### 4.3 Discussion

One of the first insights that can be gleaned from this study is how little texts change as a response to AWE feedback. Existing literature has noted that students tend to ignore a substantial amount of the feedback they receive from AWE programs (Bai & Hu, 2017; Chodorow et al., 2010; Koltovskaia, 2020; Ranalli et al., 2017). The amount of feedback ignored ranges from 24% of feedback ignored (Link et al., 2020) to 75% of feedback ignored (Tian & Zhou, 2020). This study found that students ignored 68% of the feedback they received, which is on the higher end of the spectrum reported in the literature. One reason for this might be that, while most of the literature on the use of AWE focuses on in-class interventions, where students are required to use the AWE program to revise their work as part of their classwork, this study did not include any such requirements. Participants were free to use the AWE program for their own purposes without being required to do so.

This freedom of usage accounts for two interesting results obtained in this study. First is the fact that over half of the submitted texts (N = 40) had to be discarded because the participants made no changes to them as a response to the AWE feedback. This is in line with Attali's (2004) findings that 71% of the 33,171 texts submitted by 6<sup>th</sup> to 12<sup>th</sup> grade students presented no changes as a response to the AWE feedback. It could be assumed that, since the participants in this study were not required to use the AWE tool, they simply did not make any changes when they felt the feedback was inadequate. In contrast to this, however, is the fact that some participants continued to use the AWE program to revise their texts almost a year after the end of the workshop through which most of the data was collected.

Student willingness to use the AWE program for revising texts can also be deduced from their behaviours in using the AWE tool, even when they were not making any changes to their texts. As mentioned in the results section, many of the students who did not make any changes to their texts using the AWE tool submitted at least two texts, suggesting they would be willing to give the tool a second opportunity before presumably deciding the feedback offered was not useful to them.

The results of this study also seem to confirm one of the greatest fears surrounding the use of AWE programs to provide formative feedback or student grading: that its feedback tends to be mostly mechanical in nature and therefore the revisions students make tend to be superficial (Dikli, 2010). This leads to concerns that students are being trained to focus on mechanical errors rather than other aspects of writing such as conveying meaning, argument, genre conventions, and structure (Cheville, 2004). This is something that has come up in existing AWE research, as there seems to be "only modest evidence that AWE feedback has a positive effect on the quality of the texts that students produce using AWE" (Stevenson & Phakiti, 2014, p. 62), and research into this topic has often been fragmented and inconclusive (Hibert, 2019b).

Long (2013), for example, found only marginal improvement in confused words, spelling, missing commas, and short sentences when students used the Criterion program to revise their writing.

Chodorow et al. (2010) found that using Criterion and ESL Assistant resulted in error reduction, especially because participants tended to ignore incorrect feedback, but they only looked at mechanical errors such as article and prepositions. Similarly, Ranalli (2017) found that students could use Criterion feedback to correct errors 55-65% of the time, but also focused on mechanical errors such as confused words, extra commas, article errors, and preposition errors. The results of the study presented in this chapter seem to agree with this research, as they show that students ignore most of the feedback they receive from the AWE program, and what feedback they do accept is generally mechanical in nature and does not lead to meaningful changes in the text themselves.

This seems to be in line with previous research which found that students tended to find AWE feedback “vague, fixed, and sometimes repetitive” (Lai, 2010, p. 443); confusing and generally unhelpful to the revision process (M. J. Wang & Goodman, 2012); as well as mechanistic and unspecific (Xia & Zhong, 2017). While the current study did not specifically ask for student perceptions on the usage of the AWE program, it is worth mentioning that a lot of the texts collected by the web application designed for this study had to be discarded from analysis because no changes were made by the students, either as a response to the AWE feedback or their own revisions. It is reasonable to assume that this was because the students did not find the feedback given by the AWE program helpful, and therefore made no corrections or changes to their texts. In fact, out of 183 unique texts submitted through the web application, only one-sixth (N = 30) exhibited changes between the first and last version of said text, which suggests that in most cases participants did not find the feedback helpful enough to accept any of it. To further this idea, some of the texts were accessed by the students at different times and yet presented no changes, which seems to suggest they did not find what they were looking for in the feedback.

On the other hand, Liao (2016b) found modest gains in accuracy over time, and Wang et al. (Y.-J. Wang et al., 2013) did find significant improvements in human-assigned scores after an AWE intervention, although this last result can be attributed to pedagogical interventions and not

necessarily the AWE feedback itself, a conclusion shared by Lai (2010), Chen & Cheng (2008), and Tang & Rich (2017). This last point is especially interesting, as the use of AWE feedback in the current study was not accompanied by a pedagogical intervention. Even though the functions of the AWE program were explained to participants, they were mostly allowed to use it on their own time without any specific guidance due to time constraints and the format of the workshop, which was extracurricular and only lasted five sessions. This may explain the low uptake and overall superficial revisions that were observed.

#### **4.4 Conclusions**

The results of this study suggest that providing students with AWE feedback, by itself, is not going to help students revise their drafts. Overall, the changes made by students were superficial and responded more to mechanical errors than actual restructuring of the texts to better convey meaning or improve lexical sophistication, a feature that has been found to correlate to human ratings of writing proficiency. Furthermore, most of the texts submitted to the web application did not present any changes, even when the text was accessed by the students more than once, suggesting students found the feedback unhelpful.

Tools are only useful when people know how to make good use of them, as tools can only be successfully used by students when they are aware of the occasions where a tool can be used and are able to use it skilfully (Winne, 2006). Previous studies have suggested that pedagogical interventions are needed to guide students into making good use of AWE feedback (C.-F. Chen & Cheng, 2008; Lai, 2010; Tang & Rich, 2017), as revision skills are important regardless of the source of feedback.

The main limitation of this study was that the use of the AWE program was not implemented as part of a pedagogical strategy, as there was no institutional support to carry out this study as part of a writing program offered to students. Instead, students were recruited to an extracurricular workshop lasting a total of 5 sessions, in which there was little time to implement a pedagogical

strategy, and the extracurricular nature of the workshop meant the participants preferred to prioritise their academic workload over the workshop's assignments. Another important limitation is that participants were not interviewed about their usage of the platform as part of the data collection for this study, and therefore any assumptions about student motivations are speculation. However, it is valuable to examine in depth the changes that were made by students as a result of AWE usage, even if those changes are limited and only resolve mechanical mistakes within the text. It gives further theoretical support to the notion that AWE programs by themselves cannot help ESL students to improve their writing skills or even their texts by using the program without teacher support. Future studies might benefit from exploring different ways in which this pedagogical support can help students make better use of AWE tools, instead of focusing on the effectiveness of the tools by themselves.

While these results shed light on the way texts change as a response to AWE feedback, the limitations of this study mean that we can only speculate about the intentions and thoughts of the participants. In the next chapter, I report on the second study that aimed at addressing some of the limitations on this study by reporting the results of a separate study carried out to understand the decision-making processes that lead students to decide whether they accept or reject the feedback they receive. By using qualitative data gathered from think-aloud protocols and screen recordings, we can add depth to the findings of these chapters by delving into the *why* of feedback uptake and the process of revision itself.

## Chapter 5 – Student engagement and decision-making

### 5.1 Introduction

As we mentioned in the literature review, there has been little research into AWE usage on how students engage with these programs or what self-regulation strategies they use when revising their texts using AWE feedback. I discussed how engagement is an important concept in understanding the impact of AWE feedback on student's writing skills, but it has been understudied in current research on the use of AWE programs. Zhang and Hyland (Zhang & Hyland, 2018) and Zhang (Zhang, 2017) have done the most extensive work on student engagement with the feedback provided by these programs using Fredricks, Blumenfeld and Paris's (2004) tripartite conceptualisation of engagement, understanding the concept as encompassing behavioural, emotional and cognitive dimensions, and found that "engaged learners thus tend to spend more time working with feedback, show more positive attitudes toward it, and employ more revising strategies, demonstrating that behavioral, affective, and cognitive engagement dynamically interact with each other" (p. 100).

Regarding self-regulation, in the literature review I discussed how this could be a useful framework in understanding student engagement with AWE programs, and especially cognitive engagement, given that cognitive engagement has been regarded as a key component of self-regulated learning (Wolters & Taylor, 2012), while the use of regulation strategies is one of the components that define cognitive engagement (Fredricks et al., 2004). Therefore, it seems appropriate to examine student engagement with the use of AWE programs through the lens of self-regulated learning.

While there have been some calls for using think-aloud protocols to examine student reasoning behind their revision decisions and engagement with AWE programs (Link et al., 2020; Zhang, 2020), to my knowledge, only Cotos (2012) has used this tool to examine how students use these programs. This study therefore contributes to expand our knowledge of how students engage with AWE

programs by using think-aloud protocols and screen captures to examine student engagement and understand the decision-making process behind feedback uptake.

The research presented in this chapter focuses on answering the second research question in this thesis, which is:

2. How do students engage with an AWE program?
  - What, if any, self-regulatory strategies do students employ when engaging with AWE feedback?
  - What decision-making process do students use to decide whether to accept or reject feedback given by an AWE program?
  - Do students' decision-making processes help them make good decisions when accepting or rejecting AWE feedback?

The methodology designed to answer the research questions that were the focus of this study can be found in section 3.3. All names reported in this chapter are pseudonyms.

## **5.2 Results**

### **5.2.1 Self-regulatory strategies**

During their think-aloud sessions, participants displayed many self-regulatory strategies, although the most used strategies related to metacognition. 43 out of the 44 total sessions were used for this analysis, as one of the participants felt uncomfortable doing the think-aloud during the last session and therefore there was no data on their thought processes for that particular session. Figure 16 shows the global frequency of the use of these strategies.



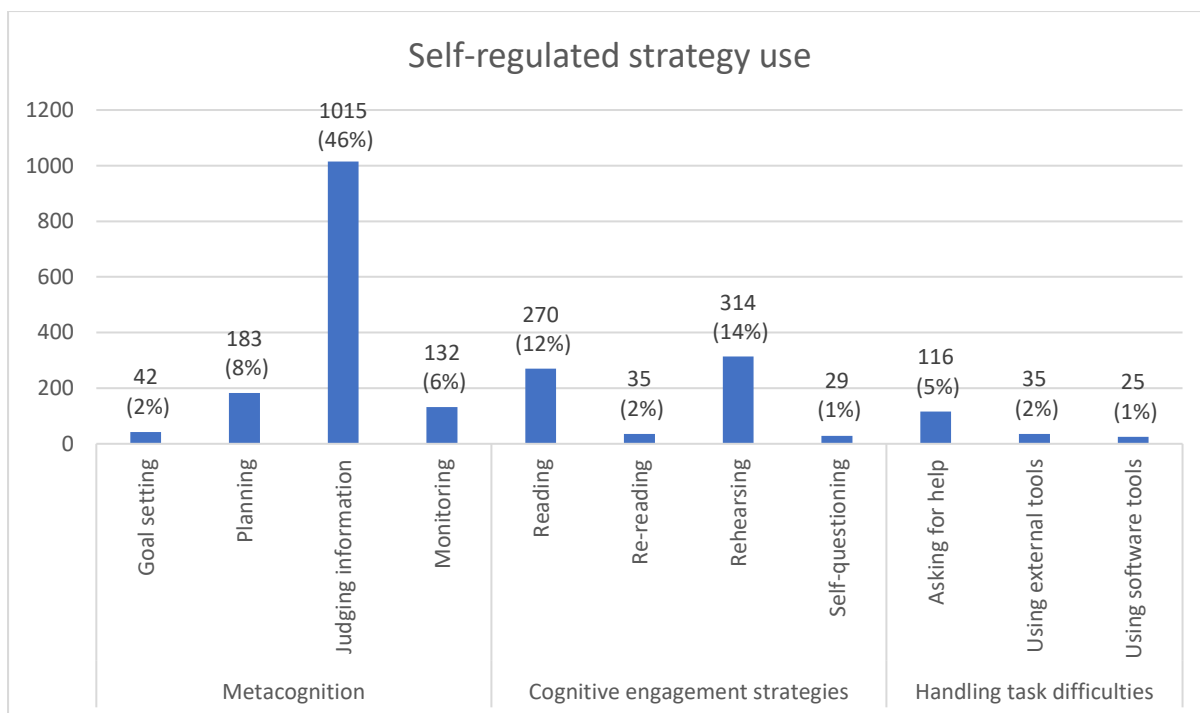


Figure 16 - Self-regulated macroprocesses

As can be seen in Figure 16, the most commonly expressed self-regulation processes were those related to metacognition, with cognitive engagement strategies in second place and handling task difficulties third. In the following subsections, I will be describing in detail how each of these strategies was used by the participants to get a better understanding of how they engaged with the AWE program.

**5.2.1.1 Metacognitive strategies.** Students tended to begin their sessions by establishing plans for what they would do during the session (27 out of 43 cases). There were three reference points used by the students when establishing these plans. As indicated by the use of software tools that can be seen connected to planning in the process map, the first was the “Summary” section provided by the program, which gives a summary of all the feedback the program can give and assigns scores to grammar, spelling and style. Several students used these scores as a way to figure out which parts of

their writing needed more attention, like Daniella<sup>4</sup>: “Wow... for less than 1,000 words, 82 spelling problems... So I check grammar first”.

Starting from the second sessions, many of the students made their initial plans for the revision session by considering the feedback they encountered in previous sessions, like Melissa: “So last time I went through the... I think grammar part, so this time I’m going to start with Overused”. Third, students made their plans based on an overall assessment of their writing, like Violet: “First, I will look at the grammar, because it always um... it’s the worst part of my dissertation.”

These planning utterances were not only confined to the start of their sessions. Whenever the students would finish using one of the program’s functions, they would either go back to the summary to look at other possible functions they could use or looked at the list of available functions to try to determine which could be the most useful. Owen, for example, decided to move to grammar once he was finished with the style check, stating that “I always make lots of grammar mistakes”. Not all students did this, however, as some, like Melissa, simply went through all the functions in order without having a concrete plan and decided on the fly whether each function would be useful to them or not.

Perhaps unsurprisingly, the most common microprocess reported in our sample was judging information as the students attempted to judge the quality or relevance of the feedback given by the AWE program. The process through which students judged the feedback they received from ProWritingAid will be expanded upon in section 5.2.2 and the results of this process will be explored in section 5.2.3.

In most cases, however, students judged that the feedback the AWE program gave them was not applicable to their situation. For example, Rachel evaluated the feedback given to her to delete discourse markers such as “firstly” and judged it inapplicable to her situation: “Sometimes it suggest

---

<sup>4</sup> All names reported in this chapter are pseudonyms

me to delete the discourse markers. I think um... I will not take this kind of advice because that might change my... how to say? change my cohesion". In other cases, students would go through an almost dialogic process with the feedback given by the AWE program to judge the information received, as this extract from Iain's think aloud demonstrates: "Contrit... contractinoary... what? You don't know the word? How come you don't know? Contradictionary... contra... Oh, yeah, I spelled it wrong".

Another metacognitive strategy used by the students was that of monitoring. Students monitored their revision by using their supervisors' feedback as a standard and by comparing their writing to other academic texts. For example, the program's feedback on repeated sentence starts helped him monitor the state of his text against previous advice given to him by his supervisor:

"The program tells me that I started three sentences in a row with the same word like I start a sentence with the word 'a'. I think this is interesting. Not only because here but also sometimes my supervisor also points out that I like to use 'and' to start a sentence."

Owen and George likewise used their supervisors' comments to monitor their revisions, in their case previous feedback given to them about their issues with writing sentences that were too long. Both participants spent a long time using the "sentence length" function so they could revise their texts according to the standards set to them by their mentors.

Other students monitored their progress in revision by comparing the AWE feedback, and their own writing, to academic texts they had read in the past. Cristina, for example, was confused by ProWritingAid's continued insistence that she use less passive voice in her writing. She used Google to find an article that explained how to use passive and active voice in academic writing and used that knowledge to monitor the state of her writing and decide whether she needed to remove instances of passive voice or not.

Students also set themselves goals for future revision sessions, which mainly fell under two categories: looking for better vocabulary to express their ideas and finding new ways of rephrasing things, and making note of common issues pointed out to them by ProWritingAid so they could keep them in mind when further revising the text without the program. The first kind of goal setting tended to happen when students did not feel like they had the skills or the time to carry out a specific revision during the session, and therefore they made notes for themselves to go back to that part of the text later on.

For example, when the AWE program pointed out her overuse of vague words, Melissa reflected that “I may need to spend more time on it, because from the... for the explanation of this program, I might need to use word explorer to find another word whose meaning is similar to the vague word”, suggesting that she didn’t have the time nor the knowledge to make the correction at that moment. Alan also cited time constraints when considering limiting his number of adverbs: “I don't think I can keep it a reasonable number of adverbs in my writing in a short time, so I think I'll check the use of adverbs later, and now I'll move to other aspects”.

The program helped Owen realise that he had a problem with using vague words in general, and he made a note of the program’s feedback so he could take it into account when revising future texts. George also made note of several sentences that he needed to restructure in the future and stated that he needed to keep his overuse of clauses in mind when producing future texts.

**5.2.1.2 Engagement strategies.** Engagement strategies were also commonly used by the participants to engage with the feedback and help students decide whether and how to apply the feedback they received from the AWE program. These strategies included reading parts of the text they wrote, re-reading parts of the text to enhance comprehension. rehearsing different ways of applying feedback or rephrasing sentences, and self-questioning their intended meaning in light of the AWE feedback.

Reading snippets of text for comprehension was the second most commonly used strategy. Students tended to stop and read entire sentences, paragraphs or passages in order to understand the context of their writing to decide whether to accept the feedback. Sometimes merely reading things out loud was not enough for them to reach a decision, so they re-read the same passage several times in order to contextualise the AWE feedback and try to understand what their original intended meaning had been.

The most commonly used engagement was rehearsing, which students used to see whether applying the feedback offered by the AWE program would fit in their writing or whether it “sounded right”. This happened especially if students were not completely sure about the feedback they were receiving from the AWE program. For example, in the sentence “The first time I asked my kid whether he want to study here or not”, Alice was told by the AWE program to change “want” to “wants”. She rehearsed both options by saying them out loud and eventually decided to accept the feedback. Likewise, when ProWritingAid suggested Cristina change “failed to” to “didn’t” in the sentence “educational research failed to cure”, she rehearsed both options and decided not to make any changes because her wording sounded more “academic”.

Finally, self-questioning happened when the feedback forced them to look at their writing and question what their intended meaning was. This was the least used engagement strategy, as the students seemed to be clear in what they were trying to convey with their writing. For example, Alan initially disagreed with the program’s suggestion that he add a determiner before one of his nouns. The original sentence read “The other is balanced corpus or sample corpus approach”, and Alan first reflected “I don't think I need to add the article here because I use "approach" later”. However, he almost immediately began questioning whether his intuition was wrong, and eventually ended up adding “the” before “balanced”.

**5.2.1.3 Handling task difficulties.** Another component of self-regulation that was coded for in this study was the handling of task difficulties. There were three types of utterances that fell under this

broad theme: asking for help, using external resources to solve questions or clarify concepts, and using the tools provided by the AWE program to understand the feedback provided.

Students tended to ask for help from the researcher during the sessions for two main reasons. The first of these was because they had questions about the technical functioning of the AWE program, and the second was because they needed help understanding the feedback they received or weren't sure how to carry out a specific revision.

Jaime asked several questions about what each of the functions of ProWritingAid did, such as the "All repeats" function and the "Readability" function. She also asked questions about the meanings of the different-coloured lines used by the program to indicate different types of feedback (see section 3.1.3 for examples). Alan also asked for help with changing the revision language from British to American English so the program would stop flagging his use of American spelling as an error.

However, the overwhelming majority of instances where the participants asked for help happened when they did not understand the feedback they received from the AWE program or were not sure how to carry out the revisions it suggested. Iain wasn't sure whether he should take the program's advice on reducing his use of passive voice and asked the researcher for her opinion on the matter. Susan also asked for help, often in understanding the feedback she received. For example, she did not understand what the word "sparingly" meant, and so when ProWritingAid suggested she used adverbs sparingly, she turned to the researcher to ask what that meant. Given that Susan asked for advice at a considerably higher rate than other participants, I will be looking more in-depth into her help-seeking behaviours in section 5.2.1.5.

Some students turned to external sources of knowledge, such as Google or multilingual dictionaries, to make sense of the feedback they received. The queries could be as simple as confirming the spelling of words, like Daniella did when the program suggested she change "stage" into "stag", and looked up the definitions of both words to make sure she used the correct one. Participants also used Google to make sense of the feedback they received. For example, Cristina used Google to look

up whether active voice or passive voice was better for academic writing, and found an article titled “How to effectively use active and passive voice in academic writing”, which she used to determine whether to take the AWE program’s suggestions on reducing the amount of passive voice she used.

**5.2.1.4 Student attitudes.** Student perceptions and affective reactions toward ProWritingAid were mixed. Many of them acknowledged the usefulness of the program when revising mechanical aspects of their writing. Rachel, for example, mentioned that she was sceptical at first, but the more she used the program, the more she noticed recurring issues in their work: “it points out the things that I’m never aware of”. Alan also pointed out that “it’s good to be made aware”, and George praised the program for helping him solve his recurring problem of writing sentences that were too long by using visual feedback to help him find which sentences needed fixing (I will come back to George’s specific case in section 5.2.3).

However, negative attitudes toward ProWritingAid were expressed because some students seemed to expect more substantial help from the program. Susan, for example, complained that “sometimes it was wrong because it’s a machine and not intelligent enough. Yeah, it was positive, my impression was positive, but I was expecting more”. Cristina also noted that: “Does it give any suggestions? No, it just says overused something but it’s not suggestion why should I modify those writing. Actually, I expected that it would.” This disconnect between student expectations and the reality of the feedback provided by ProWritingAid was one of the most common sources of frustration in the transcripts.

As evidenced by Cristina’s quote in the previous paragraph, ambiguity of the feedback was another reason why the students seemed to be dissatisfied with the feedback given by the AWE program and expressed negative attitudes towards it. Nancy complained “this is hard to use” when faced with ambiguous feedback, and Owen also pointed out that the program “did not tell me how to revise it or change it, or giving some reasons. There’s something wrong, or it’s good, it’s positive or negative. I don’t know”.

Affective reactions were therefore dependent on students' ability to make sense of the feedback and their knowledge of how to carry out revisions based on said feedback. They expressed positive opinions regarding the program's ability to help them notice recurring mistakes or fix mechanical errors, but also expressed that the support they received from the program was insufficient and they expected to receive more help from it when revising their texts.

**5.2.1.5 Individual use of self-regulated strategies.** When looking at self-regulated strategies individually, it seems that most of the students used a similar number of strategies, and there are few outliers in the use of these strategies. This suggests that self-regulation is very similar across the sample, and therefore participants were very similar in their use of self-regulated strategies. Table 18 shows the individual number of self-regulated strategies employed by each student. In this section, I will look individually at the outliers in each of these categories to understand how they used these strategies to gain a better understanding of how students engaged with the AWE program. The outliers (highlighted in the table) have been chosen because their general usage of these strategies are similar to the rest of the participants, but their increased use of these strategies helps paint a clearer picture as to how they were used and how they can be developed, which I will discuss in Chapter 6.

**Table 18 - Individual use of self-regulated strategies**

<i>Microprocess</i>	Melissa	Alice	Jaime	Alan	Daniella	Nancy	Owen	Susan	George	Cristina	Iain
Goal setting	6	1	1	8	17	0	5	0	1	2	1
Planning	27	5	5	6	22	19	47	1	30	16	5
Judging information	66	58	12	55	122	46	178	16	243	146	73
Monitoring	16	14	0	16	14	7	19	1	18	20	7
Reading	22	34	0	1	34	31	29	4	28	65	22
Rewriting	0	0	0	0	1	0	0	6	0	9	1
Re-reading	0	0	0	0	2	1	6	0	12	9	5
Rehearsing	2	39	0	0	14	20	18	0	140	64	17
Self-questioning	0	8	0	1	3	5	1	0	3	5	3
Asking for help	0	0	5	11	1	5	2	55	2	1	34
Using external tools	4	4	1	0	7	11	4	0	0	4	0
Using software tools	4	4	0	0	0	9	4	0	0	3	1



According to the data, Daniella appears to be an outlier in that she set more goals than the rest of the participants in this study. She tended to mark parts of their writing that needed to be expanded on, or parts of her writing that she did not have the knowledge or tools to revise. For example, she remarked that some of her writing was not “scientific” or “proper” and made a note to herself so she would Google better terminology later. She also tended to mark for later the clarifying of certain terms in her writing, such as changing “British English” to “Chinese English” and “Chinese learners” to “ethnic Chinese learners” to better reflect the nature of her participants. In comparison, Susan and Nancy did not set any goals for further revision, and Alice, Jamie, George, and Ian each had just one instance of goal setting. This suggests that few participants made wider goals for later revision of the texts using the AWE feedback, even though the setting of these goals helps students become more aware of recurring issues in their writing and helps them develop their knowledge of how to revise texts in the future (Jiang & Yu, 2020; Palermo & Thomson, 2018).

George proved to be an outlier when it came to rehearsing. Rehearsing different possibilities for each sentence and saying it out loud helped him, for example, notice that he was using false cognates from Spanish, and this was causing grammar errors in his writing: “‘Conformed’ .... Confirmed? Ah, no, that’s... I think in Spanish I wanted to say ‘conformar’, like... ‘group’. ‘The participants teachers’ ... ‘formed’, yeah, that’s a better word”. By rehearsing, he also identified common problems in his writing: “‘To then identify’ ... Ah, okay, here is the split infinitive. This one also... always... always... This always happens. ‘To identify their beliefs’ ... Okay. ‘These instruments are suitable as they mmmm to identify their beliefs’. Okay. Yes.” This constant rehearsing of possibilities is perhaps one of the reasons why George was the participant who accepted the most ‘good’ feedback in our sample, something I will come back to in sections 5.2.3 and 5.2.4.

Susan and Ian were outliers in their asking for help. Susan, in particular, asked for help more than anyone else in the group. Her difficulties related not only to the use of the AWE program, asking

about different functions, but to the usage of the computer in general, with questions about being able to save and using shortcuts such as ctrl+v. This suggests that she had low computer literacy, which caused her additional problems when using ProWritingAid for revision.

However, she also had problems with understanding the feedback itself and with organising her writing. For example, when looking at the amount of feedback she received, she asked: “I think it would take maybe five or six hours if I tried to do all. How should I do this?” instead of prioritising certain areas of her writing like Daniella, Melissa, or Violet, who were mentioned in the previous sections. She often asked for clarification of the feedback or suggestions on how to implement it in her work, as well as asking the researcher whether the issues marked by the AWE program “were really a problem”. Susan did not feel confident in her own judgement when it came to evaluating AWE feedback, but neither did she use the software resources or external resources to find answers to the questions, instead relying on the researcher. Asking the researcher rather than using available resources might have also been related to the lack of computer literacy she displayed during the revision sessions.

Iain seemed to have similar issues as Susan, often asking for help with the technical aspects of using ProWritingAid. He also preferred to ask the researcher for clarification about the AWE feedback rather than using the tools the software offered or using external tools to research the feedback he was receiving. However, he did seem to show more confidence in his own judgement of the feedback provided by the program, asking the researcher questions to confirm his intuition: “Which one do you think is correct? [...] ‘Leave it’, right?”.

Finally, Nancy seemed to be an outlier when it came to using external tools. Very few participants used external tools to make sense of the feedback they received from the AWE program, ranging from 0 to 7 compared to Nancy’s 11. Nancy mostly used Google to look up definitions of words she didn’t understand, especially words that were being suggested to her as replacements by the AWE feedback. She also used Google a couple of times to look up the spelling of words when she wasn’t

sure if the AWE feedback was giving her the correct spelling. A similar pattern was found in other participants. Jamie and Cristina, for example, used Google once to check the spelling of a word they were unsure of, although Cristina also used Google to deepen her understanding of how passive verbs should be used in academic writing after being told by ProWritingAid that she should limit her use of the passive voice.

This analysis of individual cases gives a picture of how the participants used self-regulation strategies to engage with the AWE program. These cases showcase how some of the self-regulation strategies were being used by participants, and even though we only examined some outliers, the rest of the participants show similar patterns in using self-regulation strategies to engage with the feedback they received. Therefore, it can be said that participants were fairly uniform in their usage of self-regulation strategies.

### 5.2.2 Decision-making processes

As discussed during the data analysis section, coding the think-aloud transcripts yielded a total of six main themes which gathered fourteen codes related to the decision-making processes used by participants when engaging with AWE feedback. The frequencies can be found in Figure 17.

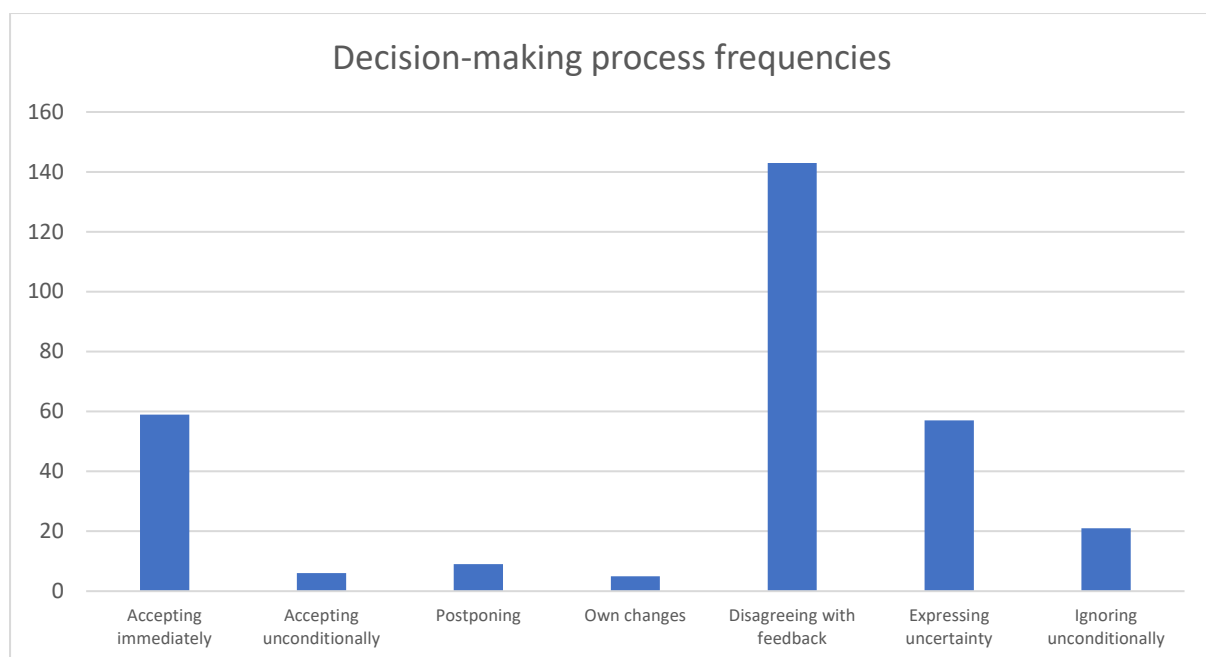


Figure 17 - Decision-making utterances

While students immediately or unconditionally accepted feedback because they thought it was correct, one initial insight from this coding process was that participants had a wide variety of reasons for deciding to ignore the feedback provided by the AWE program. Participants not only ignored feedback that they disagreed with or deemed incorrect, they also tended to ignore feedback they were unsure of. I will explore these reasons for disagreeing with feedback in this section and, through the analysis of these findings, a general decision-making process emerged from the data.

**5.2.2.1 Disagreeing with feedback.** As can be seen in Figure 17, students disagreed with most of the feedback they received from the process. Examining the think-aloud data reveals a variety of reasons for this. This could include that accepting the feedback would change their intended meaning, that the suggestion didn't fit with their personal writing style, that the feedback did not relate to issues important enough to fix immediately, or that the program's suggestions were not 'academic' enough. Analysing the reasons why students disagreed with the feedback also revealed that students' decisions to disagree with the feedback were informed by a variety of resources.

One important reason why participants would disagree with the feedback given by the AWE program if they perceived that it changed their intended meaning. For example, when Melissa was suggested to remove 'relatively' to keep the number of adverbs in their writing down, she disagreed: "I cannot say that my research is not... is fully comprehensive, so I... I'll just keep the 'relatively'."

Rachel also had similar assessments of some of the feedback he received: "it gives me some advice, I think um... it is good but if I make changed so the sentence might not express what I want it to express. So sometimes I think I'll leave what I have written and not make any changes". A common complaint among the students was that the program tended to misinterpret what they meant, something clearly expressed by Violet: "I want the emphasise the responsibility so I will remain this. Sometimes it will mistake what I intended to mean". The same feeling was shared by Cristina: "It don't understand what I'm saying, so it's suggestion is useless."

In general, participants had a good idea of what they meant to write and showed that their selection of certain terms or phrases was quite deliberate, as shown in the following excerpts from the think-aloud:

Violet: I would like to use quite, emphasise the degree, but it ask me to delete it. So I would like to remain it.

Melissa: Commonly? Mmm this program suggests me to delete this word 'commonly', but I think 'commonly' is um... is very important because uh... it shows that the first type is um... the most like... widely used type, so...

Rachel: And the first case it suggest me to use comma between Chinese and English. But actually I want to express that Chinese in... "Chinese English major undergraduates conversation". I think I use both the word... both of the word as adjectives. So you can ignore this.

George: But if I delete the 'freely', I feel it doesn't quite fit. It doesn't... not the sense I wanted to give it. I'll just leave it there.

Cristina: I... I'm using "in order to" because I want to explain that's my purpose, so I'm not going to delete it.

In a similar vein, participants seemed to have developed a personal style and tended to ignore the feedback they perceived as interfering or changing said style. Jamie, for example, disagreed with the readability feedback that suggested simpler words: "sometimes maybe I use some more complicated words and it recommends me don't use that, just use some simple words... Sometimes I find it's useful, but sometimes I want just... I won't delete it." Rachel had a similar issue: "I find it gives me some kind of alternatives um... with more uh... with easier words, like it suggests me to use "used" to replace "employed". Um... I don't think I will make changes here."

There were also occasions where the participants acknowledged the AWE program had valid feedback to offer but did not consider it a pressing enough issue to make any changes. Violet, for example, ignored most of the capitalisation offered by the program: "... I think capitalisation, this part is kind of tricky because sometimes I will use capital letters but no, sometimes no. So there's

nothing to worry about.” Melissa also had the same opinion for some readability enhancements suggested by the program: “It suggests me to change the word "indicate" to the word show, but I think um... this is not a problem, so I'm just going to ignore this one”.

One of the most interesting themes that emerged from this analysis was that participants drew on a variety of sources of previous knowledge to make a decision. These sources include academic papers read by the participants, advice they have received either from their dissertation supervisors or teachers they have had previously, and their knowledge of English grammar received from classes they have taken in the past.

For example, many students seemed to have a clear idea of what “academic writing” looked like from reading journal papers and other materials in their class. When the feedback offered by the AWE program clashed with their conception of what academic writing should look like, they tended to ignore it. Melissa, for example, complained that “this program keeps telling me to change the word 'indicate' to the word 'show', but I think the word 'show' is kind of word that um... not to be used too often in an academic essay”. The same logic led Rachel to disagree with feedback related to reducing the use of passive verbs in his writing: “I don't agree with the suggestions on passive verbs, because I'm thinking in academic writing sometimes we need to use the passive verbs to express that we are expressing our opinions”.

Supervisors and teachers were also a frequent source of previous knowledge. When using the tool provided by the AWE to visualise sentence lengths, Violet remarked: “I... my sentence structure is really low, it's just 10 words. And my supervisor also pointed out this problem, said I should make my sentences longer so that it's... maybe have higher English proficiency”. George, however, had the opposite issue: “And they've told me, that my sentences are too long, because I try to put too many ideas in them, or using words... not that complex, but words that could be said... the same message, but in a simpler way. So... that's what I'm seeing here.”

**5.2.2.2 Expressing uncertainty.** In general, participants seemed to expect the program to offer concrete solutions to their writing problems and got frustrated when they did not receive it. “I don’t understand” and “I don’t know what to do with this” were common complaints from the students as they tried to navigate this ambiguous feedback. This frustration had a strong effect on the uptake of the feedback, which will be explored in the results in section 5.2.3.

Related to the previous point, sometimes the students did understand the feedback even though it offered no concrete suggestions but doubted their ability to make any meaningful changes to their texts in response to said feedback. Daniella clearly expressed this frustration: “I don’t really know if I should make the change, when the work already represents the highest level of me. I need to find someone else to fix this. I’m going to ignore this right now. Let my proofreader worry about it.”.

Other times, even if the AWE program offered specific corrections to the participants, they were not always sure of the quality of said feedback, which also led them to reject it. Not being able to judge the quality of the feedback correctly also led participants to ignore it. When offered to change “on the basis of” to either “based on” or “because of” to improve readability, Daniella re-read the passage and concluded: “I don’t know if it’s a problem so... ignore”.

Several participants, however, used external tools such as search engines to look up specific feedback they were unsure of, especially when they suspected the feedback may be wrong but were not quite sure. For example, when the program flagged Alice’s spelling of “Kungfu”, she was sceptical: “Kungfu? Kungfu? I’m not sure how this is spelled... it is right”. Nancy had a stronger opinion when faced with similarly erroneous feedback: “Homeopathy... wrong? Let’s just check. That’s ridiculous.”

**5.2.2.3 Ignoring unconditionally.** As I explored in the last section, one of the main complaints of students was that the AWE program did not always give them a clear direction on how to revise their writing. In cases when students faced a piece of feedback that was ambiguous or hard to understand, they tended to ignore it rather than attempt to make sense of it. For example, the

readability section highlights passages and sentences that the algorithm has deemed 'hard to read' using the Flesch Reading Ease score but gives the students no guidance other than stating: "This paragraph is hard to read. Consider revising it". Melissa expressed her frustration with this ambiguity: "I think it's quite helpful for me to have a very general overview of my readability, but I think I'll have to see more specific details why it's very, very hard to read. So I'll just move to the second part."

Another common feedback that gave students little guidance was centred on the split infinitive, with the program issuing a message that read "some people strongly object to this type of construction, especially in formal writing. If your reader is likely to object to the split infinitive, rewrite the sentence to avoid its use". Faced with this feedback, Rachel said: "it suggests to rewrite the sentence if your reader don't like this kind of grammar, but I think if they are some suggestions on how to revise for or what kind of alternatives can provide... I think if there are more suggestions it would be better." Later, when talking about a different piece of ambiguous feedback, she elaborated on the same idea: "it makes sense but I don't know which word I can choose to replace these three words. So I think it is good to raise the awareness, but maybe the user needs more advice, I think."

George appeared similarly concerned with his own ability to revise his text based on the feedback, specifically for the use of passive voice: "The thing about passive voice... it says that it makes writing less direct. And that's an active verb. But there are things that sometimes I can't figure out how to make it... how to turn it around". Iain had a similar issue with changing their verbs from passive to active: "Passive verbs make your writing less direct. Really? Try to use an active verb instead. How am I supposed to change this around?" This lack of confidence in their own abilities led students to reject the feedback they received.

**5.2.2.4 A model for decision-making.** In general, by examining these different strategies used by the students to engage with the feedback, we can start to intuit a decision-making process. First, students tend to look at whether they agree with the feedback. If it's a piece of feedback they've



received before and consistently agreed with, they tend to accept it automatically whereas, in other circumstances, they stop to ponder the merit of said feedback. If they disagree with the feedback, for any of the aforementioned reasons, then they either go straight to ignoring the feedback or look up the feedback in order to reach a decision. The resulting process is detailed in Figure 18.

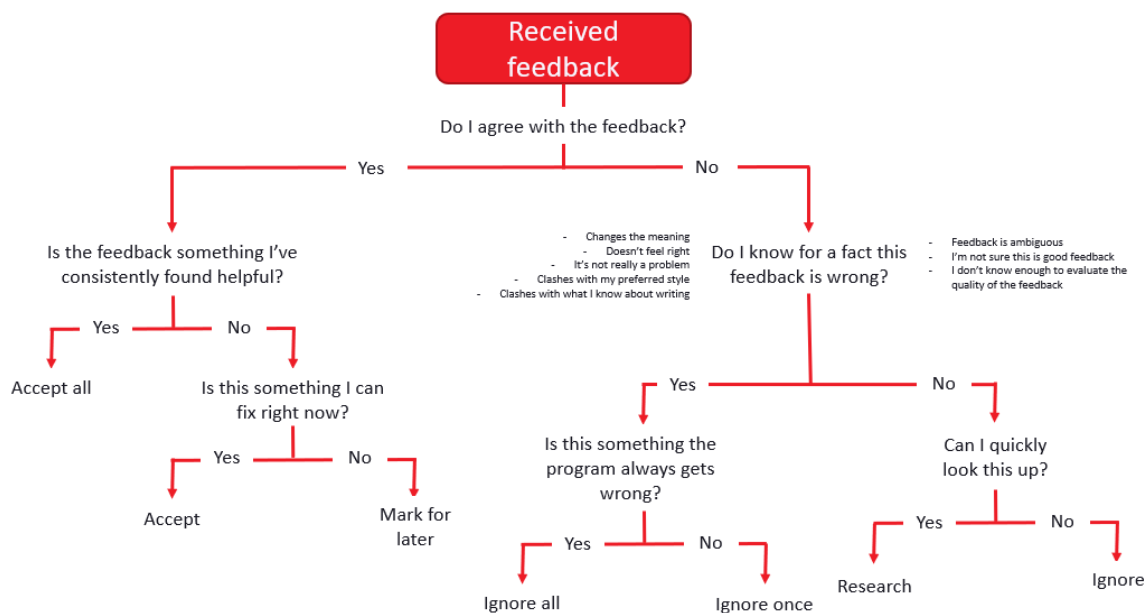


Figure 18 - Participant decision-making process

The result of this decision-making process is that students, in general, tend to make good decisions regarding their uptake of feedback. This will be discussed in the next section.

### 5.2.3 Quality of uptake

In the previous two sections, I talked about self-regulation strategies students used to engage with the AWE program and the decision-making processes with which students judged the AWE feedback to decide whether they accept or reject it. So far, the analysis shows that students use many strategies to engage with the AWE feedback, and that they draw upon a variety of resources to judge the feedback they receive. In this section, I will explore the results of this engagement and decision-making process. That is, I will look at the changes students made in response to the AWE

feedback and analyse the quality of these revisions to see what the product of this self-regulation and decision-making looks like.

After running frequency analyses on student uptake, one interesting thing to note is that most of the feedback (n = 68%) was ignored by the students, as shown in Figure 19.

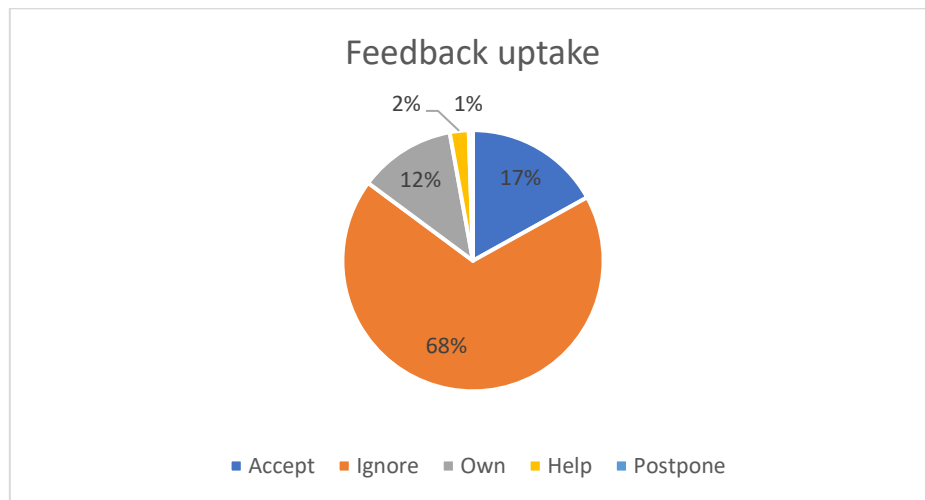


Figure 19 - Student uptake of AWE feedback

One probable reason for participants ignoring over two thirds of the feedback they received was the quality of the feedback itself. 30.5% of the feedback given by the AWE during the revision sessions could be considered “bad”, while only 5.2% of the feedback could be considered “good”. The rest of the feedback was either “neutral” (27.6%), “mixed” (8%), or the program flagged a possible issue but offered no concrete corrections (28.7%).

As discussed in the previous section, students tended to get frustrated when they received no specific feedback, so the large proportion of “no feedback” or “bad” feedback in the sample might account for the high percentage of feedback that was ignored by the students. This seems to be corroborated by the fact that, when analysing student uptake per quality of feedback, it is very clear that students tend to accept more “good” feedback, while ignoring most “bad” feedback, as can be seen in Figure 20.

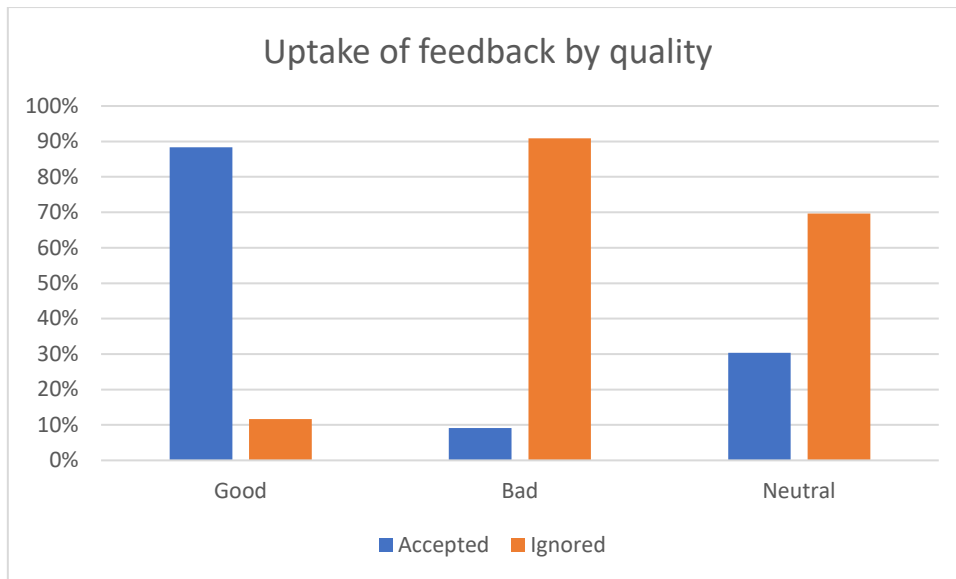


Figure 20 - Feedback uptake by quality of feedback

It was also found that the uptake of feedback was quite consistent among participants, as can be seen in Figure 21.

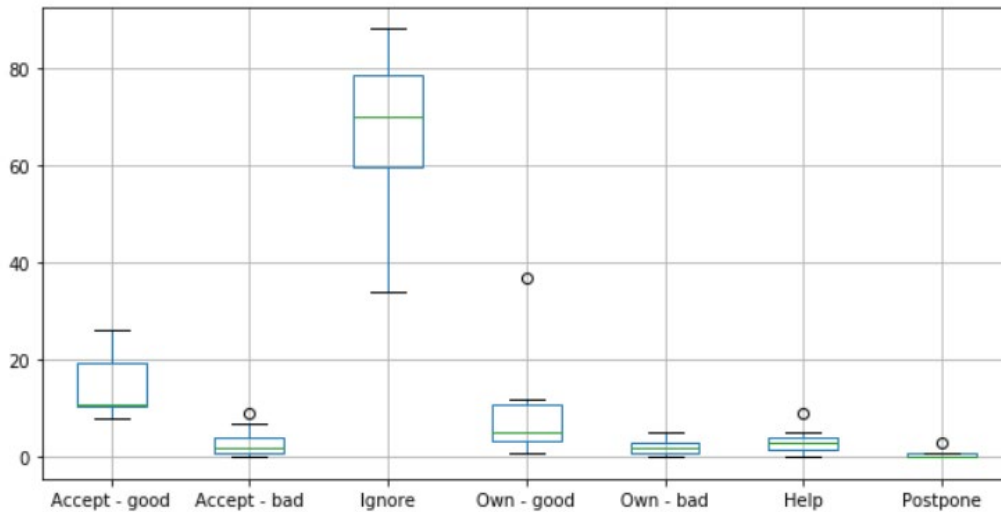


Figure 21 - Participant uptake of feedback by quality

The only exception is when students initiated their own corrections, George made significantly more “good” self-initiated corrections than the rest of the participants did. On the other side of the spectrum, Melissa turned out to be an outlier in accepting “bad” feedback and engaging in help-

seeking activities. There is also a lot more variance in feedback “ignored”, but George was again an outlier, since he ignored a lot less feedback than other participants. Individual student uptake can be seen in Table 19. Here, it is interesting to note that George was the participant who had the highest initial IELTS score (8, against an average of 7).

**Table 19 - Feedback uptake per quality per student**

	Melissa	Alice	Jamie	Rachel	Daniella	Nancy	Violet	Susan	George	Cristina	Iain
Action											
Accept - good	17%	25%	11%	11%	13%	11%	26%	8%	22%	10%	9%
Accept - bad	9%	2%	1%	0%	1%	3%	5%	0%	1%	7%	2%
Ignore	60%	56%	81%	77%	70%	67%	59%	88%	34%	72%	80%
Own - good	1%	12%	5%	5%	11%	11%	4%	1%	37%	5%	3%
Own - bad	3%	1%	1%	2%	2%	4%	3%	0%	5%	2%	1%
Help	9%	4%	1%	4%	0%	3%	2%	2%	0%	3%	5%
Postpone	1%	0%	0%	1%	3%	0%	0%	0%	1%	0%	0%

Looking into more detail at student responses to “bad” feedback, it is interesting to note there were differences in uptake between all three different kinds of feedback (see Table 13 on section 3.4.2 for an explanation of these codes). Most of the “bad” feedback received by the students (70.8%) occurred when the program incorrectly identified a mistake and offered a suggestion that would introduce a mistake into the text. When faced with this sort of feedback, students tended to simply ignore it, as they ignored this feedback 84.5% of the time.

Of the rest of the “bad” feedback, 21.7% of it occurred when the program incorrectly flagged an issue but offered no concrete feedback. As with any other type of feedback that offered no concrete suggestions, this had the highest rate of being ignored by students at 89.7%. 5.4% of the times students initiated their own corrections and managed to improve the original quality of the text, and 2.7% of the time they initiated their own corrections and introduced a new mistake to their writing.

Only 7.5% of the “bad” feedback occurred when the AWE program correctly identified a mistake but offered a suggestion that introduced a new mistake in its stead. However, when confronted with that type of feedback, 38% of the students initiated their own corrections and managed to improve the original quality of the text. This is in contrast with the amount of times they ignored this specific type of feedback, which was 35% of the time. This is the only category of feedback where students were more likely to engage in some action other than ignoring the feedback they received.

It can be concluded that the way students approach the feedback they receive tends to yield positive results, as despite the questionable quality of the feedback received, students tended to ignore the “bad” feedback and to accept the “good” feedback.

Another interesting insight is that there were certain types of feedback that students tended to accept more than others. As can be seen in Figure 22, which considers categories of feedback that offer concrete suggestions and had more than 100 incidences in the dataset, students tended to accept more feedback in the “grammar”, “punctuation” and “spelling” categories, while rejecting more feedback from the “readability”, “passive” and “adverb” categories.

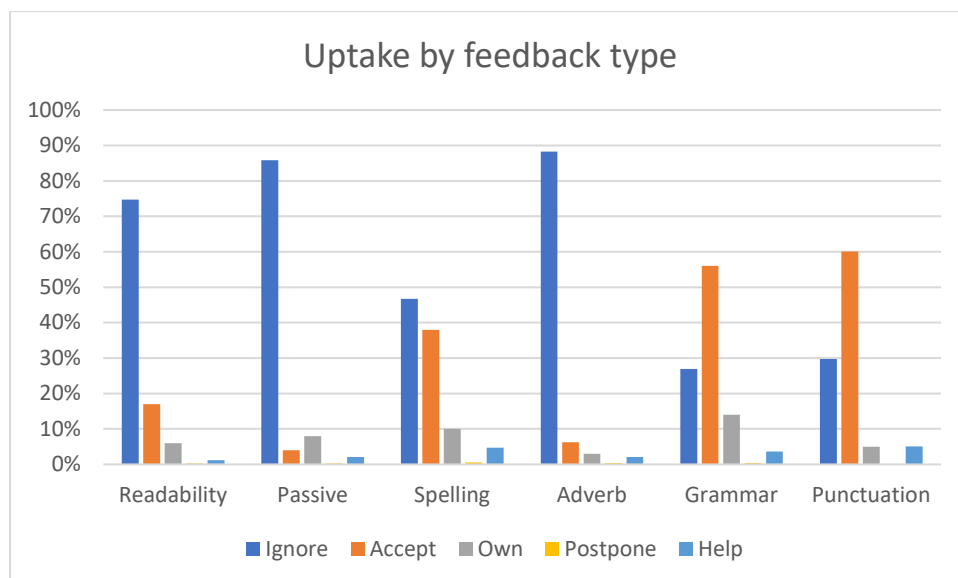


Figure 22 - Uptake by feedback type

This is an interesting observation, as the categories that were most widely accepted by the students were those which purported to identify errors in the texts, while the categories that were most widely rejected tended to make suggestions regarding style. This may suggest that, again, students use a very conscious decision-making process when interacting with the feedback they receive, and that they have a strong sense of their own style which they do not want to change at the program's request.

#### **5.2.4 The use of the AWE program**

Sections 5.2.1 to 5.2.3 gave an overview of how students used AWE programs through three different lenses: that of the self-regulated strategies they used to engage with the feedback they received from the program, that of the decision-making process that led them to judge the feedback and choose whether to accept or rejected, and that of the quality of the revisions students made to their texts as a response to the feedback they received.

So far, the results of this study show that students engage cognitively and behaviourally with the AWE programs, and that they draw upon sources of knowledge such as genre knowledge, previous lessons on English writing, and others, to judge the AWE feedback they receive. This decision-making process usually leads to students using more 'good' feedback they receive from the AWE than 'bad' feedback.

Although so far I have looked at self-regulation, decision-making and quality of uptake in general terms, this section will look more closely into the revision behaviours of two students: George and Melissa. These two students were chosen because the results presented in section 5.2.3 showed them to be outliers for opposite reasons. George had a higher percentage of "good" feedback uptake than most of the other participants, while Melissa had a higher percentage of "bad" feedback uptake. Therefore, it seems interesting to probe deeper into the characteristics of both these students and how they approached their usage of AWE programs, given that they had such contrasting results in their use of ProWritingAid.

**5.2.3.1 George.** George was a male student whose first language was Spanish. As was mentioned in the previous section, George reported the highest IELTS score of the sample, a total of 8 points. An IELTS score of 8 means that he is a “very good user” of English, and he has “a full operational command of the language with only occasional unsystematic inaccuracies and inappropriate usage”, and that he handles complex argumentation well (British Council, 2022). Therefore, George can be said to be a proficient user of the English language.

The way in which George approached the use of the AWE program was very methodical. In the first session, he reflected on previous feedback he had received on the quality of his writing. He identified “words that were repeated, expressions that I feel I am not really sure what I was trying to say” as key weaknesses that he wanted to address using the AWE program. He therefore had a clear and concise plan about how his revision session would go and what he wanted to accomplish by using ProWritingAid. In planning out his session this way, he displayed a clear use of metacognitive strategies to organise his ideas around revision and set his expectations toward the use of the AWE program.

During this first session, George used the AWE feedback as a starting point to reflect on the quality of his writing in general and identify specific weaknesses that he wished to address. The feedback even helped him find names for issues he knew he had in his writing but didn’t know how to articulate: when presented with the concept of hidden verbs, George found this feedback very helpful, because he identified his writing as being too wordy. “That thing about the hidden verbs, I think I didn’t... I didn’t know about it as such. And it makes sense now, yes, because like I said, it helps me make the phrases shorter when they’re unnecessarily long. And because, of course, English is also very direct to say things, and sometimes we [Spanish speakers] are very long-winded”.

Once he had revised the text using the functions he knew would help him achieve his goals, he accessed the summary report to gain an overview of the feedback given to him by ProWritingAid and find other functions that he believed might help him address the perceived weaknesses in his

writing. For example, he was especially interested in the function showed sentence length as a visualisation, stating that “I do tend to make my sentences too long”. George found this function especially useful:

“I find this function very... very useful. My sentences tend to be very long and I never notice. I always tend to... every time I develop an idea on a sentence I sometimes put in many words, like commas or... relative clauses, and then the sentence becomes huge [...] and now I see that I can start separating the sentences and the idea is still as developed [...] you can see it as organisation by line. What the average is and everything, so I can see where to separate”.

Figure 23 shows the visualisation that was presented to George by ProWritingAid. The sentences that were deemed ‘too long’ by the program were highlighted in the text so George could easily find them, but I am not including a screenshot of his text to preserve his privacy and anonymity.

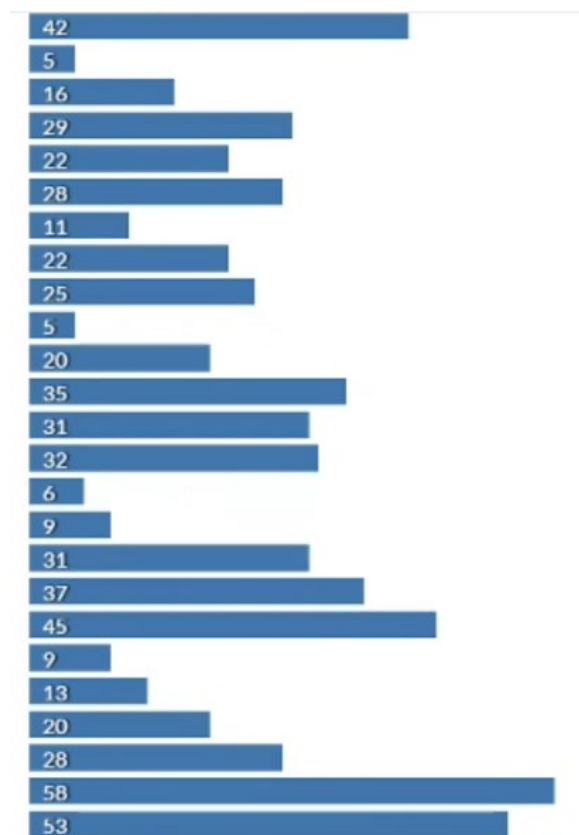


Figure 23 - Sentence length visualisation



The second session was very similar to the first. George began his session with a clear idea of what aspects of his text he wanted to revise and what kind of feedback he wanted to elicit from

ProWritingAid: "Today I'm going to look over the methodology [...] I feel like I wrote it in a hurry.

Let's see... Style, grammar... I think style [...] I want to see how consistent it is in relation to my other chapter".

Having identified his overuse of hidden verbs, his wordiness and his issue with making his sentences too long, George specifically sought out the feedback ProWritingAid provided on those three issues, working methodically through the feedback to revise his text and ensure consistency between the different chapters. As he did so, he reflected on the causes of these issues, and in doing so revealed a deep knowledge of grammar that he applied to make sense of the feedback and use it to revise his writing, as well as using previous feedback given by his supervisors to guide his revision:

"I use a lot of 'in order to' and it's always correcting me, telling me to get rid of it. I think it's going to make me delete more here because there's a lot of words, and my teacher did recommend I lower the count."

"Hmm... 'in a period of one month' ... 'in one month'... yeah, it was too redundant, and unnecessarily long. Yeah, I don't know why I tend to make my writing so long."

"I might have to cut this because otherwise I'd have two adjective clauses and it's a bit disruptive, I think. [...] there was too much information before I said what I actually wanted to say."

"Better this way. I don't know if it's more direct, but it's more simple. [My supervisors] told me to make this chapter really direct with everything I've done."

The third and fourth weeks were slightly different from the previous two. Like the first and second sessions, George began his revision with a clear plan of what he wanted to do and which

ProWritingAid functions he wanted to use to achieve his goals. As he became comfortable with the program and established a revision process, he reflected less on the quality of his writing, and most of his utterances were related to either planning, reading, or rehearsing.

For the third session, he also chose a shorter text so he could play around with other functions of the text he had not yet used and see if he could identify another one that was useful. Unlike his use of the program up until that point, which had been quite structured and purposeful, this was a more exploratory exercise that did not have a great impact on his revision. This exploration only accounted for 10 out of a total of 116 revision actions he performed during that week.

The fourth session also began with a clear plan for his revision:

“Since this is the last session, I want to revise the conclusion. I still have to revise the other thing properly, but I want to look at the last section, because this is the last session and the other one I’ve already revised here and I more or less know how to fix it.”

This, along with the clarity of purpose he demonstrated at the start of each session, seems to indicate that George reflected on the feedback he received and learned from it so he can use it to further revise his text, even without the help of the AWE program.

Related to this finding, one theme that emerged from a closer analysis of George’s work was noticing. When faced with some AWE feedback, George used that feedback to reflect on his writing, and this reflection helped him notice recurring issues with his writing. He expressed a lot of utterances along the lines of “oh, I always do that”, or “yeah, that’s a problem I have”, and learned from the feedback in such a way that it informed subsequent revision sessions and allowed him to have a clear idea of what work needed to be done to improve the quality of his texts.

Another emerging theme was metalinguistic knowledge. In his reflections about the feedback he received and about the quality of his writing, George displayed knowledge of grammatical terms and applied it to his revision and to understanding the feedback. As illustrated in the quotes presented in

this session, George was familiar with terms such as “adjective clauses” and “relative clauses”, and through the feedback he received from the AWE program he learned about other grammatical terms such as “hidden verbs” that he used to guide his revision sessions. This metalinguistic knowledge helped him both make sense of his feedback and clearly articulate the issues he found in his writing, thus giving him a valuable revision tool.

**5.2.3.2 Melissa.** Melissa was a female student whose first language was Chinese. She reported an IELTS score of 7.5, which is classified as a ‘good user’ and means that a student has “an operational command of the language, though with occasional inaccuracies, inappropriate usage and misunderstandings in some situations”, and that in general they “handle complex language well and understand detailed reasoning” (British Council, 2022). Melissa can therefore also be considered a proficient user of the English language based on her IELTS score, but while using the program, she described herself as having a “poor level of English writing” and identified this as a challenge to the revision process.

Like George, Melissa began her first revision session by planning what she was going to do. Unlike George, however, she did not have a clear idea of where she wanted to direct her revision or what the specific issues might be with her writing, so she started by accessing the “Summary” section to get an overview of all the feedback provided by the program to try to get a sense of where to start revising her text. When she did not understand a term, she would use a dictionary to look it up, as she did when faced with terms she did not understand. Once she had taken in an overview of all the feedback, she decided to go through each of the feedback functions in the order in which they were presented.

Melissa brought the same text she had used in the first session to the second session, and she continued where she left off. In the first session she had used the “style” and “grammar” functions,

and in the second session she continued with the third function offered by the program, “overused”<sup>5</sup> and, once she had finished with it, kept on moving through the functions in order, without any particular purpose in mind. Figure 24 shows the order in which ProWritingAid presents its functions, which was the order in which Melissa used them to revise her text.



Figure 24 - Order of functions in ProWritingAid

Melissa brought a different chapter for the third session, and once more she began her revision session by looking at the feedback summary to decide where to start. It was in this session that she began identifying recurring feedback that the AWE program was giving her, and therefore identifying some weaknesses in her writing such as it being hard to read, overusing long sentences and lack of vocabulary to avoid constant repetition. She also identified that she had an issue with missing determiners and began accepting all feedback the AWE program gave her in that category.

However, identifying these weaknesses was not enough. Melissa kept going back to ProWritingAid identifying her text as ‘hard to read’ but she struggled to apply the feedback, stating “I think maybe for this problem I need to find... I need to improve my vocabulary to find other adverbs or other words to replace them. So for now I can’t... I cannot do something about it”, and later stated “I think grammar is kind of basic and um... basic skills of English learning, so sometimes... it’s not... it’s not that easy for me to just improve it”, suggesting that she lacked confidence in her English knowledge and was not sure how to be improving her skills.

This was reflected in the lack of metalinguistic knowledge she displayed. Throughout all four sessions, Melissa consulted her dictionary many times, accounting for the fact that she was an

---

<sup>5</sup> For an overview of each of these function, see section 3.1.3

outlier in the “help” category. For example, she used her dictionary to look up the meaning of terms such as “inconsistent” and “redundant”, as well as looking up the meaning for some concepts other participants struggled with, such as “hidden verbs”.

Despite this lack of confidence, she seemed to have a clear idea of what ‘academic writing’ should look like, constantly disagreeing with the program’s suggestions to reduce the use of passive voice in the text, stating that “passive verbs sound more scientific... or objective”, therefore ignoring most of the ‘style’ feedback the program gave her.

On week 4, she tackled her revision session the same way as weeks 1 and 3. She began her session by opening the feedback summary and looking at the scores the program gave her to decide which aspect to address first. As happened with the previous sessions, the lowest score was related to the ‘style’ of her writing, mostly because the program considered that she overused the passive tense and adverbs.

However, as happened with the previous three sessions, once she looked at the specific feedback she received from the AWE program she realised some of it was too vague and she did not know how to begin fixing the issues it pointed out, or she did not think were big enough problems to warrant revision. Since she did not have a clear idea of what she needed to revise or what she wanted to get out of the program, she spent a lot of time looking at feedback that in previous sessions she had decided was not useful to her, before realising that this was still the case.

**5.2.3.3 Comparing George and Melissa.** Some interesting insights might be drawn from this in-depth look at the way in which George and Melissa used the AWE program. First of all, both of them used similar self-regulation strategies when engaging with the feedback, beginning their sessions by forming a plan for revision. However, George came to all sessions with a good idea of what specific aspects of his text he wanted to revise, and therefore quickly established a plan of action and sought out the particular functions of ProWritingAid that would help him achieve his goals. Melissa, on the other hand, did not have a clear idea of which aspects of her text she should focus on, and therefore

spent the first few minutes of the revision session looking at the feedback summary before deciding where to start her revision.

Another thing that stands out in the way George and Melissa approached revising their texts was how they judged the feedback they received and the decision-making process they underwent to decide whether they would accept or reject the feedback. Both of them drew heavily on their knowledge of what 'academic writing' should look like and feedback they had previously received from their supervisors, but George also had a wider metalinguistic knowledge that he could apply to understand the feedback he received and revise his text as a response to it. Melissa, on the other hand, had more limited metalinguistic knowledge and struggled with understanding the terms used by ProWritingAid when providing the feedback, and therefore was unsure of the quality of the suggestions and felt like she did not have the ability to carry out the required revisions.

### **5.3 Discussion**

Engagement is a complicated construct to measure because there are many conflicting theoretical approaches to understanding the concept (Fredricks et al., 2004). As a construct, however, engagement is quite closely linked to self-regulation, as self-regulated learners are those who engage with their studies in various ways, and the use of cognitive and metacognitive strategies has often been considered a sign of engagement (Wolters & Taylor, 2012). Furthermore, in the context of revision rather than learning-based activities, the act of changing or revising a text can be considered an indicator of behavioural engagement (Zhang & Hyland, 2018). The results of this study help shed light into how strategies for self-regulation may be used by students to engage both cognitively and behaviourally with their texts.

Two metacognitive strategies employed by students were goal setting and planning. Many participants, for example, planned their revision sessions by drawing upon previous feedback they had received from the AWE program, their own assessment of their writing strengths and weaknesses, or the analytics presented to them in the 'summary' section. This planning process was

also related to the goals students set for their own texts, for example, increasing grammatical accuracy, reducing the length of one's sentences, removing overused words, etc. Establishing goals related to improving specific parts of writing or specific forms that students struggle with has been found to be an important factor in the successful use of AWE programs for revising (Liao, 2016a; Zhang & Hyland, 2018). However, as the comparison between the planning strategies of George and Melissa in section 5.2.3 showed, the quality of the planning undertaken by the participants was not the same: George had a plan based on what he already knew about his strengths and weaknesses, while Melissa made up a plan on the spot based on the AWE feedback she received. While feedback that promotes self-regulation is often considered the most useful for students (Hattie & Timperley, 2007), my results suggest that students use self-regulatory strategies in spite of the AWE program and not necessarily because its feedback is conducive to developing self-regulation.

The most commonly used strategy in the sample was judging information. This makes sense, given that the task required students to look at the feedback given to them by the AWE program and determine whether to accept or reject said feedback. By examining the student's think-aloud protocols, we could delve deeper into how students made that judgment, as eight separate criteria for judging information were found in our sample (see Figure 18). Other studies have also found that students are selective when accepting AWE feedback (Chodorow et al., 2010), drawing on various sources of knowledge on English grammar, vocabulary and mechanics in order to judge the reliability of AWE feedback, which tends to be quite low (Bai & Hu, 2017). General research on WCF for acquiring L2 writing skills also seems to corroborate this, as the effectiveness with which students use feedback to revise their work depends on their level of confidence in prior language instruction and levels of motivation (Ferris, 2019), and my findings support this literature by making explicit the specific mechanisms at work in the use of AWE feedback in particular.

However, my results revealed that, alongside this previous knowledge of English writing, students also looked at their own text to determine whether the feedback they received was consistent with

their own personal style and the message they were trying to convey. This might explain why there was a greater rate of uptake in feedback that corrected mechanical errors such as grammar or punctuation than feedback related to the style, and as such was neither “good” nor “bad” according to the criteria used in this study.

Furthermore, students are not necessarily constrained by either accepting or rejecting the feedback offered by the AWE program, as said feedback can also elicit reformulations or deletion of sections of text (Ranalli et al., 2017). Own corrections in our sample were fewer than outright accepting the existing feedback, but they still comprised 41% of revisions undertaken by the participants, often in response to bad feedback given by the AWE program. Most of these revisions were made in response to the program correctly identifying a mistake but giving bad feedback and, as such, were mechanical in nature.

A common worry about the use of AWE has been that students will only make surface-level revisions to their texts (CCCC, 2006; C.-F. Chen & Cheng, 2008), and while my findings suggest that even those mechanical revisions are made taking into account the intended meaning of the text, it is also true that the vast majority of corrections made by the participants were superficial in that they did not necessarily change the content or structure of the text.

This might be because, when it came to suggesting broader corrections, the feedback tended to be vague. One of the program’s functions pointed out sentences that might be considered too long, and while one participant made extensive use of that feature to revise significant portions of their texts, most other participants who used that function expressed that they weren’t sure what to do with it. Another function pointed out paragraphs which were deemed hard to read according to the Flesch ease of reading score, but even though some participants used that tool, it elicited no revisions, as the AWE program offered no guidance on what made those paragraphs hard to read. This is consistent with general literature on the use of feedback, since the most effective type of feedback is that which promotes a dialogue and a negotiation of meaning (Boud & Molloy, 2013; Carless &



Winstone, 2020), something AWE programs are incapable of doing. In most cases, the ability of students to make sense of the feedback was influenced by how explicit it was, which has been shown to be a determining factor in student uptake of feedback (Heift, 2004; Ranalli et al., 2017).

Despite these issues, students tended to have positive perceptions toward the use of ProWritingAid, even if they were aware of its limitations. Violet, who was studying a postgraduate Teaching English as a Second Language (TESOL) course, noted that she might get her future students to use an AWE program to revise their first drafts before submitting to them for teacher feedback, something that has been noted to help students revising their writing (C.-F. Chen & Cheng, 2008). Overall, students were satisfied with ProWritingAid as a way to revise spelling and grammar errors in their writing, but expressed frustration when it came to use the program for more substantial revisions, which is consistent both with the uptake rate observed in our sample and the findings of previous research (e.g. Chen & Cheng, 2008; Jiang & Yu, 2020; E. L. Wang et al., 2020).

In fact, in order for students to make good use of technological tools, they need to recognise that the tool will be useful for their task and be able to use it skilfully, as well as have the motivation to actually use said tool (Winne, 2006). While some of the students acknowledged the usefulness of knowing whether their paragraphs were difficult to read or their sentences too long, they lacked the skills to make any meaningful use of that information. The only exception to this was George, who had previously been told his sentences were too long, and therefore specifically sought out the 'sentence length' feature of the AWE program and made extensive revisions to his paragraphs based on that feedback. Other students also sought out these features based on previous feedback but ultimately did not carry out any revisions because they lacked the skill to do so. This has important implications for pedagogy because instructional conditions are important for learner uptake of technological tools (Gašević et al., 2017), and teacher scaffolding has been specifically observed as crucial in use of AWE programs (Jiang & Yu, 2020; J. Li et al., 2015; O'Neill & Russell, 2019).

Therefore, it can be said that participants engaged cognitively with the feedback offered by the AWE program, by employing metacognitive and cognitive strategies to make sense of the feedback they received (Zhang & Hyland, 2018). While most of this cognitive engagement did not lead to behavioural engagement (i.e. making changes to the text), the cognitive engagement led to a thorough decision-making process that enabled participants to make good decisions on which feedback to accept and which to reject. The end result was that most of the feedback accepted by the students tended to be “good” according to the criteria used in this study.

## **5.4 Conclusion**

The question of whether AWE programs can be useful tools in offering formative feedback to ESL/EFL students is a complex one. While previous studies focused more on whether these programs resulted in a measurable gain in writing scores or an overall reduction of errors (Hibert, 2019b), there is mounting evidence that the way students are prompted to engage with their own texts as a response to AWE feedback also seems important (Zhang, 2017, 2020; Zhang & Hyland, 2018).

The study in this chapter investigated how students engage with an AWE program by examining thoughts and attitudes elicited from participants through a think-aloud protocol, and the actions performed by participants as a response to the AWE program taken from the screen recordings. The study reported in this paper yielded evidence that students engage with AWE feedback both cognitively and behaviourally, using metacognitive strategies to mediate between their engagement and their decision-making. It also adds to mounting evidence that students do not uncritically accept feedback given to them by AWE programs, but draw on their previous knowledge of English writing (Bai & Hu, 2017; Chodorow et al., 2010; Zhang, 2017). The current study also yielded evidence that students also consider their intentions and the overall meaning of their text when deciding whether to accept AWE feedback, even for mechanical errors.

An important limitation in this study is the lack of focus on affective engagement beyond student perceptions, which has been found to correlate to quality of student interaction with AWE programs

(Zhang & Hyland, 2018). Because of the controlled conditions in which students interacted with the AWE program, the focus was on behavioural and cognitive engagement. While some insights of affective engagement related to student perceptions were gleaned from the semi-structured interviews, findings were limited, and future research must take into account the role played by this type of engagement in AWE usage.

Another limitation is that think-aloud protocols might provide incomplete information “when learners do not or cannot verbalise ongoing thoughts” (Veenman, 2011). This was an issue in this study, as several participants expressed difficulty in verbalising their thoughts, and even with neutral prompting from the researcher, some were unable to produce verbalisations consistently, explaining that not being native speakers made it harder for them to verbalise thoughts in English. Although the effects of these issues were somewhat offset by using screen recordings to capture the actions made by the students, future research should address this issue.

In the next chapter, I will bring together the results and discussions of the studies in both Chapter 4 and Chapter 5 to provide an in-depth discussion of how students use and engage with AWE programs as tools for formative feedback.

## Chapter 6 – Discussion

The previous two chapters detailed the results of two studies carried out as part of this thesis. As mentioned in the methodology chapter, these two studies were designed to research the usage of an AWE program, ProWritingAid. The study reported in Chapter 4 was designed to answer the first research question, investigating what changes students made in response to AWE feedback. That study examined the product of student revisions with the help of an AWE program, focusing on the texts themselves and what an analysis of their features could tell us about the effects of using AWE programs for revision. The study reported in Chapter 5 was designed to ask the second research question by looking at how students engaged with AWE programs. It focused on the process of revising with an AWE program, using qualitative methods to understand the strategies and decision-making processes students used to engage with AWE programs. This chapter brings together the findings of both studies to answer the research questions postulated in Chapter 3 and discuss the general picture of how this AWE program was used as a tool for revision in both of the studies, and the implications on the broader usage of AWE programs as tools for formative feedback in the ESL/EFL classroom.

### **6.1 The process and product of AWE interaction**

In the literature review, I pointed out how academic writing skills are taught and evaluated through product and process approaches. Product approaches to evaluating writing are the more traditional, and the focus of these approaches is the production of an error-free, coherent text. However, most modern literature on acquiring writing skills has moved beyond simply looking at the product of writing itself, and instead considers writing as a cyclical process which focuses on planning, drafting, editing and revising as essential parts of the writing process (Neupane, 2017). In this study, I decided to use both product and process approaches to gain a more comprehensive picture of how students used AWE feedback to revise her texts. Given that feedback is a process which “has a role in

developing students' continued evaluative judgement" (Molloy & Boud, 2013, p. 17), both dimensions are important in understanding the impact of the use of AWE programs as tools for formative feedback, especially in ESL and EFL classrooms where adherence to grammatical conventions is an important factor in academic success (O'Neill & Russell, 2019). Therefore, one of the main contributions of this thesis is delving into both the product and process approaches and bringing them together to form a holistic understanding of the impact of AWE feedback on students' writing abilities.

### **6.1.1 The product of revision with AWE feedback**

As mentioned in the literature review, a substantial amount of research into the use of AWE programs for formative purposes has focused on studying how AWE feedback affects the final texts produced by students. For example, some research has looked into whether using an AWE program would lead to a reduction in the number of errors present in a text (e.g. Chodorow et al., 2010; Lavolette et al., 2015; Li et al., 2015; Long, 2013; Ranalli et al., 2017), or whether using AWE programs has led to an increase in writing scores, either provided by the AWE itself or by human raters (e.g. El Ebyary & Windeatt, 2010; Huang & Renandya, 2018; Saadi & Saadat, 2015; Tang & Rich, 2017; Wang et al., 2013).

Results of this research is quite diverse, from studies finding that the use of AWE feedback greatly reduced error rates or increased accuracy (Chodorow et al., 2010; J. Li et al., 2015; Z. Li et al., 2017), to studies finding that the improvements were only marginal (Liao, 2016a; Long, 2013; Ranalli, 2018), to studies finding no significant improvements in scores or reduced error rates (Dikli, 2010; Huang & Renandya, 2020; Lavolette et al., 2015). On the other hand, there is evidence that AWE can help with reduction of error rates and increase in accuracy on the long, rather than the short term (Liao, 2016a; Link et al., 2020).

Given this diversity of findings, it is hard to draw any hard conclusions on whether the use of AWE feedback can help ESL/EFL students reduce the number of errors in their writing. However, one

common issue that can be gleaned from the literature is that most of the revisions that students make are superficial and only attend to mechanical errors (Dikli, 2010; Koltovskaia, 2020; Ranalli, 2018), as well as ignoring most of the feedback they receive (Koltovskaia, 2020; Tian & Zhou, 2020). The study in Chapter 4 contributes to this understanding by using research questions 1.2 and 1.3 to understand how substantial their revisions were, and whether the changes made by the students improved the quality of texts as measured through linguistic factors that have been found to correspond to human judgements of writing quality. The study in Chapter 5 contributes to the literature by answering research question 2.3 of whether their revisions help them make good decisions in their feedback uptake by looking at the amount of “good” and “bad” feedback they accept.

**6.1.1.1 Feedback uptake.** The first notable finding from both the Chapter 4 and Chapter 5 studies was that most of the feedback was disregarded by the participants. Many other studies have also found that a large proportion of AWE feedback gets ignored (Bai & Hu, 2017; Chodorow et al., 2010; Koltovskaia, 2020; Lai, 2010; R. Li et al., 2019; Link et al., 2020), and the reason for this has been thought to be the quality of the feedback itself, which is quite generic (Dikli, 2010), and students are not always sure how to apply it to their own writing (Ranalli, 2018).

The studies detailed in Chapters 4 and 5 found that 67% and 68% of the feedback, respectively, was ignored by the participants. This is more than the proportion of feedback ignored by Link et al.’s (2020) participants, who ignored 24% of AWE feedback, but closer to Tian and Zhou’s (2020) participants, who ignored around 75% of the feedback they received. This may be due to the use of different AWE programs using different scoring and feedback engines. Link et al. (2020) used Criterion, while Tian and Zhou (2020) used Pigai, and I used ProWritingAid. However, there seems to be some further nuance to this rejection. The literature, and my own findings from Chapter 5, show that different kinds of feedback are taken up at different rates. Bai and Hu (2017), for example, found that students tended to accept most grammar and mechanics feedback, while ignoring most

of the feedback related to collocations and synonyms. Li et al. (2019) also found that surface-level AWE feedback tended to be accepted with much more frequency than meaning-level AWE feedback, something that did not hold true for either peer or teacher feedback. My own results are in line with this research, as student uptake was high for spelling, grammar, and punctuation, but low on non-mechanic feedback such as readability improvements, eliminating passive voice and reducing the use of adverbs in the text. The study reported in Chapter 4 offers similar results, with participants being more likely to accept feedback related to spacing, spelling, and use of determiners, while less likely to accept feedback related to redundancy, use of passive verbs, use of prepositions, and suggestions for stronger wording. This discrepancy in feedback uptake by type might be explained by the perceived complexity of it, as in many cases the students reported in Chapter 5 were not sure what the feedback meant, but the AWE programs offer little in the way of dialogue and negotiation of meaning, which have been found to be important in helping students make sense of the feedback and successfully apply it to their learning (Boud & Molloy, 2013; Carless & Winstone, 2020). Further supporting this insight from broader feedback literature was the fact that some of the students who struggled the most with the feedback, like Susan, preferred to ask the researcher for help rather than use the tools offered by the AWE program, since she and the researcher could engage in dialogue and clarify the feedback to ensure comprehension.

Chapter 4 offers a further insight into the way students use the AWE tool and their use of the feedback. Out of the 70 texts that fulfilled the criteria of being longer than 100 words, only in 30 of them had the participants made any changes. However, by examining the texts where participants made no changes, an interesting pattern emerged. Many of the participants accessed their texts several times, at different times, and most of those occasions they spent time looking at their texts and used the program's summary function, but ultimately took no actions to revise the texts, either using the AWE feedback or making revisions of their own. In a study of 33,171 texts submitted by 6<sup>th</sup> to 12<sup>th</sup> grade students to Criterion, Attali (2004) found that 71% of those texts were only submitted once, indicating that students had not used the program to make any revisions. The study in Chapter

4, however, went one step further than Attali, since we could see that many students returned to the text and used the summary function more than once without necessarily resubmitting. This failure to use any AWE feedback at all has been underexplored in current literature, perhaps because most studies were made as part of an instructional plan, and therefore made revision of texts mandatory (Bai & Hu, 2017; El Ebyary & Windeatt, 2010; Liao, 2016a) or encouraged student revision by requiring them to get a certain score from the AWE program before submitting for teacher feedback (C.-F. Chen & Cheng, 2008; Jiang et al., 2020; J. Li et al., 2015). The study detailed in Chapter 4, on the other hand, gave students complete liberty to use the AWE program as they saw fit, as it was not tied to a specific class but rather an elective.

**6.1.1.2 Text changes.** Feedback uptake, however, does not paint a full picture of the way in which students use AWE programs to revise their texts. As explained in Chapter 4, linguistic markers of complexity were chosen for this study because they can help understand the text in a more holistic manner, given that research has shown linguistic features in a text can be reliable predictors of L2 writing proficiency (Ryu, 2020). While the study detailed in Chapter 4 looked at feedback uptake, the approach was novel in that it also sought to understand how texts changed as a response to AWE feedback, not only in terms of error reduction, but in terms of the linguistic composition of the texts, as error rates or scores by themselves are a limited measure of whether an AWE intervention has been successful (Hibert, 2019b).

Both the study reported in Chapter 4 and the study reported in Chapter 5 show that the text changes carried out by the participants were mostly superficial and not substantial enough to affect the linguistic composition of texts. In fact, the findings of Chapter 4 showed that the changes to linguistic composition were not uniform, with some texts increasing their scores and others decreasing the scores. There was no evidence that any of the characteristics measured in the study were significantly impacted by the AWE feedback.



The results from this study, therefore, contribute a new dimension to the idea that AWE feedback is often underutilised, and gives strength to the argument that students often do not find the generic and redundant feedback useful (C.-F. Chen & Cheng, 2008; Dikli, 2010). The study also adds nuance to Wang et al.'s (2013) findings that the longer students used the Criterion program, the less useful they found it. This corresponds to the findings of the study detailed in Chapter 5, where the participants expressed their frustration with the lack of specificity in the feedback they received from the AWE program. For example, Alice expressed frustration when ProWritingAid told her that her text was too difficult to read but offered her no guidance as to why. Rachel expressed a similar frustration: "And now it gives me a suggestion that I cannot understand. I don't know what is wrong with my writing, I can't understand whether it's suggesting me to change the... how to say? the markers, or to change the words". This led to them not knowing how to implement the feedback into their own writing and ultimately disregarding most of it.

Since dialogic conditions are essential for students to understand feedback and apply it to their own learning and revision (Boud & Molloy, 2013; Carless & Winstone, 2020), and the AWE program does not allow for students to engage in dialogue and clarification about the feedback they receive from it, it makes sense that the students in our sample struggled to use it. The fact that AWE provides unfocused feedback might also contribute to low uptake, given that research into the use of feedback for promoting L2 writing skills has shown that focused feedback might be better to help students notice forms and transfer the feedback to new contexts (Bonilla López et al., 2018; Ellis et al., 2008; Kang & Han, 2015)

### **6.1.2 The process of revision with AWE feedback**

However, while our findings about student uptake of feedback are consistent with both the literature on the use of AWE programs and broader literature on feedback and second language acquisition, looking at the product of revision with AWE feedback alone is not going to help explain in detail how and why students accept or reject the feedback they are given. To better understand

why feedback uptake is so low among participants, and the dynamics of student interaction with AWE programs, it is important to look beyond the product of the revision and examine the process of revision itself. The writing process encompasses planning, drafting, revising and editing (Neupane, 2017), and those two last processes were the focus of the first research question posited in Chapter 4 about student revision behaviours and the study detailed in Chapter 5, whose first research question sought to understand how students engaged with the AWE program and whose second question investigated how students make decisions about which feedback to use and how to revise their texts.

**6.1.2.1 Use of 'summary' function.** While the study reported in Chapter 4 did not delve qualitatively into student engagement with the platform, it found that students who made changes to their texts often used the 'summary' function, which gives an overview of feedback going beyond grammar and style and offering information on sentence length, sentence variety, repeated phrases, etc. The results from Chapter 5 delved a bit deeper into this behaviour, finding that students used the 'summary' function to plan out their revision sessions. The 'summary' function, however, was not the only source that students drew upon when planning how to revise their texts; they also drew upon previous feedback they had received from both the AWE program and their supervisors.

While research into student engagement with AWE programs has been scarce, some findings suggest that the students who are more successful at using these programs to revise engage in planning and reflexion of their feedback (Zhang & Hyland, 2018). This is consistent with broader literature on the use of feedback for learning and revision, as having a clear direction by answering the three clear questions of where the student is going, how they are going, and where to next are essential in student use of feedback (Hattie & Gan, 2015; Hattie & Timperley, 2007). This was observed in Chapter 5, as George took his time to plan out all his revision sessions and reflected on the feedback he received from the program in between sessions to inform this planning process. While no causality can be established based on the data available, George used more 'good' than

'bad feedback in his revisions, which suggests that these strategies might help with success because his planning of revision sessions can be seen to be consistent with Hattie and Timperley's (2007) three questions that guide successful uptake of feedback. A less successful student in Zhang and Hyland's (2018) study did not use the available tools for planning and as a result felt overwhelmed by the amount of feedback the AWE program gave him, which led him to rely heavily on teacher guidance, both characteristics that Susan displayed during her interaction with the AWE program. This suggests that planning is an essential part of successfully using AWE feedback, and merits further investigation.

#### **6.1.2.2 Proficiency level.**

The ability of students to make sense of feedback and engage with it seems to vary depending on their language proficiency (Koltovskaia, 2020; Tian & Zhou, 2020; Zhang & Hyland, 2018). Although in both studies participants seemed to have a similar proficiency level, given that they all had IELTS scores between 6.5 and 8, their actual ability to use and understand AWE feedback did vary in several ways. The participants of the study related in Chapter 5 expressed frustration at not being able to understand the feedback proffered to them by the AWE program, especially in categories such as sentence length, ease of reading, use of passive voice, and other issues that would require rewriting of sentences or paragraphs to resolve, which students revising with AWE programs rarely do (El Ebyary & Windeatt, 2010; Jiang & Yu, 2020). In fact, 19% of all examined utterances expressed some form of uncertainty about the feedback received. Given the amount of 'bad', 'mixed' or 'neutral' feedback given by the program, it is not surprising students struggled to make sense of it (Bai & Hu, 2017; Dikli, 2010; O'Neill & Russell, 2019). As has been discussed, an explanation for this can be found in general literature on feedback uptake, which has found that students are less likely to make sense of feedback and therefore close the feedback loop if they cannot engage in dialogue with the feedback provider (Boud & Molloy, 2013; Hattie & Timperley, 2007; Molloy & Boud, 2013; Nicol & MacFarlane-Dick, 2006), which AWE programs don't allow as of yet.

**6.1.2.3 Engagement with feedback.** Chapter 5 also looked at other factors that affect students' decision-making processes when faced with AWE feedback. It found that students engage with the feedback by drawing from different sources to determine its accuracy. Those sources included things they had learned in previous English classes, comments made by teachers or supervisors, their own observations from reading academic literature, and external tools such as google searches. Students' previous learning and comments made by teachers have been found to help them make sense of feedback in L2 writing settings (Ferris, 2019), and my results from Chapter 5 add further nuance to this understanding by including students' observations of academic writing conventions from reading literature and the use of external tools to help them understand the feedback.

Research into the different sources of knowledge that students draw upon is an area that has been neglected in current AWE literature. Koltovskaia (2020) and Zhang & Hyland (2018) found that, out of the two participants in their studies, only the high proficiency participant made use of external tools to make sense of the feedback they received, and in the case of Koltovskaia's participant it consisted of a single Internet search. Both Koltovskaia (2020) and Zhang & Hyland (2018), however, focused only on using technological tools, and while our findings support the idea that students do use these resources to revise, there are other sources of previous knowledge that students use to contextualise and make sense of the feedback they receive, such as knowledge of the genre, previous teacher feedback, and grammar knowledge. This is an important point to consider, because cognitive engagement with feedback can be mediated by prior linguistic knowledge like familiarity with grammar rules and collocations, among others (Tian & Zhou, 2020), and my results support the idea that this prior linguistic knowledge is essential in making sense of the feedback students receive from AWE programs.

The results from Chapter 5 suggested that students engage both behaviourally and cognitively with AWE feedback, and that metacognitive strategies were used to make sense of said feedback, which

has been a recurring theme of AWE literature that deals specifically with engagement (Koltovskaia, 2020; Zhang, 2017, 2020; Zhang & Hyland, 2018). This allows for a revision process that yielded a reduction in error rates for the revised texts, even if the study in Chapter 4 found that this does not necessarily have a measurable impact on linguistic markers of proficiency in their texts because most of the changes students make in response to AWE feedback are superficial (C.-F. Chen & Cheng, 2008; Jiang & Yu, 2020; Roscoe et al., 2017). However, the fact that the findings in the study presented in Chapter 5 corroborate previous research showing that students make thoughtful decisions that result in discarding 'bad' feedback (Bai & Hu, 2017; Chodorow et al., 2010), means that more research is needed to understand how these decision-making processes can be supported to help students make sense of AWE feedback (Jiang & Yu, 2020).

## **6.2 Potential of AWE as a tool for formative feedback**

One important draw for the use of AWE programs in classrooms is the possibilities it offers for providing students with continuous formative assessment, encouraging students to review their work and increasing their motivation for writing (Hockly, 2019). Our findings from Chapter 4 and Chapter 5 can help understand how students use the feedback provided by AWE programs for formative purposes by examining how students use AWE programs for revising and how they engage with the feedback they receive. There are two main areas where it is believed that AWE programs might help for formative assessment that the results from this study can help shed light on. The first is the possibility of submitting multiple drafts, gaining instantaneous feedback on each of them, and therefore being able to immediately address the issues found by the AWE program (Woodworth & Barkaoui, 2020). The second point, related to the first one, claims that the use of AWE might promote noticing of form (Cotos, 2011), which draws attention to language use and so is a key mediator in second language acquisition (Chapelle, 2010).

These two areas are key in understanding not only how AWE programs are used by students when revising their texts but are also important in helping understand which pedagogical strategies might

help introduce AWE programs into classrooms in ways that help in the provision of formative assessment, and therefore it is worth discussing them in detail. The research questions of this thesis focused on how students used AWE programs to revise their texts and, in this section, I will be discussing how my findings on student revision behaviours and student engagement with AWE programs can help further current understanding of how students can use this feedback to create multiple drafts and notice linguistic forms for future revision.

### **6.2.1 Multiple drafting**

Multiple drafting and revising are key components of process approaches to writing, as they allow students to reflect on their texts and seek input from various sources in order to rework the text (Neupane, 2017). There is some evidence that the use of AWE programs helps motivate students to revise and redraft their writing (El Ebyary & Windeatt, 2010; J. Li et al., 2015; P. Wang, 2013), and the results of Chapter 4 seem to slightly support this notion since several of the students went back to revise their texts several times even though, unlike other research, they were not specifically asked to do this. El Ebyary & Windeatt (2010), however, found that, while students did use the AWE feedback to make multiple drafts, during revision, they tended to adopt avoidance strategies such as using simpler sentences or deleting parts of their text to artificially inflate the scores given by the AWE program. In contrast, the findings of Chapter 4 show that, while there were no significant changes or improvements in the texts revised with the AWE program, they were also not significantly simpler, so avoidance strategies were not an important factor in this study.

There is also evidence that the use of AWE programs to help with the drafting process depends on how the program is introduced into the classroom (C.-F. Chen & Cheng, 2008; J. Li et al., 2015; Tang & Rich, 2017), and in general feedback works best when it is given as part of a well-designed instructional environment in which teachers set up the conditions for students to appreciate and use feedback (Carless & Winstone, 2020). The results from Chapter 5 seem to support this notion, given that participants often referred to feedback they had received from teachers and supervisors when

planning their revision sessions and making decisions about which functionalities of the AWE program to use. In that chapter, I examined the specific case of George, and how he used previous feedback given to him by his supervisors to carefully plan out how to revise his texts and seek out only the AWE functions that would help him achieve his goals. While not all the students in the sample used these strategies as effectively, they nonetheless did refer to previous feedback by supervisors to make decisions, or even asked the researcher for help directly if they felt they couldn't understand the feedback on their own, supporting existing literature about the importance of teacher guidance in the use of these programs (J. Li et al., 2015; Tang & Rich, 2017).

Chapter 4 also provided an insight into the usefulness of the AWE tool for drafting purposes: more than half of the texts submitted to the platform were submitted more than once, meaning that students came back to revise their texts more than once before being satisfied with them. Notably, this interest in revisiting text and revising several times came about unprompted, as the participants in the Chapter 4 study were not required to achieve a certain score nor use the AWE program in a certain way, unlike most of the research carried out into this topic.

In fact, user logs analysed in Chapter 4 also showed that sometimes the participants would come back to their texts several days after first submitting them to continue revising, something Zhang & Hyland (2018) found their high proficiency student did to give herself time to reflect on the feedback and the text itself before continuing to revise. However, this does not necessarily mean that students made substantial revisions to their texts. The results in Chapter 4 also show that most of the revisions undertaken by the students were superficial and did not significantly change the texts themselves, with only an average of 19 changes per text, most of them pertaining to fixing grammar and spelling mistakes.

The fact that participants in Chapter 4 were willing to use the AWE program to revise their texts and made multiple drafts but were not able to make any significant changes to it suggests that students have the motivation to revise and the willingness to use the AWE programs but not the skills to

make good use of the AWE feedback. This is in line with existing literature, which often reports the superficial quality of revisions made by students with AWE programs (Stevenson & Phakiti, 2014, 2019). Chapter 5 helped shed light on this phenomenon, as students often expressed uncertainty about how to use the feedback they were offered by the AWE program and frustration about the ambiguity of some of the feedback they received. That students experience frustration and ignore feedback they do not understand is well documented in existing literature on written corrective feedback for second language learning and literature on feedback in general (Hattie & Timperley, 2007; Truscott, 1996; Zimmerman & Kitsantas, 1999), with some research into the use of AWE finding that participants tend to lose their excitement in the use of AWE programs upon encountering vague and repetitive feedback (Lai, 2010; Roscoe et al., 2017; M. J. Wang & Goodman, 2012; Y.-J. Wang et al., 2013).

One of the ways in which this dissonance between student motivation and ability to act on it is resolved through a carefully thought-out pedagogical strategy implemented along with teacher training (Tang & Rich, 2017). I discuss how the results from my research can inform the particulars of such a strategy and its implications in section 6.4, as they encompass more than just the support for multiple drafting.

### **6.2.2 Noticing**

The second area where the use of AWE programs might help as tools for formative feedback is noticing, which was one of the themes that emerged from my analysis on self-regulation and decision-making reported in Chapter 5. Recent second language acquisition theory has emphasised the importance of noticing, defined as a conscious attention to the form of input (Robinson, 1995), since interaction draws attention to language and therefore makes the acquisition of form easier (Chapelle, 2010). This is especially important when it comes to acquiring writing skills and especially when revising texts, as students often have a mental representation of what they mean to say and



sometimes skip over errors because they know what their own meaning is supposed to be (Flower et al., 1986).

It is believed that the constant and immediate provision of feedback afforded by AWE programs can help draw attention to common mistakes and therefore promote noticing of form in students (Barrot, 2021; Koltovskaia, 2020; Lavolette et al., 2015), something that I observed in several cases reported in Chapter 5. This was the case of Rachel, who noticed she had recurring issues with pluralising nouns, and George, who noticed he had an issue with including too many clauses in his sentences, making them too long. On the other hand, some fear that the amount of feedback offered by these programs, some of which is repetitive or inaccurate, might overwhelm students and impede the noticing process (Liao, 2016a), which was observed with students such as Susan, who was overwhelmed with the feedback she received and often had to ask the researcher for help in understanding and prioritising the AWE feedback she received.

The contrasting cases of George and Susan show that the matter of using AWE programs to notice form is not straightforward. Task demand seems to play an important role in this noticing, since cognitively demanding tasks lower the chances of consciously noticing input (Bonilla López et al., 2018). However, while the necessity of evaluating AWE feedback for correctness does seem to impede students' ability to use it (Ranalli, 2018), there is no evidence that this process of evaluation overtaxes students' cognitive capabilities (Ranalli et al., 2017).

The concept of noticing is an under researched area in the domain of AWE studies, and as it showed up as an emerging theme in my research, I did not have the data to delve deeper into how and what students noticed when using the AWE feedback. My results, however, show that students did in fact begin to notice certain linguistic forms that recurred in their feedback, and they used this knowledge to plan further revision sessions. These results are supported by the existing literature, given that the impact of noticing form seems to be mediated by metacognition (Zhang, 2020) and the degree of autonomy experienced by the student (Liao, 2016a). Since AWE programs are mostly used on the

students' own time, autonomy, and the ability to self-regulate are important factors in their successful use of AWE programs to revise texts and improve writing accuracy (Hibert, 2019b; Zhai & Ma, 2021). Therefore, if we are to delve into how students can use AWE programs to promote both multiple drafting and noticing of form, it is important to talk about self-regulation and student autonomy in this context. In the next section, I will be discussing how the results of this thesis further an understanding of the importance of these concepts in the use of AWE programs.

### **6.3 Self-regulation in the use of AWE programs**

The results of Chapter 5 showed that students use self-regulatory strategies, such as judging information, planning, goal setting, rehearsing, and monitoring, among others, when interacting with the feedback they receive from the AWE program. The idea of self-regulated learning “is associated with forms of learning that are metacognitively guided, at least partly intrinsically motivated, and strategic” (Winne & Perry, 2000, p. 533), and as such provides a framework to understand cognitive, emotional and motivational aspects of learning (Panadero, 2017). Self-regulation is an essential factor in students' ability to effectively use feedback in general (Carless, 2019), and there is evidence that several self-regulatory strategies, such as goal setting, planning, and judging information, are important factors in the adoption and successful use of AWE programs as tools for formative feedback (Hibert, 2019b), and that the use of both cognition and metacognition can help students in the adoption of AWE as tools for formative feedback (Zhang & Hyland, 2018). There is also a strong link between self-regulation and cognitive engagement (Wolters & Taylor, 2012) which has been linked to better outcomes when using AWE programs (Cotos et al., 2020; Liao, 2016a; Xia & Zhong, 2017). The case of George, which I expanded on in section 5.2.3, seems to be consistent with the literature, as he successfully used metacognitive strategies to revise beyond the surface level, rewriting and restructuring sentences based on the AWE feedback and his own understanding of what needed to be improved in his text. While other participants in Chapter 5 did not use the strategies as successfully, they did use most of these in

some capacity, suggesting that improved adoption of AWE feedback could be promoted by helping students develop these skills.

### **6.3.1 Metacognition and cognition in the use of AWE**

Metacognition can be defined as “the awareness learners have about their general academic strengths and weaknesses, cognitive resources they can apply to meet the demands of particular tasks, and their knowledge about how to regulate engagement in tasks to optimise learning processes and outcomes” (Winne & Perry, 2000, p. 533). Liao (2016a) found that the activation of metacognition was key in student adoption of AWE feedback, and the more successful students were those who used metacognitive strategies to mediate the process of writing. The ability to use metacognitive strategies to engage with AWE feedback also seems to be mediated by the proficiency level of individual students (Koltovskaia, 2020).

This seems to be consistent with the findings of the study reported in Chapter 5, which used think-aloud protocols to delve into the self-regulation strategies used by the students to engage with feedback from the AWE program. The findings suggest that students use both metacognitive and cognitive strategies to engage with the AWE feedback, and that the use of these strategies led to the students accepting the ‘good’ feedback offered by the program, while ignoring or disregarding ‘bad’ feedback. Participants would, for example, begin their revising sessions by reflecting on previous feedback to decide which areas to concentrate on. Zhang and Hyland (2018) and Jiang & Yu (2020) noted how students who reflected on feedback and used it for planning revisions were more successful than students who didn’t. My study’s findings went a little further, suggesting that students would also draw on previous knowledge, either of academic papers they had read, or advice they had been given by their supervisors, to make decisions not only of which feedback to accept, but how to guide their revision sessions.

The most commonly used metacognitive strategy used by participants in the study, however, was that of judging information. An important issue with the feedback given by AWE programs is that it is

often unreliable (Bai & Hu, 2017), and in fact the study reported in Chapter 5 found that 30.5% of the AWE feedback received by participants could be considered 'bad' in that it either incorrectly flagged something as a mistake or it correctly flagged a mistake but gave a suggestion that introduced a new mistake. One of the main contributions of Chapter 5 was in using the data from the think-aloud protocols to model the way in which participants judged information, providing an understanding of how students make decisions with regard to the feedback they receive. Among the reasons for rejecting feedback was because it changed their meaning, they do not consider the issue to be a real problem, the feedback was too ambiguous, they do not have the knowledge to evaluate the quality of the feedback, it clashed with their style, among other reasons (see section 5.2.2 for a complete list).

While other research has also looked into feedback uptake, in-depth research into this decision-making process like the one I undertook in chapters 4 and 5 has been scarce. Bai & Hu (2017), for example, researched how students respond to fallible AWE feedback, but only elicited participants' overall perceptions of the quality of the feedback rather than their reasons for accepting or rejecting specific feedback. Chodorow et al. (2010) also studied quality of uptake and found that users were selective in their uptake of feedback, but the study only looked at user actions and did not investigate participants' perceptions or how they reached their conclusion regarding the quality of the feedback like I did by examining the decision-making processes of students in the study reported in Chapter 5. Koltovskaia (2020) did investigate why students made decisions regarding the feedback they received from an AWE program, using stimulated recall to have students talk through their decision-making process when engaging cognitively with AWE feedback. They found that the use of cognitive and metacognitive strategies to make sense of the feedback they received was influenced by their language proficiency, although their sample size was limited to only two participants.

This research is also consistent with the results of the study reported in Chapter 5. Although all the participants had a similar level of proficiency, some like Susan or Melissa seemed to struggle with

metalinguistic knowledge more than others. A very important part of the participants' decision-making process, which I examined in detail in section 5.2.2, relies on drawing upon previous knowledge of English grammar, academic writing conventions, and previous feedback from tutors and supervisors to apply self-regulating strategies such as planning, monitor and judging information, giving strength to the argument that student uptake of feedback is affected by complex factors (Jiang & Yu, 2020). These different factors are important sources of information that the students use to engage with the feedback and judge its applicability to their own writing, as my results suggest. Other cognitive engagement strategies undertaken by the participants in my study included reading, re-reading, rehearsing, and self-questioning, all of which are self-regulatory microprocesses related to cognition (Bannert et al., 2014).

Given the important relationship between feedback provision and self-regulation (Butler & Winne, 1995; Carless, 2019), the results from the study reported in Chapter 5 contribute to existing literature by modelling the self-regulation strategies and decision-making processes of students by looking at the specific case of AWE feedback. However, the results of that study and the study reported in Chapter 4 suggest that, even though students successfully use these strategies to distinguish good feedback from bad, this does not translate in any substantial revision practices, as most of the feedback gets ignored and texts are not changed in any meaningful way. Knowledge of the strategies students use to engage with AWE feedback is important when considering how pedagogical strategies and pedagogical interventions can be used to support the successful integration of AWE programs into the classroom as tools for giving students formative feedback and helping them improve their academic writing skills. Before these pedagogical interventions can be discussed, however, we should discuss what it means to 'successfully adopt' a technological tool and how my findings on self-regulation strategies can help contextualise this process.

### 6.3.2 Self-regulation and technology adoption

Much research has focused on the relationship between self-regulation and the adoption of new technologies in the classroom (Bartolomé & Steffens, 2011; Kitsantas, 2013). In general, for a learner to use a tool they must first recognise that the tool will be useful for their particular learning task, they must be able to use the tool skilfully and they must be motivated to spend the effort required to use it (Winne, 2006). In order for students to successfully adopt the use of technologies for learning, or in this case, revising, they need to be aware of the usefulness of the tool, they need to recognise that the tool can be applied to the task they want to perform, they have to have the skills to use the tool effectively and they need sufficient motivation to use the tool (Gašević et al., 2017). The study in Chapter 4 was consistent with this research in that the participants were indeed aware of the usefulness of the AWE tool and they recognised that the tool could be applied to the task they wanted to perform, given that they uploaded texts to the platform of their own volition. However, the fact that most texts were not changed, or not significantly changed with the use of the AWE program suggests that the students did not have the skills to correctly use the AWE program, a notion supported by the think-aloud protocols collected for the study in Chapter 5 and the general literature on AWE uptake (Bai & Hu, 2017; R. Li et al., 2019; Tian & Zhou, 2020).

There are other factors that influence whether students will adopt new technologies for learning. The technology acceptance model (TAM), for example, posits that social influence processes and cognitive instrumental processes determine the perceived usefulness of technology and intention to use (Venkatesh & Davis, 2000). This model has sometimes been used to understand the factors that influence student adoption of AWE feedback (Lai, 2010; R. Li et al., 2019; Zhai & Ma, 2021). Zhai & Ma, for example, examined environmental, individual, educational, and systemic factors in the adoption of technologies, finding that subjective norm, perceived trust, and cognitive feedback positively influenced the perceived usefulness of the AWE program. In the in-depth examination of student use of AWE programs reported in Chapter 5, perceived usefulness tended to vary between students, with some finding the AWE program very useful for revision and other finding it

frustratingly vague and unhelpful. Perceived usefulness being the first crucial step in technology adoption in the classroom (Winne, 2006), it is worth examining each of these factors through the lens of self-regulation and wider literature on the use of AWE programs, as well as my own findings.

**6.3.2.1 Subjective norm.** Subjective norm refers to the degree to which a student perceives that others believe they should use AWE feedback (Zhai & Ma, 2021). These ‘others’ are usually figures of authority or people who are respected by the student, and therefore their opinions are highly valued. Findings across the literature have very consistently noticed the role of teacher perceptions of the usefulness of AWE tools in student use of these programs. Chen & Cheng (2008), for example, found that teacher attitudes toward the usefulness of AWE programs was a vital factor in student adoption of the tool. On a similar note, Tang & Rich (2017) found that teacher guidance and instruction were key in students being able to use the AWE program to guide their own writing, and noted that teacher training was crucial in the successful introduction of AWE technologies into the classroom.

While neither of the studies that were conducted for this thesis was done as part of an instructional program, some key findings from the study detailed in Chapter 5 are still relevant in the context of the aforementioned literature. First, students referred back to teacher and supervisor feedback not only to judge the validity of the AWE feedback they received, but also to plan their sessions and seek out which AWE functions to use. Second, when encountering difficulties in understanding the feedback or using the program, the participants tended to ask for help from me rather than use the self-help resources offered by ProWritingAid or using a search engine to find answers themselves. Put together, these results hint at the importance of the role of teachers in the use and adoption of AWE programs for revision, as even teachers who were not present or part of the study influenced the perceived usefulness of certain functions within the program. For example, in the study reported in Chapter 5, the student identified as ‘George’ had been constantly told he had an issue with writing sentences that were too long, and so he deliberately sought the functions of the AWE program that

showed him sentence lengths so they could attend to the feedback previously given to them by their supervisors.

**6.3.2.2 Perceived trust.** Perceived trust refers to the confidence students have in the reliability of AWE feedback (Zhai & Ma, 2021). This is an important issue in the use of AWE, as most students find the feedback to be unreliable, generic, and not very useful (Bai & Hu, 2017; C.-F. Chen & Cheng, 2008; Lai, 2010; Zhang, 2017), making it harder for students to use it (Ranalli, 2018). Developing metacognitive abilities is important for students to consider quality of the information they are presented with (Gurbin, 2015), and therefore are important for students to use when faced with AWE feedback that may or may not be reliable. Koltovskaia (2020), for example, found that metacognitive and cognitive engagement with AWE programs helped students make independent judgments on AWE feedback and therefore selectively incorporate that feedback into their texts.

This is consistent with the findings of the study reported in Chapter 5, which showed that ‘judging information’ was the metacognitive strategy most commonly used by participants. This seemed to be necessary since most of the feedback participants received from the AWE program was either “bad” (30.5%) or did not offer any specific actions participants could take to correct flagged mistakes (28.7%). In contrast, only 5.2% of the feedback they received could be considered “good” in that it correctly identified an issue with the writing and offered a way to fix it. Participants expressed their frustration with the quality and ambiguity of the feedback, and this lack of quality is perhaps what led to a relatively low feedback uptake, something that has been widely observed in the literature (Attali, 2004; Bai & Hu, 2017; Chodorow et al., 2010; Koltovskaia, 2020; Ranalli et al., 2017; Tian & Zhou, 2020). Something similar could be observed in the results of the study reported in Chapter 4. While no student perceptions were collected for that study, the uptake of feedback was practically identical to that of the study in Chapter 5, with 67% of feedback being ignored in the former and 68% being ignored in the latter, results that are consistent with some previous studies (Koltovskaia, 2020; Tian & Zhou, 2020), but higher than the 24% ignored feedback reported by Link et al. (2020).



Therefore, it is possible to assume that their perceptions of the validity and trustworthiness of the feedback might have been similar. Furthermore, the fact that most students uploaded several texts to the platform but did not use any of the feedback could be related to this lack of trust in its quality and its dampening effects on student motivation to use the AWE program.

The perceived usefulness of AWE programs as tools for formative feedback is also influenced by how teachers perceive and introduce these programs to the classroom, as well as how much students believe they can trust the feedback given to them. This second point is also related to whether students have the skill to productively use these tools for their revision, drawing attention to the importance of having previous knowledge from various sources to contextualise AWE feedback, as discussed in Chapter 5. It has been argued that students perceive the usefulness of AWE programs and they recognise the situations where they can be used, but sometimes lack the skills to use them and that has an impact on motivation to revise using these tools (El Ebyary & Windeatt, 2019; Huang & Renandya, 2020), which might explain why students in Chapter 4 accessed the program several times but ultimately made very few changes. Therefore, for AWE programs to be successfully integrated into classrooms, any sufficient pedagogical design must take these gaps into account.

#### **6.4 Importance of pedagogical approaches in implementing AWE in the classroom**

The discussion on this chapter has so far been centred around the main research questions guiding this research, namely the usage of AWE programs by students, and some of the theoretical underpinnings through which this usage might be understood and evaluated, supported both by the general literature around the use of AWE and the results of both of the studies conducted as part of this thesis. This theoretical background is important in understanding not only how students are using AWE programs but also in using student experiences to map out ways in which these programs could be implemented as part of pedagogical interventions that are centred on maximising the benefit that students can get from AWE feedback, because students' use of feedback depends on it being introduced in a way that sets up conditions for its use and appreciation (Carless & Winstone,

2020). This section brings all the previous discussion together and uses the results of my research to inform and discuss what these interventions might look like.

First, the discussion will focus on what my results can tell about how teachers and instructors can help scaffold students into taking full advantage of the possibilities of AWE programs as tools to aid in the writing process. Second, I discuss how my results can help inform pedagogical interventions that help students engage cognitively, behaviourally and emotionally with the feedback itself to improve the quality of the revisions they make with the help of these programs.

#### **6.4.1 Using AWE as a tool for providing formative feedback**

The first section of this chapter discussed how my results can be understood from the perspectives of the AWE program's effects on the writing process and the end product of the writing. From a process writing perspective, AWE tools have shown potential in helping students make multiple drafts of their writing (Jiang et al., 2020; J. Li et al., 2015; Zhang & Hyland, 2018) as seen in the results for Chapter 4, as well as helping students with noticing of linguistic forms that can contribute to helping their language proficiency (Liao, 2016b; Parra & Calero, 2019; Xia & Zhong, 2017), as seen in the results from Chapter 5. The question is then, what would a successful integration of AWE programs into the writing process look like.

Chen & Cheng (2008) found that, while the teacher's opinion on the usefulness of AWE programs strongly influenced student perceptions of it, instructional approach was also important, something which has also been noted in general literature on feedback use and integration (Carless & Winstone, 2020; Tracii Ryan et al., 2021; Sutton, 2009). The most successful implementation in Cheng and Chen's (2008) study was that which used AWE to facilitate an early drafting and revising process, rather than as a surrogate writing coach. This is consistent with other literature, which had found that AWE programs can encourage students to revise more often (J. Li et al., 2015; Tang & Rich, 2017) and foster their autonomy and agency (Liao, 2016a; Xia & Zhong, 2017; X. Yang & Dai, 2015).

The findings of this thesis also support that idea. It was found that students used previous teacher feedback to contextualise the feedback they received from the AWE program and make decisions on which areas of the text to focus on when revising, which is consistent with broader literature on feedback, which posits that creating a dialogue between teachers and students helps them make sense of and use the feedback they receive (Carless & Winstone, 2020; Sutton, 2009). While full dialogue could not be carried out in the context of AWE program use, my participants found referring to previous conversations and dialogues about their writing to be very useful in contextualising AWE feedback. This approach led the students to make good decisions in general about which feedback to accept and which to ignore. The participant who made the most extensive revisions to their text, George, did so based on previous teacher feedback about the length of his sentences, which led him to rewrite entire paragraphs using the program's 'sentence length' option. This happened even without specific teacher feedback or intervention related directly to the texts which were being revised, which suggests that specific teacher guidance on which AWE functions may be useful to individual students might help scaffold their use of these tools (Gašević et al., 2017) and help students to use the feedback more effectively to make more substantial revisions.

In fact, research on the use of AWE supports general research on how feedback should be scaffolded by teachers (Dawson et al., 2018), since most successful implementations using AWE in the classroom seem to be those which integrated AWE into a process writing approach, with AWE feedback helping students work on the mechanical and grammar aspects of their texts in the early stages of drafting and revising, and then having teachers provide further feedback on content and organisation after that first stage (Tang & Rich, 2017). This particular approach has been found to lead not only to more accurate texts (Koh, 2017; Liao, 2016b; Y.-J. Wang et al., 2013) but to a better retention of linguistic forms for the students (Z. Li et al., 2017; Link et al., 2020), although the significance of these improvements is still a matter of debate (Lavolette et al., 2015; Long, 2013; Ranalli, 2018).

As discussed in previous sections, the findings of both studies conducted as part of this thesis suggest that tertiary students have some of the prerequisites for using AWE to draft and revise their writing. That is, they recognise the utility of the tool and have the motivation to use it, although they do not necessarily have the skills to use the program to its full potential because of the unreliable nature of the feedback (Bai & Hu, 2017). The need to critically evaluate AWE feedback and successfully apply the many decision-making strategies identified in Chapter 5 means that teachers should introduce these tools in pedagogically sound ways (Hockly, 2019) which create the right environment for students to be able to appreciate and judge AWE feedback (Carless & Winstone, 2020). My results therefore suggest that a holistic approach to introducing AWE programs to the classroom should help students develop the metacognitive skills they need to engage with the feedback and evaluate their own writing, as my results coincide with previous evidence that integrating AWE tools as part of a process writing approach can activate metacognitive and cognitive processes (Liao, 2016a).

These metacognitive and cognitive processes are closely linked to the linguistic proficiency of the students and may be linked to the process of noticing linguistic forms, as our results showed that the noticing process helped students plan which AWE functions they were going to use for their revision sessions. For example, Liao (2016a) noted the importance of developing metacognitive skills to help what they termed 'late bloomers' to notice learning gaps and subsequently activate goal setting and other self-regulatory behaviours. Low proficiency students struggle with critically evaluating and understanding feedback, and their lack of metacognitive skills to engage with this feedback often leads to avoidance strategies (El Ebyary & Windeatt, 2010; Zhang & Hyland, 2018). This could also be seen in participants from the study in Chapter 5, many of whom admitted they did not have the capacity to critically analyse whether the AWE feedback could help them carry on further revisions. This may also explain why the participants in the study reported in Chapter 4 made only superficial revisions, which result from poor linguistic proficiency and lack of self-regulatory strategies to use the feedback (Zhang & Hyland, 2018).

Student ability to engage with feedback does not only rely on self-regulation, but also on a course design that creates the appropriate conditions for its uptake (Carless & Winstone, 2020; Dawson et al., 2018) and on learning strategies that foster engagement, such as requiring minimum AWE scores before submitting to teachers and critically analysing the feedback they receive from the AWE program (C.-F. Chen & Cheng, 2008; Tang & Rich, 2017). Therefore, it is important to consider not only how teaching strategies can scaffold the use of AWE programs as drafting and revising tools, but how these can foster engagement with the feedback itself.

#### **6.4.2 Student engagement with AWE feedback**

Student engagement with the use of AWE programs is an area of research that has been underexplored in current literature (Zhang, 2017), and one of the contributions of this thesis is to explore how self-regulatory processes can help with student engagement with AWE feedback. Engagement is commonly understood as having cognitive, behavioural, and emotional components (Fredricks et al., 2004), all of which feed into each other to bring together different facets of students' academic functioning (Wolters & Taylor, 2012). The multidimensional nature of this construct makes it hard to measure (Whitton & Moseley, 2014), and one of the main hurdles in studying engagement in the field of education is the lack of definitional and conceptual operationalizations of the concept (Azevedo, 2015).

In the field of AWE research, Zhang & Hyland (2018) have operationalized cognitive engagement as the metacognitive and cognitive strategies used to attend to feedback, behavioural engagement as revision actions and time spent on revision, and affective feedback as students' emotional responses and attitudinal reactions to the program. These three types of engagement interact with each other, with cognitive engagement informing behavioural engagement, and both of these being mediated by affective engagement (Zhang, 2020). In Chapter 5, I examined these dimensions of engagement with AWE programs to understand how students used the feedback to revise their texts and, in this

section, I will relate these findings to the wider literature of student engagement with AWE programs.

**6.4.2.1 Cognitive engagement.** When students engage cognitively with the feedback they receive, they make better decisions regarding their uptake (Tian & Zhou, 2020), something that has been noted in previous research (Bai & Hu, 2017; Chodorow et al., 2010) but not explicitly linked to cognitive engagement in AWE research, but has been linked to it in wider research on feedback uptake (Butler & Winne, 1995; Nicol & MacFarlane-Dick, 2006). The findings of the study in Chapter 5 support this notion: students engaged cognitively with the feedback they received from the AWE program, and that engagement manifested itself in a decision-making process that led them to make informed choices regarding the feedback they incorporated into their writing.

The impact of cognitive and metacognitive strategies for the use of AWE feedback has already been discussed, and the use of these types of strategies is considered evidence of cognitive engagement (Wolters & Taylor, 2012). Of these strategies, the most important operations seem to be planning, goal setting, monitoring, and judging information (Liao, 2016a; Zhang, 2020; Zhang & Hyland, 2018). This is consistent with the findings reported in Chapter 5, where judging information was the most commonly-used metacognitive strategy, followed by planning and then monitoring.

Planning allows students to use previous knowledge to inform their use of AWE and the revision sessions they carry out with this tool. In the study reported in Chapter 5, participants used planning strategies to focus on parts of their writing which they had identified as problematic or in need of revision, either because of previous AWE feedback they'd received or areas their supervisors had pointed out needed work. This process is also related to goal-setting. Goal-setting behaviour was the least observed metacognitive strategy in the Chapter 5 findings, but it has also been found to mediate successful use of AWE programs for revision (Liao, 2016a; Zhang & Hyland, 2018), especially when establishing specific goals about improving specific parts of one's writing or noticing specific forms students struggle with.

Judging information was analysed thoroughly in Chapter 5, where the different strategies and decision-making processes used by students to determine the usefulness or appropriateness of AWE feedback were discussed. Students used several sources of previous knowledge to make these judgments, including previous English courses they had attended, feedback from teachers and supervisors, and knowledge about academic writing conventions gained from reading articles and journals. To my knowledge, this is the first study that has attempted to delve into these decision-making processes and understand how students judge information they receive from AWE, but the results of this process are on par with previous studies that have found students tend to make good decisions regarding their uptake of AWE feedback (Bai & Hu, 2017; Chodorow et al., 2010).

Finally, monitoring, or identifying the gap between their current performance and their desired performance, is another crucial component of cognitive engagement. Monitoring is considered to be the gateway to self-regulation due to the internal feedback it generates, creating a standard against which to enact regulation (Winne & Perry, 2000). One of the students who accepted the most ‘good’ feedback in the sample reported in Chapter 5, George, constantly monitored his progress against the standards set to him by his supervisors, and used this knowledge to plan out his revisions sessions. Other students also monitored their progress by comparing their writing not only to the standards set by their supervisors, but to standards set by their knowledge of what academic writing “should look like”. Therefore, in order to help students to cognitively engage with AWE feedback, an instructional strategy must first focus on helping them develop monitoring strategies to identify the gap between their current writing and the desired product, and using their knowledge of that gap to plan not only the drafting process but also the revision process. By developing these monitoring strategies, students can go on to develop other self-regulatory strategies that will help them cognitively engage with the feedback. Second, it should focus on giving students the resources and knowledge they need to adequately judge the relevance of the feedback they receive from AWE programs, given how important these resources were to the decision-making process identified in section 5.2.2. Further research should be made into students’ decision-making processes to

corroborate the findings in this thesis and understand what these resources and knowledge ought to be.

**6.4.2.2 Behavioural engagement.** Behavioural engagement, as has already been mentioned, is related to the specific revision actions students make as a response to AWE feedback. Feedback that does not elicit revisions or responses from students is “useless” (Sutton, 2009), as closing the feedback loop is one of the main principles of effective feedback (Brooks et al., 2019; Nicol & MacFarlane-Dick, 2006). The findings reported in Chapter 4 show that behavioural engagement with AWE feedback is quite low, as the texts analysed underwent very few changes, and the changes that were made to them tended to be quite superficial in nature. This can also be seen in the linguistic composition of the texts themselves, which did not significantly change in any of the markers of linguistic proficiency that were examined in this study. There were also very few cases where participants restructured parts of their texts or rewrote entire sections, and even those were not necessarily in response to AWE feedback, although that doesn’t necessarily mean the AWE feedback was irrelevant (Zhang, 2020). Most research on the use of AWE has found similar results. Uptake of AWE feedback tends to be low, with most of the feedback being ignored by students (Koltovskaia, 2020; Link et al., 2020; Tian & Zhou, 2020), and students usually preferring either peer or teacher feedback (Lai, 2010; Tian & Zhou, 2020).

The relationship between cognitive engagement and behavioural engagement is quite complex, as there can be behavioural engagement without cognitive engagement if students uncritically accept the AWE feedback they are given. However, the research conducted as part of this thesis, as well as existing research into the subjects, suggests that this is not the case for most students who use AWE programs to revise their texts (Bai & Hu, 2017; Chodorow et al., 2010; Koltovskaia, 2020; Tian & Zhou, 2020). Reasons for this low uptake range from the vague, repetitive nature of AWE feedback (Dikli, 2010; Ranalli, 2018), to the students not possessing sufficient knowledge to understand or apply the feedback they receive (Koltovskaia, 2020; Zhang & Hyland, 2018). The question is



therefore not whether behavioural engagement with AWE programs can be improved through instructional design, but how instructional design can foster the kind of cognitive engagement that leads to behavioural engagement (Carless & Winstone, 2020).

**6.4.2.3 Affective engagement.** The affective component was not specifically addressed in this study, although the students who participated on the study reported in Chapter 5 overwhelmingly commented that, despite their frustrations with some aspects of the feedback provided by the AWE program, they would be willing to use a similar tool in the future for revision. While no interview data was collected from participants of the study reported in Chapter 4, the fact that several of them continued using the AWE program months after the study was concluded suggests a similar attitude.

The relationship between affective engagement, behavioural engagement, and cognitive engagement, in regard to AWE usage, appears to be rather complex. While motivation, which is a component of affective engagement (Fredricks et al., 2004), appears to positively influence student adoption of AWE feedback (R. Li et al., 2019; Zhang, 2020), there is also evidence that negative affective engagement with AWE feedback leads students to engage cognitively with it (Koltovskaia, 2020). That is, when students do not blindly trust the feedback they are given by AWE programs, they are more likely to critically evaluate it before accepting or rejecting it, as was the case for the results reported in Chapter 5, where students used a complex decision-making process to evaluate the feedback they received from ProWritingAid.

While student attitudes toward AWE programs reported in Chapter 5 were somewhat ambivalent, previous literature has found that teacher attitudes toward the use of AWE, in turn, influences student attitudes toward these programs (C.-F. Chen & Cheng, 2008; Liao, 2016a; Tang & Rich, 2017). When teachers have positive attitudes toward the role of AWE programs and are able to integrate them into their teaching practices as part of a process writing approach, students tend to hold more positive attitudes toward the role of these programs, which leads to better uptake. One of the main limitations in both studies was that, since the use of the AWE program was not

implemented as part of a structured class activity, there was no insight into how teacher attitudes could impact student perceptions. On the other hand, none of the existing literature has examined how students use AWE programs when given complete freedom and not being constrained by the requirements of a class, so in that sense it helps understand student attitudes independently of any other factors.

## **6.5 Contributions**

Throughout this discussion section, I have been outlining how the results from the studies described in Chapter 4 and Chapter 5 could be contextualised within the wider AWE literature and what lessons could be learned from these results given what we already know from previous research into how students use AWE programs. In this section, I will highlight the specific contributions this thesis makes in light of the above discussion, both methodologically and theoretically.

### **6.5.1 Methodological contributions**

There are two main methodological contributions that I make in this thesis to the study of AWE programs. Each of the studies related in this thesis, reported in Chapter 4 and Chapter 5, was designed to contribute methodologically to the study of AWE programs in a different way.

First of all, the study reported in Chapter 4 innovated both by using linguistic markers of writing proficiency to analyse student texts, and by using the AWE program itself to collect data on student texts and student revision behaviours.

No AWE research to date has used linguistic markers of writing proficiency to measure the effects of these programs on the perceived quality of student revision, even though they have been extensively studied in literature about natural language processing and the development of evaluation tools for ESL and EFL students (Crossley et al., 2016b; Crossley & Kyle, 2018; Kyle, 2016).

The use of these indexes and tools can help give a different perspective of how revision of AWE programs help shape student texts, especially given the constant criticisms of AWE revision being seen as superficial (Bai & Hu, 2017; C.-F. Chen & Cheng, 2008; Z. Li, 2021; Stevenson & Phakiti,

2019). Future research should expand the use of natural language processing tools to get a more in-depth look into how different features of texts are affected by the use of AWE programs for revision.

Using statistical measures to compare changes between versions of the text is another methodological innovation of this thesis. While some previous research has looked at feedback uptake (Bai & Hu, 2017; Chodorow et al., 2010; Koltovskaia, 2020; Tian & Zhou, 2020; M. J. Wang & Goodman, 2012), none of them have made a direct quantitative comparison between pre-AWE feedback texts and post-AWE feedback texts to quantify how much they have changed with revision using AWE feedback.

In the systematic literature review I presented at the EC-TEL conference in 2019 (Hibert, 2019b), I called for the better use of the data collection tools afforded by AWE programs in studying the effects of their feedback on student revisions. Although I developed a web application that could help me collect the data generated by student usage of the AWE program because of the limitations I had in procuring access to AWE programs specifically designed for academic contexts, this bespoke web application helped me collect data not only on the texts themselves, but also some basic data on how the students used the program and how often they accessed some of its functions. As I argued in my 2019 conference paper, more research should harness the data-collection capabilities of AWE programs to gain deeper insights into student revision behaviours and analysing the changes texts undergo because of revision with AWE feedback. This research served as proof of concept that harnessing the data collection capabilities of these programs can help provide insights into student revision and student behaviours when using AWE programs.

The second main methodological contribution of this thesis was the use of think-aloud protocols to study student engagement with AWE programs and decision-making processes when judging the quality of AWE feedback. While several studies have called for the importance of using think-aloud protocols to examine student engagement with AWE programs (Bai & Hu, 2017; Lavolette et al., 2015; Link et al., 2020; Zhang, 2020), to my knowledge this is the only study so far that has used this

data collection technique to understand student engagement with AWE programs. Cotos (2011) has used think-aloud protocols to investigate AWE usage, but she focused on a bespoke, specialised AWE program that was designed to teach students academic writing structure and not general grammar/syntax/style issues like most commercially available AWE programs that are the focus of most research into the subject. El Ebyary and Windeatt (2019) used think-aloud, but it was applied during a stimulated recall rather than using it while the participants used the AWE program.

Furthermore, very little research on the use of AWE programs have used screen recordings to obtain data on how students actually use the AWE programs (Chapelle, 2010; Cotos, 2011; Cotos et al., 2020; El Ebyary & Windeatt, 2019). The combination of using think-aloud protocols and screen recordings to understand student engagement with AWE programs added methodological sophistication and allowed me to have a rich understanding not only of which self-regulatory processes the students used to engage with the feedback, but also the decision-making process they underwent while judging what to do with the AWE feedback they received. This thesis has shown that using think-aloud protocols to study the use of AWE programs is feasible, and future research could expand on the methodology presented here, using think-aloud protocols to examine AWE usage in the context of a writing class, or by setting up experimental conditions for examining revisions with or without AWE feedback.

### **6.5.2. Theoretical contributions**

As with the methodological contributions, each of the studies presented as part of this thesis was designed to make a different theoretical contribution to the overall research on the usage of AWE programs as tools for formative feedback.

The previous section already discussed how natural language processing was used as a novel methodology to understand the changes that texts undergo as a result of revision with AWE programs. While the use of natural language processing to analyse texts is a methodological contribution, the idea of using linguistic markers as proxies for writing proficiency has a strong

theoretical backing (Crossley et al., 2016b; Lu, 2017; McNamara et al., 2010), so far it has not been used at all when talking about the quality and the scope of student changes to texts using AWE feedback. Error rates (Lavolette et al., 2015; Link et al., 2020; Ranalli et al., 2017) and specific grammatical aspects (Z. Li et al., 2017; Long, 2013; Saricaoglu & Bilki, 2021) have been used to examine what students revise using AWE programs, but markers of linguistic proficiency have not yet been used as a theoretical approach to understand how texts change as a result of revision with AWE feedback.

While the results of my research showed that none of these markers changed significantly as a result of revision with AWE feedback, future research can use these to examine more targeted and intensive interventions with AWE programs, and perhaps be used to evaluate the merits of revising given different experimental conditions.

With regards to the study reported in Chapter 5, the theoretical contribution was the use of self-regulation as a theoretical model to understand how students use AWE programs. While some previous studies have used some concepts related to self-regulation to understand how students engage with AWE feedback (Jiang et al., 2020; Koltovskaia, 2020; Zhang & Hyland, 2018), to my knowledge this study is the only one who has specifically used self-regulation theory as a framework to analyse and interpret student engagement with AWE feedback. To this end, I adapted Bannert et al.'s (2014) model for self-regulated learning to make it more relevant for revision rather than learning sessions. Future studies could use and refine my adapted coding of self-regulation in revision activities to understand how students self-regulate their usage of AWE programs and engage with the feedback these programs provide them.

Finally, the results of the study related in Chapter 5 helped me to elaborate a model of how students decide what to do with the feedback they receive. Future research can help expand and refine this model to apply to the use of AWE programs in different contexts. This model can also be used to help create research-informed interventions that help support students in their use of AWE

programs by giving students the resources they need to support these decision-making processes. These resources have been outlined as part of the analysis for chapter 5 and discussed in this chapter, and include previous knowledge about grammar, previous feedback on specific textual characteristics by supervisors, as well as the ability to use different resources to seek help when the AWE feedback is insufficient.

## **6.6 Conclusion**

While most of the research into the use of AWE programs has focused on the end product of the writing process, there is a growing number of studies focusing on the use of AWE programs as part of the writing process. This thesis helps expand on both fields by examining both process, understanding the decision-making processes students use to cognitively engage with AWE feedback, and the product, using novel methods to analyse how texts change as a response to AWE feedback. This chapter has discussed how both product and process approaches help understand the place of AWE programs as tools for providing students with formative feedback and how pedagogical strategies can help students engage cognitively, behaviourally and affectively with AWE feedback.

In order to integrate AWE programs as tools for formative feedback, the most important factor that is derived from this research is helping students develop the metacognitive and cognitive strategies they need to cognitively engage with AWE feedback. Teacher attitudes and their ability to integrate the use of AWE into a process writing approach are key in helping scaffold students into using these skills to revise their work. In the next chapter, I discuss the implications of these findings for the use of AWE in the classroom, as well as some final recommendations for future directions for the field of AWE research.

## Chapter 7 – Conclusion

This thesis has looked at student use of AWE programs from the lenses of the product and the process of student interaction with AWE feedback, using novel data collection and data analysis techniques to study how students use AWE programs in the context of ESL and EFL learners in tertiary education. The preceding chapters have detailed how this research was conducted and the results that were obtained, as well as discussed what these results mean in the light of existing research into the use of AWE programs as tools for formative feedback and the theoretical and methodological contributions of this thesis to the wider realm of AWE research.

In this final chapter, I present the overall conclusions that can be gleaned from the research carried out as part of this thesis. I use the results from the research I conducted during my doctoral studies to provide suggestions for designing research-based implementations of AWE programs as tools for formative feedback in ESL and EFL classrooms. Then, I present some of the limitations of this research and briefly discuss what these limitations mean for the interpretation of these conclusions. Finally, I discuss some implications for future research that can be gleaned from the present research.

### **7.1 – Overall conclusions**

The study reported in Chapter 4, which looked at the product of student interaction with AWE feedback, showed that students were generally willing to use AWE programs to revise their texts, but the revisions were mostly superficial and did not have any significant impact on the linguistic composition of texts. That is, the revised text did not improve in any of the measures that have been found to correlate with human judgments of writing quality and proficiency, meaning that the way in which students use AWE programs does not lead them to produce texts that will help them receive better scores in their writing. While writing is a form of social interaction and the main purpose of acquiring writing skills is to be able to communicate with an audience (Hyland, 2015), it is also true that in ESL and EFL contexts, high scores and good grades put students in more advantageous

positions with regards to future prospects (Hasan & Akhand, 2010; Kadmiry, 2021) and are one of the main goals for many students in taking writing classes and in using AWE programs (Jiang & Yu, 2020), so they should not be discounted.

The study reported in Chapter 5 expands on and gives context to the findings of Chapter 4 by using think-aloud methods and screen captures to delve into student cognitive and behavioural engagement with AWE feedback. Findings from that study showed that students do engage with the feedback they receive and make informed decisions on whether to accept or reject the feedback. It was also found that students made use of a wide range of sources of previous knowledge to make sense of the feedback they received from the AWE program including, among others, teacher feedback, grammar knowledge gained through English classes, and knowledge about academic writing learned through reading academic articles. The main issue in the uptake of feedback found in this study was that most of the feedback they received from the program was either 'bad', in which case students outright ignored it, or so ambiguous that students did not have the background knowledge required to make sense of it or use it to revise their texts.

These findings support the idea that feedback uptake is mediated by complex factors (Jiang & Yu, 2020), and show that students make little use of the feedback they receive from AWE programs, ignoring most of it and making only superficial changes to their texts (Bai & Hu, 2017; Stevenson & Phakiti, 2019). The data collected for this thesis suggests that the potential to use AWE programs is there, as participants demonstrated the willingness and the capacity to engage with the feedback they received. However, lack of background knowledge and tools for critically evaluating and successfully using the feedback were identified as barriers to the participants' usage of AWE feedback.

There are two main implications that can be deduced from the data for the use of AWE programs as tools for formative feedback in ESL and EFL classrooms. First, technology use does not occur in a vacuum, learners need to have knowledge of when and how to use tools to achieve their objectives



(Winne, 2006). Students need to be given the tools, knowledge and resources that allow them to engage with and make sense of the feedback they receive from AWE programs. Second, there is evidence that students' cognitive engagement with AWE programs is an important factor in their uptake of the feedback they receive. Students need to be scaffolded in the use of the cognitive and metacognitive skills that allow them to engage with the AWE feedback. In the next section, I propose five points for guiding pedagogical practices in the use of AWE programs, based on the findings of this thesis.

## **7.2 – Implementation of AWE as tools for formative feedback**

Given the above points, a well-thought-out pedagogical strategy based on supporting student engagement with AWE feedback through the development of self-regulation skills is key for introducing AWE programs into the classroom as tools for formative feedback. While the importance of teacher intervention and pedagogical strategies have often been emphasised in research into the use of AWE programs (C.-F. Chen & Cheng, 2008; Huang & Renandya, 2020; Z. Li, 2021), guidance so far has been vague (Stevenson & Phakiti, 2014). While the research conducted as part of this thesis was not carried out as part of a specific pedagogical intervention in an official academic writing class, the insights I have provided into how students engage with AWE programs can help us glean some specific guidance on how to integrate these programs into ESL and EFL classrooms.

### **7.2.1 Using AWE programs alongside other forms of feedback**

AWE programs have limitations that mean they should not be used as a replacement for teacher and/or peer feedback in the writing process. Research has shown that, when teachers use AWE programs as the sole vehicle for providing feedback, students tend to be dissatisfied and frustrated because AWE feedback is often vague and tends to help mostly with mechanic aspects rather than helping them improve the content and organisation of their texts (C.-F. Chen & Cheng, 2008; Jiang et al., 2020).

My own findings in this thesis showed that most of the feedback that students received from the AWE program was 'bad' in that it incorrectly flagged certain aspects as errors, or it correctly flagged errors but gave suggestions that would introduce a new error into their writing. Although natural language processing has made rapid advances in the last 10 years (Kumar & Boulanger, 2020), we are still not at a stage where computers can understand such nuances as tone, audience, and creative approaches to writing (Vojak et al., 2011), and AWE programs can still be fooled by grammatically correct gibberish (Hockly, 2019). Therefore, AWE programs alone cannot give students a sense of audience or help them develop their voices as writers.

Because of this, current research shows that these programs work best in the initial stages of redrafting and revising, to be followed up by either peer or teacher feedback that allows students to develop a sense of audience and help them with content and organisation (C.-F. Chen & Cheng, 2008; J. Li et al., 2015; Liu & Kunnan, 2016). Teachers can ask students to write their first drafts and then revise them using the AWE program so most of the low-level, grammatical mistakes are dealt with before giving them more comprehensive feedback that allows them to work on the content and organisation of their work.

Using AWE programs for the first few rounds of revision can also help students and teachers become aware of any recurring writing issues identified by the AWE program which can then be addressed in class. Teachers can help promote noticing by encouraging students to evaluate the feedback they receive from the AWE program to evaluate their writing strengths and weaknesses, something that some of the participants in the study reported in Chapter 5 managed to do unprompted.

### **7.2.2 Using AWE programs to set writing and revision goals**

One of the metacognitive strategies that needs to be fostered regarding the use of AWE programs is that of goal setting. AWE programs have scoring functions that allow students to aim for a target score before submitting their work to a teacher or a peer for further feedback, and this has proven to be a good motivator for students and a way to get them to revise their drafts often (Jiang et al.,

2020; Zhang & Hyland, 2018). This can work in concert with the strategy detailed in the previous section, as students can be asked to write their first draft on the AWE program and revise it until they reach a minimum score before turning it into the teacher for additional feedback. This gives students a tangible goal they can work toward, lessens teacher load when dealing with low-level and mechanical errors, and allows students to work on multiple drafts of their texts using different sources of feedback.

However, using machine scores to give the students a goal to aspire to is not the only way in which teachers can help students with goal-setting. The data collected in my thesis showed that students plan their revision sessions around specific aspects of their text they want to correct, such as writing shorter sentences, using more 'academic' vocabulary, fixing the grammar, reducing passive voice, etc.

Research suggests that teacher interventions that are aimed toward helping students understand task conditions and set appropriate goals for them can have more learning benefits (Greene & Azevedo, 2007). Many AWE programs offer diagnostic data that breaks down types of errors and issues that are most common in specific students' writing. Teachers can use this diagnostic data to help students set specific revision goals based on their writing and the course objectives. This diagnostic data can also help teachers plan their lessons geared toward the issues that are most salient in their groups' writing and bring attention to specific writing skills. This will not only help students make better use of the AWE feedback they receive but will give students the tools and standards they need to help monitor their writing and revision processes.

### **7.2.3 Using AWE programs to apply metalinguistic knowledge**

Students need to be given the metalinguistic knowledge necessary to critically evaluate the feedback they receive from AWE programs. Participants in the study reported in Chapter 5 put the AWE feedback they received in context using their metalinguistic knowledge, previous feedback received from teachers and mentors, and the conventions of the genre they had learned from reading

academic papers. This previous knowledge helped participants decide whether to accept or reject the feedback they were given, and in general it resulted in mostly using the 'good' feedback given to them by the AWE program, while ignoring most of the 'bad' feedback.

Most AWE programs include metalinguistic explanations and handbooks that link to the feedback they give, but students do not often make use of these (Koltovskaia, 2020) or do not have the capacity to make use of this feedback (Ranalli, 2018). Teachers need to scaffold the metalinguistic feedback students receive with in-depth explanations to make sure students not only attend to the feedback they receive but also understand it (Woodworth & Barkaoui, 2020). One way to do this is by showing students examples of AWE feedback and explaining what the feedback means, why it was flagged by the AWE algorithm, and ways of fixing the issue identified. Teachers can also use the handbooks and metalinguistic explanations offered by the programs as a starting point to expand on different linguistic issues and writing conventions. By turning this metalinguistic analysis into a classroom activity, teachers can also support noticing of form when using AWE programs.

Related to this previous point, it is possible for teachers to use the failures of the AWE programs as teaching opportunities. In the studies conducted as part of this thesis, participants demonstrated awareness of the weaknesses of the AWE program and would act accordingly, ignoring certain categories of feedback that they had flagged as always being unhelpful or incorrectly flagging mistakes. The studies reported in this thesis were not carried out as part of a pedagogical intervention, but these would have been good opportunities to discuss why the feedback was unhelpful and further scaffold students into critically evaluating the feedback they received. Some teachers have had some success in this regard, having students engage critically with the way AWE feedback works to help with students' metalinguistic awareness and critical analysis of the feedback (Z. Li, 2021; Link et al., 2014).

#### **7.2.4 Teacher training**

Finally, none of this can be achieved without sufficient teacher training, and the results of the research presented in this thesis can give us some understanding of what this teacher training should look like. Tang & Rich (2017) stressed the importance of teacher training in ensuring a successful implementation of AWE programs in the classroom, giving the teachers participating in their research both technical and pedagogical training and support for the use of these programs. I pointed out in Section 2.3.4.2 of the literature review that most studies into the use of AWE either did not report having given teachers any training, specified that no training had been given to teachers, or reported having provided only technical training when implementing the use of AWE into ESL and EFL classrooms. Future implementations and future research both need to acknowledge the importance of theory-based, research-informed teacher training in the integration of AWE programs into the writing process. Training teachers on applying the implementation ideas mentioned in the previous sections can be a good starting point, but there are other areas where teacher training might help in the implementation of AWE programs in ESL and EFL classrooms.

Most AWE programs have functions beyond feedback such as tools for planning essays, dashboards to follow up on student progress, grammar handbooks, portfolios, and other tools that can be integrated into the writing process to help students develop their writing skills. Some participants in my research successfully used these tools to guide their revision processes, so this skill should be promoted in students that use AWE programs to revise their texts. Teacher training should focus not only on making teachers aware of these tools but providing strategies for integrating them into the writing process to help students develop their skills. Teachers could be taught to use AWE programs to track student progress to help them set revision goals, or they could be taught how to use the portfolio functions to help students self-evaluate and monitor their progress as writers.

Teachers could also be taught to use these tools in concert. For example, they could encourage students to make extensive use of the grammar handbooks provided by the AWE programs to make

sense of the feedback they receive or use the grammar handbooks and the AWE feedback as a starting point to discuss some of the most common writing issues faced by the students in a specific classroom, as the results of my research showed that students struggled with making sense of the feedback they received and did not always have sufficient autonomy to seek the answers to their questions for themselves.

Teacher training also needs to make teachers aware of the limitations of AWE feedback. One of the biggest issues the participants in both studies reported in this thesis was that some of the feedback was too vague for them to be able to revise their texts effectively. Teachers need to be made aware of which types of feedback might be vague; for example, in ProWritingAid's case, feedback on passive voice, readability, use of split infinitives and sticky sentences tended to be quite vague. If teachers are made aware of these particular weaknesses, they can devote time in class to discussing these aspects and show students how they can use even this vague feedback to revise. Furthermore, in the previous section I discussed how teachers have used the limitations in AWE feedback as teaching opportunities. Teacher training should focus on making teachers aware of these research-supported practices and helping them implement or adapt these practices to their own classroom environments.

However, some limitations regarding this thesis and its findings should be addressed to put the findings and recommendations of this thesis into context. The next section addresses these limitations and give a brief discussion of how these impact the interpretation of the results and the conclusions presented so far.

### **7.3 - Limitations**

There were several limitations that could affect the conclusions drawn from the analysis presented in this thesis. Some of these limitations are related to the theoretical constructs that were used to inform the design of this research, and others were related to methodological constraints. In this section, I expand on both.

### **7.3.1 Methodological limitations**

One of the main limitations in the studies carried out as part of this thesis were not integrated into existing writing classes or designed as part of pedagogical interventions. While an academic writing workshop was designed for facilitating the data collection detailed in Chapter 4, the fact that it was an extracurricular activity and awarded no university credits meant that participants did not always commit fully to producing the texts required as part of the workshop activities or using the AWE program consistently. Because of the characteristics of the workshop, students used the AWE program freely and with little guidance. Research suggests, and I have argued in this thesis, that integrating AWE programs into the structure of a well-designed instructional strategy results in better use of these tools. However, I still deemed it worthwhile to understand how students use AWE programs on their own in order to identify potential areas in which they could be supported. Furthermore, the product and process of using AWE programs for revision were examined in two different studies. While this was a methodological choice designed to highlight each of these processes and examine them separately, this creates some limitations in the interpretation of the data. For example, we do not have any qualitative data on the students who participated in the study reported in Chapter 4, and therefore we can only infer what their motivations were or why they made the choices they did with regard to the usage of the AWE program and their uptake of the feedback they received. Similarly, since no text data was collected as part of the study reported in Chapter 5, the effect of their revision on their texts could only be inferred. The results of the study reported in Chapter 5 helped interpret and expand on the results obtained in the study reported in Chapter 4, and treating process and product separately helped to examine each one of them in-depth. However, future research might consider examining both of these aspects together in one study to corroborate or disprove the interpretations put forth in this thesis.

Small sample sizes were also an issue in this research. Even though a lot of text data was collected for the study reported in Chapter 4, ultimately only a very reduced number of texts had the

characteristics required for inclusion in the analysis. For the study reported in Chapter 5, only 11 participants were recruited. Although this is a larger number of participants than most of the studies carried out on student engagement (Koltovskaia, 2020; Link et al., 2020; Zhang, 2020; Zhang & Hyland, 2018), it is still a small sample size and so generalisations into other contexts should be done with caution. Furthermore, all participants in both studies were ESL students who had been accepted at a university in the United Kingdom, and therefore can be considered high proficiency. Results may vary in EFL contexts where students are not as immersed in an English-speaking environment.

One final limitation is related to the fact that, when I began working on this thesis, I had little to no programming knowledge and had never used the Python language or the Flask framework to build any applications. While I was able to create GrammarAid with enough data-collection capabilities to support a sufficient analysis in this thesis, the data collection was necessarily limited by my technical knowledge. Future projects could avail themselves of better programming knowledge to create applications that allow for more intricate and fine-grained data collection than the one used in this study.

### **7.3.2 Theoretical limitations**

Another subsection of limitations is related to theoretical constructs that were not explored as part of the research for this thesis because of methodological or time constraints. One of these limitations is that I did not specifically study affective engagement. Since the use of the AWE program in the study reported in Chapter 4 was carried out on the participants' own time and no qualitative data was collected on their experiences, motivation could only be deduced from their activity logs. The study reported in Chapter 5 did collect qualitative data on student experiences and thoughts while using the AWE program. However, this study was carried out in an controlled environment which did not replicate normal usage conditions, as participants were recruited to



partake in four thirty-minute sessions. Therefore, utterances related to affective engagement were so few they were not included in the data analysis for this thesis.

Research on the relationship between affective, cognitive and behavioural engagement is still inconclusive, as some studies have found that negative affective engagement leads to positive cognitive engagement (Koltovskaia, 2020), while others have found that negative affective engagement leads to frustration with the AWE program (Zhang & Hyland, 2018).

Another limitation related to the previous one was that, given the experimental conditions of both studies conducted as part of this thesis and explained above, motivation was also not an element of analysis that was captured by the available data. However, motivation is an important component of technology adoption (Bannert & Mengelkamp, 2008; Clarebout et al., 2013), self-regulation (Winne & Hadwin, 1998) and cognitive engagement (Fredricks et al., 2004). Lack of motivation can lead to students deciding not to use a tool or ignore teacher guidance in the use of said tools (Winne, 2006). While there are several studies focused on student perceptions of AWE programs (Fang, 2010; Lai, 2010; O'Neill & Russell, 2019), motivation as such is a construct that is still underexplored in the literature, and therefore merits looking into.

#### **7.4 – Implications for future research**

An issue that should be considered for future research into the use of AWE programs is the divergent nature of the algorithms that power the feedback and scoring capabilities of these programs. One of the points that was raised in the discussion of this thesis was that of the percentage of feedback accepted by students. My findings show that students rejected around 67% of the feedback they received, compared to 24% in Link et al.'s (2020) study and 75% in Tian and Zhou (2020). While instructional conditions and the particular characteristics of the participants may have affected this uptake rate, another important difference is that all three studies used different AWE programs. Each program uses a proprietary algorithm to generate scores and feedback, which

makes it difficult to know exactly how the AWE program reaches the conclusions it does (Grimes et al., 2010).

Research into the use of AWE programs has focused on programs designed specifically for academic contexts, including ones initially designed for L1 speakers, such as Criterion (Z. Li, 2021; Link et al., 2014, 2020; Ranalli, 2018; Saricaoglu & Bilki, 2021, among others) and MYAccess! (Chen & Cheng, 2008; Fang, 2010, among others), those that have been designed with L2 speakers in mind like Pigai (Huang & Renandya, 2020; Jiang & Yu, 2020; R. Li et al., 2019, among others) and iWrite (Qian et al., 2020), and those that are geared toward more general audiences. Other research into the use of AWE in academic contexts has used programs which were not specifically designed for academic contexts, but with more general audiences in mind, like Grammarly (Koltovskaia, 2020; O'Neill & Russell, 2019) or ProWritingAid (Hibert, 2019a). Future research should consider the different characteristics of various AWE programs when interpreting data pertaining to the research of these programs, as the different engines and algorithms makes direct comparisons between programs and between research using different programs problematic.

Besides addressing the theoretical and methodological limitations described in the previous section, there are several other implications for future research that can be deduced from the findings presented in the previous chapters. This section details some issues that future research into the use of AWE should consider, as well as outlining some new directions that this research should follow to better understand the role of AWE programs as tools for formative feedback in ESL and EFL classrooms.

One of the future directions that research into AWE should follow is the use of NLP tools to analyse the products of student interaction with AWE programs, something I also suggested as part of the systematic literature review I presented at the EC-TEL conference (Hibert, 2019b). As part of this thesis, I used linguistic features as proxies for linguistic proficiency, based on the extensive research that has been done on finding which linguistic characteristics can predict human scores of writing

quality and proficiency. While the findings of this thesis showed that revision with AWE programs did not affect any of these characteristics in any significant way, there is still potential in using these features as a complement to scores and error rates to evaluate the effectiveness of AWE programs, given that these allow to obtain a more fine-grained analysis of how texts change as a response to revision using AWE feedback. Future research could continue exploring the use of these variables to develop more complex validity arguments to assess the impact of using AWE programs as tools for formative feedback in ESL and EFL classrooms.

Another direction that would be promising for research into the use of AWE programs is the use of self-regulation as a theory to inform research design and provide a framework to interpret research findings. Some current research has acknowledged the importance of concepts such as goal-setting (Jiang & Yu, 2020; Zhang & Hyland, 2018) and monitoring (Zhang, 2020; Zhang & Hyland, 2018) in student engagement with AWE programs, but have not explicitly addressed self-regulation. I have argued in this thesis that the constructs of self-regulation can help understand how students engage with AWE feedback, identifying areas of opportunity for scaffolding and instructional support for the use of these programs. I have used the findings from this type of analysis to begin proposing concrete research-informed pedagogical implementations that might help students make better use of AWE programs. The use of trace data to understand student engagement with AWE feedback holds promise in helping develop a more fine-grained and complete understanding of student usage of AWE programs and how to support their implementation in ESL and EFL classrooms.

Finally, future research needs to consider the importance of specific pedagogical strategies in the implementation of AWE programs and student use of the feedback they receive from them.

Targeted teacher training needs to be an integral part of future research, given that technology adoption and tool use requires scaffolding so students understand when and how to use these tools to achieve their purposes (Winne, 2006). If we are to achieve a full understanding of the role AWE programs might play as tools for formative feedback, then research needs to focus on specific

implementations and strategies to determine what works and how best to integrate these technologies into the classroom.

## **7.5 – Final thoughts**

As I mentioned in the paper I presented at the EC-TEL conference, research into the use of AWE is still in its infancy. We have only begun to understand the role these technologies might play as tools for formative feedback in ESL and EFL classrooms, and the question of whether AWE programs can actually help students become better writers is still unresolved. However, there are promising advances and new directions in research into the use of AWE programs which put the students at the centre and try to understand how they engage with and use the feedback they receive from AWE programs. While validity arguments for the use of AWE programs should not be disregarded and research should still focus on that aspect, it is also important to continue to understand how these programs are already being used by students and how we can improve their experience of using them.

The findings in this thesis contribute to the understanding of student usage of AWE programs by exploring both the process and the product of student interaction with AWE programs. On the product side, I proposed new data analysis techniques that might help obtain a broader understanding of how texts change as a response to AWE feedback. On the process side, I used novel data collection techniques such as think-aloud protocols and trace data to deepen our understanding of how students interact and engage with AWE feedback by addressing some of the limitations with data collection expressed in previous research.

The results of the research undertaken as part of this thesis show that there is potential in the use of these AWE programs to help students build revision habits and acquire metalinguistic knowledge that might help them develop their writing skills. The recommendations for integrating these programs into ESL and EFL classrooms put forth in this thesis are only the starting point, and more research needs to address whether these interventions are sufficient or whether different

interventions might be necessary to realise the potential of AWE programs as tools for formative feedback.

## Reference List

- Abuseileek, A. F. (2013). Using track changes and word processor to provide corrective feedback to learners in writing. *Journal of Computer Assisted Learning, 29*(4), 319–333.  
<https://doi.org/10.1111/jcal.12004>
- Ansari, S., Panhwar, A., & Mahesar, G. (2016). Mixed Methods Research: Ontological, Epistemological and Methodological underpinnings. *An International Research Journal of Language and Literature, 27*, 133–141.
- Aryadoust, V., & Liu, S. (2015). Predicting EFL writing ability from levels of mental representation measured by Coh-Metrix: A structural equation modeling study. *Assessing Writing, 24*, 35–58.  
<https://doi.org/10.1016/j.asw.2015.03.001>
- Attali, Y. (2004). Exploring the Feedback and Revision Features of Criterion. Paper presented at. *The National Council on Measurement in Education (NCME)*.
- Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? the role of self-regulated learning. *Educational Psychologist, 40*(4), 199–209.  
<https://doi.org/10.4324/9781315866239-2>
- Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist, 50*(1), 84–94.  
<https://doi.org/10.1080/00461520.2015.1004069>
- Bachman, L. F. (1991). What Does Language Testing Have to Offer? *TESOL Quarterly, 25*(4), 671.  
<https://doi.org/10.2307/3587082>
- Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: how do students respond? *Educational Psychology, 37*(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>
- Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction

- to think aloud and reflect when prompted. Does the verbalisation method affect learning?  
*Metacognition and Learning*, 3(1), 39–58. <https://doi.org/10.1007/s11409-007-9009-6>
- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, 9(2), 161–185. <https://doi.org/10.1007/s11409-013-9107-6>
- Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. *Computer Assisted Language Learning*, 1–24. <https://doi.org/10.1080/09588221.2021.1936071>
- Bartolomé, A., & Steffens, K. (2011). Technologies for Self-Regulated Learning. In *Self-regulated Learning in Technology Enhanced Learning Environments* (pp. 21–31).
- Berger, C., Crossley, S. A., Kyle, K., Journal, S., & April, N. (2017). Using Novel Word Context Measures to Predict Human Ratings of Lexical Proficiency. *Journal of Educational Technology & Society*, 20(2), 201–212.
- Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, 69, 65–78. <https://doi.org/10.1016/j.system.2017.08.004>
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118. <https://doi.org/10.1016/j.jslw.2007.11.004>
- Bitchener, J. (2012). Written Corrective Feedback for L2 Development: Current Knowledge and Future Research. *TESOL Quarterly*, 46(4), 855–860. <https://doi.org/10.1002/tesq.62>
- Bitchener, J., & Knoch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37(2), 322–329. <https://doi.org/10.1016/j.system.2008.12.006>
- Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31(2), 193–214.

<https://doi.org/10.1093/applin/amp016>

Bogolepova, S. V. (2016). НА АНГЛИЙСКОМ ЯЗЫКЕ : ПОДХОДЫ И ПРОДУКТЫ. *Higher Education in Russia, 1*, 87–94.

Bonilla López, M., Van Steendam, E., Speelman, D., & Buyse, K. (2018). The Differential Effects of Comprehensive Feedback Forms in the Second Language Writing Class. *Language Learning, 68*(3), 813–850. <https://doi.org/10.1111/lang.12295>

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment and Evaluation in Higher Education, 38*(6), 698–712.  
<https://doi.org/10.1080/02602938.2012.691462>

British Council. (2022). *Understand and explain the IELTS scores*.  
<https://takeielts.britishcouncil.org/teach-ielts/test-information/ielts-scores-explained>

Brookhart, S. (2008). *How to give effective feedback to your students*. Association for Supervision and Curriculum Development.

Brooks, C., Carroll, A., Gillies, R. M., & Hattie, J. (2019). A matrix of feedback for learning. *Australian Journal of Teacher Education, 44*(4), 14–32. <https://doi.org/10.14221/ajte.2018v44n4.2>

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97–113. <https://doi.org/10.1177/1468794106058877>

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65.  
<https://doi.org/10.1016/j.jslw.2014.09.005>

Burke Johnson, R., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research, 1*(2), 112–113.  
<https://doi.org/10.4324/9781315740447-13>



- Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65(3), 245–281. <https://doi.org/10.3102/00346543065003245>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/l1.1.1>
- Carless, D. (2019). Feedback loops and the longer-term: towards feedback spirals. *Assessment and Evaluation in Higher Education*, 44(5), 705–714. <https://doi.org/10.1080/02602938.2018.1531108>
- Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 1–14. <https://doi.org/10.1080/13562517.2020.1782372>
- Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing*, 44, 51–62. <https://doi.org/10.1016/j.jslw.2019.03.005>
- CCCC. (2006). *Writing assessment: a position statement*.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. *Intercultural Language Use and Language Learning*, 41–57. [https://doi.org/10.1007/978-1-4020-5639-0\\_3](https://doi.org/10.1007/978-1-4020-5639-0_3)
- Celce-Murcia, M., Dornyei, Z., & Thurrell, S. (1995). Communicative Competence: A Pedagogically Motivated Model with Content Specifications. *Issues in Applied Linguistics*, 6(2). <https://doi.org/10.5070/l462005216>
- Chapelle, C. A. (2010). Computer-Assisted Language Learning. *The Oxford Handbook of Applied Linguistics*, (2 Ed.), January 2020, 1–10. <https://doi.org/10.1093/oxfordhb/9780195384253.013.0037>
- Chen, C.-F., & Cheng, W.-Y. 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.
- Chen, C., & Cheng, W. (2006). The Use of a Computer-based Writing Program : Facilitation or Frustration? *23rd International Conference on English Teaching & Learning, May*, 1–15.  
<http://files.eric.ed.gov/fulltext/ED492078.pdf>
- Cheville, J. (2004). Automated Scoring Technologies and the Rising Influence of Error. *The English Journal*, 93(4), 47–52.
- Cho, D. (1999). A study on ESL writing assessment: Intra-rater reliability of ESL compositions. *Melbourne Papers in Language Testing*, 8(1), 1–24.
- Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, 27(3), 419–436. <https://doi.org/10.1177/0265532210364391>
- Choi, W., & Jeong, H. Y. (2016). Finding an appropriate lexical diversity measurement for a small-sized corpus and its application to a comparative study of L2 learners' writings. *Multimedia Tools and Applications*, 75(21), 13015–13022. <https://doi.org/10.1007/s11042-015-2529-1>
- Clarebout, G., Elen, J., Juarez Collazo, N. A., Lust, G., & Jiang, L. (2013). Metacognition and the Use of Tools. In *International Handbook of Metacognition and Learning Technologies* (pp. 187–195).  
[https://doi.org/10.1007/978-1-4419-5546-3\\_13](https://doi.org/10.1007/978-1-4419-5546-3_13)
- Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. *Research Methods in Education*, 245–284. <https://doi.org/10.1016/B978-0-12-813029-2.00009-8>
- Cohen, Y. (2017). Estimating the Intra-Rater Reliability of Essay Raters. *Frontiers in Education*, 2, 1–11. <https://doi.org/10.3389/feduc.2017.00049>
- Collaço, C. (2017). Increasing student engagement in higher education. *Journal of Higher Education*

*Theory and Practice*, 17(4), 40–47. <https://doi.org/10.1002/berj.3121>

Cotos, E. (2011). Potential of Automated Writing Evaluation Feedback. *CALICO Journal*, 28(2), 420–459.

Cotos, E. (2012). Towards effective integration and positive impact of automated writing evaluation in L2 Writing. In *English Publications* (Vol. 60, pp. 81–112).

Cotos, E., Huffman, S., & Link, S. (2020). Understanding graduate writers' interaction with and impact of the research writing tutor during revision. *Journal of Writing Research*, 12(1), 187–232. <https://doi.org/10.17239/JOWR-2020.12.01.07>

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41(4), 965–981. <https://doi.org/10.1016/j.system.2013.08.002>

Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing*, 38, 46–50. <https://doi.org/10.1016/j.asw.2018.06.004>

Crossley, S. A., Kyle, K., Allen, L. K., Guo, L., & McNamara, D. S. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in Automated Writing Evaluation. *The Journal of Writing Assessment*, 7(1), 1–34. <https://doi.org/10.5204/ssj.v7i1.330>

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods*, 51(1), 14–27. <https://doi.org/10.3758/s13428-018-1142-4>

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016a). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1–16. <https://doi.org/10.1016/j.jslw.2016.01.003>

- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016b). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48(4), 1227–1237. <https://doi.org/10.3758/s13428-015-0651-7>
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18(2), 119–135. <https://doi.org/10.1016/j.jslw.2009.02.002>
- Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, 21(2–3), 170–191. <https://doi.org/10.1504/IJCEELL.2011.040197>
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115–135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66–79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243–263. <https://doi.org/10.1177/0265532211419331>
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561–580. <https://doi.org/10.1177/0265532210378031>
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561–580.

<https://doi.org/10.1177/0265532210378031>

Dawson, P., Henderson, M., Ryan, T., Mahoney, P., Boud, D., Phillips, M., & Molloy, E. (2018).

Tecnology and Feedback Design. In J. M. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design and Technology: An International Compendium of Theory, Research, Practice and Policy*. Springer.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1), 1–35.

Dikli, S. (2010). The Nature of Automated Essay Scoring Feedback. *CALICO Journal*, 28(1), 99–134.

<https://doi.org/10.1017/CBO9781107415324.004>

El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, 10(2), 121–142.

<https://doi.org/10.6018/ijes/2010/2/119231>

El Ebyary, K., & Windeatt, S. (2019). Eye tracking analysis of EAP Students' regions of interest in computer-based feedback on grammar, usage, mechanics, style and organization and development. *System*, 83, 36–49. <https://doi.org/10.1016/j.system.2019.03.007>

Ellis, R. (1998). Teaching and Research : Options in Grammar Teaching. *TESOL Quarterly*, 32(1), 39–60.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107.

<https://doi.org/10.1093/elt/ccn023>

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36(3), 353–371. <https://doi.org/10.1016/j.system.2008.02.001>

Fan, Y., Saint, J., Singh, S., Jovanović, J., & Gašević, D. (2021). A learning analytic approach to

- unveiling self-regulatory processes in learning tactics. *ACM International Conference Proceeding Series*, 184–195. <https://doi.org/10.1145/3448139.3448211>
- Fang, Y. (2010). Perceptions of the computer-assisted writing program among EFL college learners. *Educational Technology and Society*, 13(3), 246–256.
- Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4(1), 6–16. <https://doi.org/10.1177/1558689809349691>
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1–11.
- Ferris, D. (2004). The “Grammar Correction” Debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime ...?). *Journal of Second Language Writing*, 13(1), 49–62. <https://doi.org/10.1016/j.jslw.2004.04.005>
- Ferris, D. (2019). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing: Contexts and Issues* (pp. 81–108). Cambridge University Press. <https://doi.org/10.4324/9781410607201-9>
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 34(4), 365–387.
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, Diagnosis, and the Strategies of Revision. *College Composition and Communication*, 37(1), 16–55. <https://doi.org/10.2307/357381>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>

- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1). <https://doi.org/10.1007/s11528-014-0822-x>
- Gašević, D., Mirriahi, N., Dawson, S., & Joksimović, S. (2017). Effects of instructional conditions and experience on the adoption of a learning tool. *Computers in Human Behavior*, 67, 207–220. <https://doi.org/10.1016/j.chb.2016.10.026>
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36(2), 193–202. <https://doi.org/10.3758/BF03195564>
- Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, 77(3), 334–372. <https://doi.org/10.3102/003465430303953>
- Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, 34(1), 18–29. <https://doi.org/10.1016/j.cedpsych.2008.05.006>
- Greene, J. A., Robertson, J., & Costa, L. C. (2011). Assessing Self-Regulated Learning Using Think-Aloud Methods. In *Handbook of Self-Regulation of Learning and Performance* (pp. 313–328).
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. <https://doi.org/10.3102/01623737011003255>
- Grimes, D., Warschauer, M., & Russell, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment*, 8(6), 4–43.
- Gunther, C. W., & Rozinat, A. (2012). Disco: Discover your processes. *BPM (Demos)*, 40–44.
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in

- both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18(3), 218–238. <https://doi.org/10.1016/j.asw.2013.05.002>
- Gurbin, T. (2015). Metacognition and Technology Adoption: Exploring Influences. *Procedia - Social and Behavioral Sciences*, 191, 1576–1582. <https://doi.org/10.1016/j.sbspro.2015.04.608>
- Gwet, K. L. (2008). Intrarater Reliability. *Wiley StatsRef: Statistics Reference Online*. <https://doi.org/10.1002/9781118445112.stat06882>
- Ha, H. S. (2019). Lexical Richness in EFL Undergraduate Students' Academic Writing. *English Teaching*, 74(3), 3–28. <https://doi.org/10.15858/engtea.74.3.201909.3>
- Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2(2–3), 107–124. <https://doi.org/10.1007/s11409-007-9016-7>
- Hall, R. (2013). Mixed Methods: In Search of a Paradigm. *Conducting Research in a Changing and Challenging World*, 71–78. <https://doi.org/10.1007/s13384-015-0169-0>
- Harrits, G. S. (2011). More than method?: A discussion of paradigm differences within mixed methods research. *Journal of Mixed Methods Research*, 5(2), 150–166. <https://doi.org/10.1177/1558689811402506>
- Hasan, M. K., & Akhand, M. M. (2010). Approaches to writing in EFL/ESL context: Balancing product and process in writing class at tertiary level. *Journal of NELTA*, 15(1–2), 77–88.
- Hattie, J., & Gan, M. (2015). Instruction Based on Feedback. *Handbook of Research on Learning and Instruction*, 249–271. <https://doi.org/10.4324/9780203839089.ch13>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, 29(3), 369–388.



<https://doi.org/10.1177/0741088312451260>

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, 16(2), 416–431.

<https://doi.org/10.1017/S0958344004001120>

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers and Education*, 90, 36–53.

<https://doi.org/10.1016/j.compedu.2015.09.005>

Hibert, A. I. (2019a). Metacognitive Processes and Self-regulation in the Use of Automated Writing Evaluation Programs. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming Learning with Meaningful Technologies .EC-TEL 2019. Lecture Notes in Computer Science* (Vol. 11722, pp. 665–668). Springer International Publishing.

<https://doi.org/10.1007/978-3-030-29736-7>

Hibert, A. I. (2019b). Systematic Literature Review of Automated Writing Evaluation as a Formative Learning Tool. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming Learning with Meaningful Technologies .EC-TEL 2019. Lecture Notes in Computer Science* (Vol. 11722, pp. 119–212). Springer International Publishing.

<https://doi.org/10.1007/978-3-030-29736-7>

Hockly, N. (2019). Automated writing evaluation. *ELT Journal*, 73(1), 82–88.

<https://doi.org/10.1093/elt/ccy044>

Hongwei, W., & Liqin, Y. (2013). A Computational Analysis of Textual Features and L2 Writing Proficiency. *International Journal of Academic Research in Progressive Education and Development*, 2(4), 170–185. <https://doi.org/10.6007/ijarped/v2-i4/396>

Horowitz, D. (1986). Process , Not Product : Less Than Meets the Eye. *TESOL Quarterly*, 20(1), 141–144.

Huang, S., & Renandya, W. A. (2020). Exploring the integration of automated feedback among lower-

- proficiency EFL learners. *Innovation in Language Learning and Teaching*, 14(1), 15–26.  
<https://doi.org/10.1080/17501229.2018.1471083>
- Hyland, K. (2003). Writing and teaching writing. In K. Hyland (Ed.), *Second Language Writing* (pp. 1–30). Cambridge University Press. <https://doi.org/10.1080/13586840701442950>
- Hyland, K. (2015). *Teaching and Researching Writing*. Routledge.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>
- Hyland, K., & Hyland, F. (2019). Contexts and Issues in Feedback on L2 Writing. In K. Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing* (pp. 1–22). Cambridge University Press.
- Jaiswal, N. (2019). SequenceMatcher in Python. *Towards Data Science*.
- Jiang, L., & Yu, S. (2020). Appropriating automated feedback in L2 writing: experiences of Chinese EFL student writers. *Computer Assisted Language Learning*, 0(0), 1–25.  
<https://doi.org/10.1080/09588221.2020.1799824>
- Jiang, L., Yu, S., & Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System*, 93, 102302.  
<https://doi.org/10.1016/j.system.2020.102302>
- Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education*, 33, 74–85.  
<https://doi.org/10.1016/j.iheduc.2017.02.001>
- Jung, Y. J., Crossley, S. A., & McNamara, D. S. (2019). Predicting second language writing proficiency in learner texts using computational tools. *Journal of Asia TEFL*, 16(1), 37–52.  
<https://doi.org/10.18823/asiatefl.2019.16.1.3.37>
- Kadmiry, M. (2021). The Comparison between the Process-Oriented Approach and the Product-

- oriented Approach in Teaching Writing: The Case of Moroccan EFL Students in Preparatory Classes for the Grandes Ecoles. *Arab World English Journal*, 12(1), 198–214.
- Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal*, 99(1), 1–18.  
<https://doi.org/10.1111/modl.12189>
- Karadenizli-Çilingir, M. N. (2019). *The effect of standardisation sessions conducted before English language writing exams on inter-rater and intra-rater reliability* (Issue August). Middle East Technical University.
- Kardong-Edgren, S., Oermann, M. H., Rizzolo, M. A., & Odom-Maryon, T. (2017). Establishing inter- and intrarater reliability for high-stakes testing using simulation. *Nursing Education Perspectives*, 38(2), 63–68. <https://doi.org/10.1097/01.NEP.0000000000000114>
- Kato, T. (2019). *Constructing a Measurement Model of L2 Complexity in Automated Essay Scoring for Japanese EFL Learners' Writing : Toward a Qualitative and Analytic Evaluation*. 1351–1356.
- Kayapinar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113–136. <https://doi.org/10.14689/ejer.2014.57.2>
- Kim, J. (2014). Predicting L2 Writing Proficiency Using Linguistic Complexity Measures: A Corpus-Based Study. *English Teaching*, 69(4), 27–51. <https://doi.org/10.15858/engtea.69.4.201412.27>
- Kim, M., & Crossley, S. A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing*, 37, 39–56. <https://doi.org/10.1016/j.asw.2018.03.002>
- Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical Sophistication as a Multidimensional Phenomenon: Relations to Second Language Lexical Proficiency, Development, and Writing Quality. *Modern Language Journal*, 102(1), 120–141. <https://doi.org/10.1111/modl.12447>

- Kitsantas, A. (2013). Fostering college students' self-regulated learning with learning technologies. *Hellenic Journal of Psychology, 10*(3), 235–252.
- Koh, W.-Y. (2017). Effective Applications of Automated Writing Feedback in Process-based Writing Instruction. *English Teaching, 72*(3), 91–118. <https://doi.org/10.15858/engtea.72.3.201709.91>
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*.  
<https://doi.org/10.1016/j.asw.2020.100450>
- Kuiper, C., Smit, J., de Wachter, L., & Elen, J. (2017). Scaffolding tertiary students' writing in a genre-based writing intervention. *Journal of Writing Research, 9*(1), 27–59.  
<https://doi.org/10.17239/jowr-2017.09.01.02>
- Kumar, V. S., & Boulanger, D. (2020). Automated Essay Scoring and the Deep Learning Black Box: How Are Rubric Scores Determined? *International Journal of Artificial Intelligence in Education*.  
<https://doi.org/10.1007/s40593-020-00211-5>
- Kyle, K. (2016). *Measuring syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication (doctoral dissertation)* [Georgia State University]. [https://scholarworks.gsu.edu/alesl\\_diss/35](https://scholarworks.gsu.edu/alesl_diss/35)
- Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly, 49*(4), 757–786. <https://doi.org/10.1002/tesq.194>
- Kyle, K., & Crossley, S. A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing, 34*, 12–24.  
<https://doi.org/10.1016/j.jslw.2016.10.003>
- Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods, 50*(3), 1030–1046.  
<https://doi.org/10.3758/s13428-017-0924-4>

- Lai, Y. H. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, 41(3), 432–454. <https://doi.org/10.1111/j.1467-8535.2009.00959.x>
- Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, 19(2), 50–68.
- Lehmann, C. (2007). Linguistic competence: Theory and empiry. *Folia Linguistica*, 41(3–4), 223–278. <https://doi.org/10.1515/flin.41.3-4.223>
- Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18. <https://doi.org/10.1016/j.jslw.2014.10.004>
- Li, R., Meng, Z., Tian, M., Zhang, Z., Ni, C., & Xiao, W. (2019). Examining EFL learners' individual antecedents on the adoption of automated writing evaluation in China. *Computer Assisted Language Learning*, 32(7), 784–804. <https://doi.org/10.1080/09588221.2018.1540433>
- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, 102505. <https://doi.org/10.1016/j.system.2021.102505>
- Li, Z., Feng, H. H., & Saricaoglu, A. (2017). The short-term and long-term effects of AWE feedback on ESL students' development of grammatical accuracy. *CALICO Journal*, 34(3), 355–375. <https://doi.org/10.1558/cj.26382>
- Liao, H. C. (2016a). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62, 77–92. <https://doi.org/10.1016/j.system.2016.02.007>
- Liao, H. C. (2016b). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, 70(3), 308–319. <https://doi.org/10.1093/elt/ccv058>

- Lichtinger, E., & Kaplan, A. (2011). Purpose of engagement in academic self-regulation. *New Directions for Teaching and Learning*, 126, 9–19. <https://doi.org/10.1002/tl.440>
- Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards best ESL practices for implementing automated writing evaluation. *CALICO Journal*, 31(3), 323–344. <https://doi.org/10.11139/cj.31.3.323-344>
- Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2020.1743323>
- Liu, S., & Kunnan, A. J. (2016). Investigating the application of automated writing evaluation to Chinese undergraduate English majors: A case study of writetolearn. *CALICO Journal*, 33(1), 71–91. <https://doi.org/10.1558/cj.v33i1.26380>
- Long, R. (2013). A review of ETS's Criterion online writing program for student compositions. *The Language Teacher*, 37(3), 11. <https://doi.org/10.37546/jalttl37.3-2>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493–511. <https://doi.org/10.1177/0265532217710675>
- Luo, Q., & Zhou, J. (2017). Data-driven learning in second language writing class: A survey of empirical studies. *International Journal of Emerging Technologies in Learning*, 12(3), 182–196. <https://doi.org/10.3991/ijet.v12i03.6523>
- Mahfoodh, O. H. A. (2017). “I feel disappointed”: EFL university students’ emotional responses towards teacher written feedback. *Assessing Writing*, 31, 53–72. <https://doi.org/10.1016/j.asw.2016.07.001>

- Matthews, J., & Wijeyewardene, I. (2018). Exploring relationships between automated and human evaluations of L2 texts. *Language Learning and Technology*, 22(3), 143–158.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86. <https://doi.org/10.1177/0741088309351547>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press.
- Mehrabi-Yazdi, O. (2018). Short communication on the missing dialogic aspect of an automated writing evaluation system in written feedback research. *Journal of Second Language Writing*, 41, 92–97. <https://doi.org/10.1016/j.jslw.2018.05.004>
- Molloy, E., & Boud, D. (2013). Changing conceptions of feedback. In *Feedback in higher and professional education* (pp. 11–33). Routledge.
- Neupane, P. (2017). Approaches to Teaching English Writing : A Research Note. *Studies in Foreign Language Education*, 39, 141–148.
- Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- O’Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students’ perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42–56. <https://doi.org/10.14742/ajet.3795>
- Page, E. (1966). Grading Essays by Computer: Progress Report. *Invitational Conference on Testing Problems*, 87–100. [www.journal.uta45jakarta.ac.id](http://www.journal.uta45jakarta.ac.id)
- 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(July), 255–270. <https://doi.org/10.1016/j.cedpsych.2018.07.002>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00422>
- Pardo, A. (2018). A feedback model for data-rich learning experiences. *Assessment and Evaluation in Higher Education*, 43(3), 428–438. <https://doi.org/10.1080/02602938.2017.1356905>
- Parra, L., & Calero, X. (2019). Automated Writing Evaluation Tools in the Improvement of the Writing Skill. *International Journal of Instruction*, 12(2), 209–226.
- Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>
- Pillar, G. (2011). A framework for testing communicative competence. *The Round Table: Partium Journal of English Studies*, 2, 1–17.
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92, 544–555. <https://doi.org/10.1037/0022-0663.92.3.544>
- Pluye, P., & Hong, Q. N. (2014). Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annual Review of Public Health*, 35, 29–45. <https://doi.org/10.1146/annurev-publhealth-032013-182440>
- Python Software Foundation. (2020). *difflib — Helpers for computing deltas*. Python Documentation. <https://docs.python.org/3/library/difflib.html#difflib.SequenceMatcher.ratio>
- Qian, L., Zhao, Y., & Cheng, Y. (2020). Evaluating China’s Automated Essay Scoring System iWrite.



*Journal of Educational Computing Research*, 58(4), 771–790.

<https://doi.org/10.1177/0735633119881472>

Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it?

*Computer Assisted Language Learning*, 31(7), 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), 8–25.

<https://doi.org/10.1080/01443410.2015.1136407>

Robinson, P. (1995). Attention, Memory, and the “Noticing” Hypothesis. *Language Learning*, 45, 283–331.

Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior*, 70, 207–221. <https://doi.org/10.1016/j.chb.2016.12.076>

Ryan, Thomas, & Bagley, G. (2015). Nurturing the integration of technology in education. *Journal of Theory & Practice in Education*, 11(1), 33–50.

[http://proxy.lib.odu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=e  
hh&AN=100932723&site=ehost-live&scope=site](http://proxy.lib.odu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=e<br/>hh&AN=100932723&site=ehost-live&scope=site)

Ryan, Tracii, Henderson, M., Ryan, K., & Kennedy, G. (2021). Identifying the components of effective learner-centred feedback information. *Teaching in Higher Education*.

<https://doi.org/10.1080/13562517.2021.1913723>

Ryu, J. (2020). Predicting Second Language Writing Proficiency in the Different Genres of Writing Using Computational Tools. *Korean Journal of Applied Linguistics*, 36(1), 141–170.

<https://doi.org/10.17154/kjal.2020.3.36.1.141>

- Saadi, Z. K., & Saadat, M. (2015). EFL Learners' Writing Accuracy: Effects of Direct and Metalinguistic Electronic Feedback. *Theory and Practice in Language Studies*, 5(10), 2053.  
<https://doi.org/10.17507/tppls.0510.11>
- Saint, J., Whitelock-Wainwright, A., Gašević, D., & Pardo, A. (2020). Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning from Trace Data. *IEEE Transactions on Learning Technologies*, 13(4), 861–877. <https://doi.org/10.1109/TLT.2020.3027496>
- Saldaña, J. (2016). *The coding manual for qualitative researchers*. Sage Publications.
- Saldaña, J., Leavy, P., & Beretvas, N. (n.d.). *Fundamentals of Qualitative Research*. Oxford University Press.
- Samuel Barkin, J. (2015). Translatable? On mixed methods and methodology. *Millennium: Journal of International Studies*, 43(3), 1003–1006. <https://doi.org/10.1177/0305829815581534>
- Saricaoglu, A., & Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL*, 1–13. <https://doi.org/10.1017/S0958344021000021>
- Savignon, S. J. (2005). Communicative language teaching: strategies and Goals. In *Handbook of research in second language teaching and learning* (pp. 659–676). Routledge.
- Savignon, S. J. (2017). Communicative Competence. *The TESOL Encyclopedia of English Language Teaching*, 1–7. <https://doi.org/10.1002/9781118784235.eelt0047>
- Schellings, G., & van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: Theoretical and empirical considerations. *Metacognition and Learning*, 6(2), 83–90.  
<https://doi.org/10.1007/s11409-011-9081-9>
- Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers and Education*, 128, 13–35.

<https://doi.org/10.1016/j.compedu.2018.09.009>

Shao, X. (2015). On written corrective feedback in L2 writing. *English Language Teaching*, 8(3), 155–168. <https://doi.org/10.5539/elt.v8n3p155>

Sheehan, S. (2013). *British Council ELT Research Papers*.

Steele, V. (1992). *Product and process writing: a comparison*. Newbury House.

Stenlund, T. (2013). Agreement in assessment of prior learning related to higher education: an examination of interrater and intrarater reliability. *International Journal of Lifelong Education*, 32(4), 535–547. <https://doi.org/10.1080/02601370.2013.778077>

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>

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65. <https://doi.org/10.1016/j.asw.2013.11.007>

Stevenson, M., & Phakiti, A. (2019). Automated feedback and second language writing. In *Feedback in Second Language Writing: Contexts and Issues* (pp. 125–142). <https://doi.org/10.1017/9781108635547.009>

Sutton, P. (2009). Towards dialogic feedback. *Critical and Reflective Practice in Education*, 1(1), 1–10.

Tang, J., & Rich, C. S. (2017). Automated writing evaluation in an EFL setting: Lessons from China. *JALT CALL Journal*, 13(2), 117–146.

Thierbach, C., Hergesell, J., & Baur, N. (2020). Mixed methods research. In *SAGE Research Methods Foundations* (pp. 1–26). <https://doi.org/10.1002/9781119410867.ch12>

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, 91.

<https://doi.org/10.1016/j.system.2020.102247>

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272. <https://doi.org/10.1016/j.jslw.2007.06.003>

Truscott, J., & Hsu, A. Y. ping. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17(4), 292–305. <https://doi.org/10.1016/j.jslw.2008.05.003>

Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning*, 6(2), 205–211. <https://doi.org/10.1007/s11409-011-9080-x>

Veenman, M. V. J. (2013). Assessing Metacognitive Skills in Computerized Learning Environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 157–168). Springer. <https://doi.org/10.1007/978-1-4419-5546-3>

Venkatesh, V., & Davis, F. D. (2000). Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>

Vojak, C., Kline, S., Cope, B., McCarthey, S., & Kalantzis, M. (2011). New Spaces and Old Places: An Analysis of Writing Assessment Software. *Computers and Composition*, 28, 97–111. <https://doi.org/10.1016/j.compcom.2011.04.004>

Waer, H. (2021). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching*, 0(0), 1–25. <https://doi.org/10.1080/17501229.2021.1914062>

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(May 2019), 100449.  
<https://doi.org/10.1016/j.asw.2020.100449>
- Wang, M. J., & Goodman, D. (2012). Automated writing evaluation: students' perceptions and emotional involvement. *English Teaching & Learning*, 36(3), 1e37. *English Teaching & Learning*, 36(3), 1–37. <https://doi.org/10.6330/ETL.2012.36.3.01>
- Wang, P. (2013). Can automated writing evaluation programs help students improve their English writing? *International Journal of Applied Linguistics and English Literature*, 2(1), 6–12.  
<https://doi.org/10.7575/ijalel.v.2n.1p.6>
- Wang, S., & Li, R. (2019). An Empirical Study on the Impact of an Automated Writing Assessment on Chinese College Students' English Writing Proficiency. *International Journal of Language and Linguistics*, 7(5), 218. <https://doi.org/10.11648/j.ijll.20190705.16>
- Wang, Y.-J., Shang, H.-F. F., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 234–257. <https://doi.org/10.1080/09588221.2012.655300>
- White, R. (1988). Academic Writing: Process and product. In *Academic writing: Process and product* (pp. 4–16).
- Whitton, N., & Moseley, A. (2014). Deconstructing Engagement: Rethinking Involvement in Learning. *Simulation and Gaming*, 45, 433–449. <https://doi.org/10.1177/1046878114554755>
- 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>

Winne, P. H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist*, 41(1), 5–17. [https://doi.org/10.1207/s15326985ep4101\\_3](https://doi.org/10.1207/s15326985ep4101_3)

Winne, P. H. (2011). A Cognitive and Metacognitive Analysis of Self-Regulated Learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 15–32). Routledge. <https://doi.org/10.4324/9780203839010.ch2>

Winne, P. H. (2013). Learning Strategies, Study Skills, and Self-Regulated Learning in Postsecondary Education. In M. B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research* (pp. 377–403). [https://doi.org/10.1007/978-94-007-5836-0\\_8](https://doi.org/10.1007/978-94-007-5836-0_8)

Winne, P. H. (2017). Learning Analytics for Self-Regulated Learning. In *Handbook of Learning Analytics* (pp. 241–249). <https://doi.org/10.18608/hla17.021>

Winne, P. H., & Hadwin, A. F. (1998). Studying as Self-Regulated Learning. In *Metacognition in Educational Theory and Practice* (pp. 291–318). Routledge. <https://doi.org/10.4324/9781410602350-19>

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Academic Press.

Wolters, C., & Taylor, D. (2012). A Self-regulated Learning Perspective on Student Engagement. In *Handbook of Research on Student Engagement* (pp. 635–651). Springer. <https://doi.org/10.1007/978-1-4614-2018-7>

Woodworth, J., & Barkaoui, K. (2020). Perspectives on Using Automated Writing Evaluation Systems to Provide Written Corrective Feedback in the ESL Classroom. *TESL Canada Journal*, 37(2), 234–247. <https://doi.org/10.18806/tesl.v37i2.1340>

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading?

*Language Testing*, 27(3), 291–300. <https://doi.org/10.1177/0265532210364643>

Xia, L., & Zhong, L. (2017). 作文自动评价系统在大学英语写作教学中的实证研究. In *Research in Teaching* (Vol. 40, Issue 1, pp. 57–61).

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53–67. <https://doi.org/10.1016/j.jslw.2015.02.002>

Yang, X., & Dai, Y. (2015). 基于批改网的大学英语自主写作教学模式实践研究. *外语电化教学*, 162, 17–23.

Yoon, H. J. (2018). The Development of ESL Writing Quality and Lexical Proficiency: Suggestions for Assessing Writing Achievement. *Language Assessment Quarterly*, 15(4), 387–405. <https://doi.org/10.1080/15434303.2018.1536756>

Zhai, N., & Ma, X. (2021). Automated writing evaluation ( AWE ) feedback : a systematic investigation of college students ' acceptance. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2021.1897019>

Zhang, V. (2017). Student engagement with computer-generated feedback : a case study. *ELT Journal*, 71(3), 317–328. <https://doi.org/10.1093/elt/ccw089>

Zhang, V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, 43(December). <https://doi.org/10.1016/j.asw.2019.100439>

Zhang, V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36(February), 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>

Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, 37, 13–24.

<https://doi.org/10.1016/j.asw.2018.03.001>

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology, 91*(2), 241–250.

<https://doi.org/10.1037/0022-0663.91.2.241>



## Appendixes

### Appendix 1 – Slides used during the academic writing workshop

#### Week 1

# Academic writing workshop

## Week 1

- Welcome
- What is an academic text?
- How is academic writing different from other genres?
- General academic writing conventions
- Anatomy of the academic writing text
- The introduction

# What is an academic text?

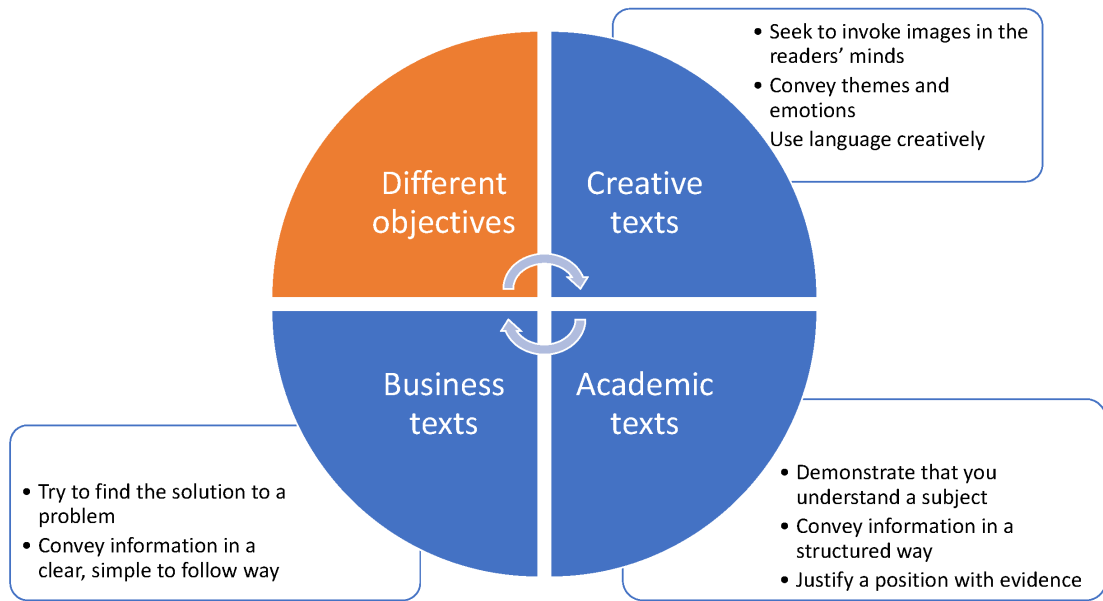
## Definition

- A text whose main purpose is to showcase knowledge and present evidence-based arguments
- It is always looking to defend a main idea as objectively as possible
- In order to be successful, there needs to be a planning process beforehand

## Types of academic texts

- Academic essay
- Dissertation
- Book chapters
- Conference papers
- Research articles

How is academic writing different from other genres?



## General academic writing conventions

## Academic English

- Should be concise, straight to the point
  - John Smith wrote in his 2005 paper “Teacher engagement in the classroom” that students tended to react more positively when teachers engaged with them.
  - Students reacted better when teachers engaged with them (Smith 2005).
- Always uses the third person
  - I will explore the implications of using blogs as a learning tool.
  - The use of blogs as a learning tool will be explored in this paper.
- Avoid being vague
  - Some people consider play as an integral part of socialization for children.
  - Child development specialists consider play as an integral part of children’s socialization.

## Academic English

- English tends to tend toward simple and concise, so don’t use complicated words just because they sound “academic”
- Avoid empty words like “really” or “very” that add nothing but wordcount
  - It is very important to take sample size into account when doing quantitative research.
  - It is important to take sample size into account when doing quantitative research.
- Avoid contractions like can’t, won’t, haven’t, etc.

## Active vs. passive voice

- Active voice places emphasis on whoever is performing the action
  - I failed my class
  - I was failed in my class.
- APA recommends using active voice for clarity
- Academic writing, however, has a few cases in which it is relevant to use passive voice

## When to use passive voice?

- When we want to focus on the action instead of who performed the action
  - Educational researchers classify Down's syndrome as a learning disability.
  - Down's syndrome is classified as a learning disability.
- To add formality or 'objectivity' to a sentence and avoid the first person
  - I designed the study focusing on a qualitative framework.
  - The study was designed focusing on a qualitative framework.
- To address complex or controversial topics
  - Learning technologies help students become more engaged with their work.
  - Learning technologies are considered to help students become more engaged with their work.

## Making claims

- Do not make strong claims
  - Homework does not help students learn better.
  - Homework might not help students learn better.
- Every statement of fact must be backed up by relevant sources

## Facts vs. anecdotes

- Academic texts should be based on facts. You might have relevant experience, but it's not admissible in an academic text
  - One person's experience might not be representative of the whole population
  - People are affected by their biases, so their casual observations might not be rigorous
- The solution is to start from your perceptions and do research to see if scientific evidence backs it up... but it may not!

# Anatomy of the academic writing text

## Parts of an academic text

- Introduction
  - Sets out the main idea and structure of the paper
- Body
  - Defends the main idea through the use of evidence
- Conclusion
  - Wraps up the main arguments used in the paper



# The introduction

## Purpose of the introduction

- Give context on the topic of the paper
  - The XXI Century has brought many changes to the roles of teachers ...
- Introduce the main idea that will be defended throughout the paper
  - However, current teacher training does not prepare teachers to face any of these challenges.
- Outline the main structure of the arguments in the paper
  - Teacher training rarely makes a connection between theory and practice
  - Teachers often feel unsupported when making the transition from studying to teaching

## Putting it all together...

The XXI Century has brought many changes to the roles of teachers. Where formerly teachers merely transmitted knowledge to the students, they are now expected to become mentors in a student-driven learning environment (citation). Teacher training seeks to prepare teachers for these tasks, however the results are sometimes inconclusive (cite). However, current teacher training does not prepare teachers to face any of these challenges. Teacher training rarely makes a connection between theory and practice. Furthermore, teachers often feel unsupported when making the transition from studying to teaching.

## Writing a good introduction

- Brainstorming questions
  - What topics do I find interesting?
  - What do I know about these topics?
  - What do I want to know about these topics?
- Brainstorming techniques
  - Word storm
    - Start with a topic-relevant word and write down all the words that come to mind
  - Mind mapping
    - Organise concepts around a central idea

## For next session...

- Write a 1-paragraph introduction for the topic you selected using the online tool provided
- Deadline: Thursday at midnight
  
- Remember...
  - You don't have to be perfect! That's what drafts are for!

**Week 2**

# Academic Writing Workshop

## Week 2

- Recap: Introductions
- How to revise an introduction
- Peer feedback: Introduction
- The body of an academic text

## Purpose of the introduction

- Give context on the topic of the paper
  - The XXI Century has brought many changes to the roles of teachers ...
- Introduce the main idea that will be defended throughout the paper
  - However, current teacher training does not prepare teachers to face any of these challenges.
- Outline the main structure of the arguments in the paper
  - Teacher training rarely makes a connection between theory and practice
  - Teachers often feel unsupported when making the transition from studying to teaching

An increasing amount of research on the use of automated writing evaluation (AWE) in writing classes suggest growing interest among programme administrators and instructors in AWE's potential to provide formative assessment; that is, assessment that can support learning and teaching. AWE tools, which employ natural-language processing, machine-learning or other computational methods in the analysis of text, can provide both scores on writing quality as well as qualitative feedback on aspects of grammar, mechanics, style, discourse and organisation. While the use of AWE for scoring purposes remains controversial because of its connection to large-scale standardised tests, formative uses as support for writing instruction are viewed more positively (Ware, 2011).

Originally developed for use by native speakers of English, the most commonly used commercial AWE systems, such as Educational Testing Service's Criterion Online Writing Evaluation Service, are increasingly marketed as useful for second language (L2) learners. Given that L2 learners have a greater need for feedback on sentence-level correctness, which AWE systems are more computationally adept at providing compared to feedback on higher level concerns (Weigle, 2013a), a case can be made for the use of AWE as a complement to instructor feedback in L2 writing classrooms. In this role, AWE promises greater autonomy for students while potentially freeing up instructors to devote their feedback efforts to aspects of writing that require human evaluation (Chen & Cheng, 2008; Li, Link, & Hegelheimer, 2015; Warschauer & Grimes, 2008) such as audience awareness and communicative effectiveness.

Arguments in favour of formative applications rest, however, on the assumption that AWE feedback is accurate and useful. While some recent research has investigated AWE accuracy and usefulness separately and in different ways (Dikli & Bley, 2014; Lavolette, Polio, & Kahng, 2015), these studies define accuracy with reference to developer-centric rather than user-centric standards and leave aside the question of how accuracy problems may affect usefulness. The present research seeks to investigate these issues from the perspective of argument-based validation (Kane, 1992, 2012, 2013, 2015), which provides a framework for combining potentially disparate forms of validity evidence in appraising particular interpretations or uses of an assessment. In two studies described here, we appraise evidence regarding two inferences in an interpretation/use argument (IUA) for the use of AWE as a formative assessment tool in college-level ESL writing courses. The first inference, called *evaluation*, focuses on the accuracy of implementation and adherence to the conditions of standardisation in automated feedback generation (Clauser, Kane, & Swanson, 2002). The second inference, *utilisation*, addresses the usefulness of AWE feedback to students in making decisions about revisions. Because 'choices made in developing the scoring algorithm may have an important impact on the strength of this aspect of the argument' (Clauser et al., 2002, p. 424), it is important to investigate accuracy and usefulness in conjunction with each other.

An increasing amount of research on the use of automated writing evaluation (AWE) in writing classes suggest growing interest among programme administrators and instructors in AWE's potential to provide formative assessment; that is, assessment that can support learning and teaching. AWE tools, which employ natural-language processing, machine-learning or other computational methods in the analysis of text, can provide both scores on writing quality as well as qualitative feedback on aspects of grammar, mechanics, style, discourse and organisation. While the use of AWE for scoring purposes remains controversial because of its connection to large-scale standardised tests, formative uses as support for writing instruction are viewed more positively (Ware, 2011).

Originally developed for use by native speakers of English, the most commonly used commercial AWE systems, such as Educational Testing Service's Criterion Online Writing Evaluation Service, are increasingly marketed as useful for second language (L2) learners. Given that L2 learners have a greater need for feedback on sentence-level correctness, which AWE systems are more computationally adept at providing compared to feedback on higher level concerns (Weigle, 2013a), a case can be made for the use of AWE as a complement to instructor feedback in L2 writing classrooms. In this role, AWE promises greater autonomy for students while potentially freeing up instructors to devote their feedback efforts to aspects of writing that require human evaluation (Chen & Cheng, 2008; Li, Link, & Hegelheimer, 2015; Warschauer & Grimes, 2008) such as audience awareness and communicative effectiveness.

Arguments in favour of formative applications rest, however, on the assumption that AWE feedback is accurate and useful. While some recent research has investigated AWE accuracy and usefulness separately and in different ways (Dikli & Bleyle, 2014; Lavolette, Polio, & Kahng, 2015), these studies define accuracy with reference to developer-centric rather than user-centric standards and leave aside the question of how accuracy problems may affect usefulness. The present research seeks to investigate these issues from the perspective of argument-based validation (Kane, 1992, 2012, 2013, 2015), which provides a framework for combining potentially disparate forms of validity evidence in appraising particular interpretations or uses of an assessment. In two studies described here, we appraise evidence regarding two inferences in an interpretation/use argument (IUA) for the use of AWE as a formative assessment tool in college-level ESL writing courses. The first inference, called *evaluation*, focuses on the accuracy of implementation and adherence to the conditions of standardisation in automated feedback generation (Clauser, Kane, & Swanson, 2002). The second inference, *utilisation*, addresses the usefulness of AWE feedback to students in making decisions about revisions. Because 'choices made in developing the scoring algorithm may have an important impact on the strength of this aspect of the argument' (Clauser et al., 2002, p. 424), it is important to investigate accuracy and usefulness in conjunction with each other.

Over the past several years, we have witnessed a growing trend for increased student demand for participation in higher education. While previous reports demonstrated the need for higher education and contrasted this with an argument surrounding the finite capacity to support such growth (OECD, 2013), it was not until 2012 and the hype linked to the massive open online courses (MOOC) that there was intensive public debate about the future role of the university and scalable education models (Kovanović, Joksimović, Gašević, Siemens, & Hatala, 2014). In essence, the rapid advances in technology and its subsequent broad scale adoption provided the necessary infrastructure, and the necessary tipping point for public acceptance of online learning, to enable the delivery of education at such a large scale. While there is much promise amidst the proliferation of MOOCs and online and blended modes of learning more generally, these models also promulgate a new suite of education challenges. For instance, the noted poor attrition rates, and the sheer volume of students enrolled in a MOOC necessitates a more independent study model that is in stark contrast to the more accepted socio-constructivist approaches to learning (Bayne & Ross, 2014).

Despite the challenges of online delivery, the adoption of educational technologies has afforded a new opportunity to gain insight into student learning. As with most IT systems, the student's interactions with their online learning activities are captured and stored. These digital traces (log data) can then be 'mined' and analysed to identify patterns of learning behaviour that can provide insights into education practice. This process has been described as learning analytics. The study of learning analytics has been defined as the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Gašević, 2012). Learning analytics is a bricolage field drawing on research, methods, and techniques from numerous disciplines such as learning sciences, data mining, information visualization, and psychology. This paper reviews the learning analytics research to outline a few of the major topics that the learning analytics field needs to address in order to deliver its oft cited promise for transforming education practice. In so doing, we argue that learning analytics needs to build on and better connect with the existing body of research knowledge about learning and teaching. Specifically, in this paper, we suggest how learning analytics might be better integrated into existing educational research and note the implications for learning analytics research and practice.

Over the past several years, we have witnessed a growing trend for increased student demand for participation in higher education. While previous reports demonstrated the need for higher education and contrasted this with an argument surrounding the finite capacity to support such growth (OECD, 2013), it was not until 2012 and the hype linked to the massive open online courses (MOOC) that there was intensive public debate about the future role of the university and scalable education models (Kovanović, Joksimović, Gašević, Siemens, & Hatala, 2014). In essence, the rapid advances in technology and its subsequent broad scale adoption provided the necessary infrastructure, and the necessary tipping point for public acceptance of online learning, to enable the delivery of education at such a large scale. While there is much promise amidst the proliferation of MOOCs and online and blended modes of learning more generally, these models also promulgate a new suite of education challenges. For instance, the noted poor attrition rates, and the sheer volume of students enrolled in a MOOC necessitates a more independent study model that is in stark contrast to the more accepted socio-constructivist approaches to learning (Bayne & Ross, 2014).

Despite the challenges of online delivery, the adoption of educational technologies has afforded a new opportunity to gain insight into student learning. As with most IT systems, the student's interactions with their online learning activities are captured and stored. These digital traces (log data) can then be 'mined' and analysed to identify patterns of learning behaviour that can provide insights into education practice. This process has been described as learning analytics. The study of learning analytics has been defined as the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Gašević, 2012). Learning analytics is a bricolage field drawing on research, methods, and techniques from numerous disciplines such as learning sciences, data mining, information visualization, and psychology. **This paper reviews the learning analytics research to outline a few of the major topics that the learning analytics field needs to address in order to deliver its oft cited promise for transforming education practice.** In so doing, we argue that learning analytics needs to build on and better connect with the existing body of research knowledge about learning and teaching. Specifically, in this paper, we suggest how learning analytics might be better integrated into existing educational research and note the implications for learning analytics research and practice.

## Questions to ask when revising an introduction

- What is the main idea in this introduction?
  - Is it clear?
  - Is it expressed as a statement of fact?
- Does the context help understand the main idea?
  - Are all sentences in the context related to the main idea?
  - Is the context well-sourced?
  - Are ideas repeated unnecessarily?
- Is it clear how the main idea is going to be defended?
  - What are the main arguments in this essay?

## Workshop Activity

- Get together in pairs
- Exchange your essay with another person and give them feedback
  - Try to identify and underline the main idea
  - Use the revision questions to try and give meaningful feedback

## The body of an essay



## Purpose of the body of an essay

- Support the main idea proposed in the introduction through
  - Analysis
  - Arguments
  - Citations

## Basic structure of an argument

- Main idea of the paper
  - It is necessary to revise the aims of integration and bilingual education for minorities, considering the ethical implications that underlie these educational efforts.
- Main point of the argument
  - One of the questions that have been pursued by modern educational philosophy is that of the right to educate.
- Supporting statements and evidence
  - While some philosophers of education have outright declared that no person or institution has the right to educate (Tamir, qtd. in Rosenow, 1993), others have dismissed this viewpoint as ignoring the fact that the self develops in response to their environment, and if educators do not guide this development, other individuals or institutions will certainly try to do so (Wringe, 2013).

## Putting it all together...

- One of the questions that have been pursued by modern educational philosophy is that of the right to educate. While some philosophers of education have outright declared that no person or institution has the right to educate (Tamir, qtd. in Rosenow, 1993), others have dismissed this viewpoint as ignoring the fact that the self develops in response to their environment, and if educators do not guide this development, other individuals or institutions will certainly try to do so (Wringe, 2013).

One of the questions that have been pursued by modern educational philosophy is that of the right to educate. While some philosophers of education have outright declared that no person or institution has the right to educate (Tamir, qtd. in Rosenow, 1993), others have dismissed this viewpoint as ignoring the fact that the self develops in response to their environment, and if educators do not guide this development, other individuals or institutions will certainly try to do so (Wringe, 2013). Wringe's point is especially interesting because it touches upon the importance of the environment in the formation of the 'self' as he questions the right of educators to transform said self by enriching and extending their education and experience (2013). However, he speaks of the transformation as a neutral process which simply opens up the range of experiences for students to decide which ones are relevant to their own interests and pursuits of a fulfilling life. Such an idealistic view of the process ignores that education cannot be neutral and is undermined by the historical realities of group relationships between dominant groups and involuntary minorities. By their very definition, involuntary minorities are created against their will and thus their struggle is something that has been imposed on them, unlike voluntary immigrants who see this struggle as a transitory adaptation phase that will lead to better opportunities than what they had "back home" (Ogbu & Simons, 1998). Namely, involuntary minorities have historically suffered through systematic discrimination and a lack of rewards in employment and wages for educational accomplishments, as well as in receiving a lesser quality of education than the dominant groups (Hickling-Hudson & Ahlquist, 2003; Nasser & Nasser, 2008; Ogbu & Simons, 1998). Children whose languages are not supported in the educational system have a more fragmented or muddled idea of their own identity (Martin, 2012), which creates long-lasting problems for identity formation, group self-esteem, and displacement, denying them the opportunities to build future expectations for themselves (Ahonen, 2001).

# Citations

## The purpose of a citation

- Knowledge does not exist in a vacuum, we build our knowledge on top of what others have researched
- Common sense and intuition might fail, only research can give us (some) certainty
- Using a wide variety of sources shows that you have engaged with the subject and done your research
- To avoid plagiarism

## When do I need to cite?

- Whenever you're presenting:
  - A fact or a claim
  - A definition of an idea or concept
- When you want to analyse or dispute a claim made by someone else
- Using primary data or previous research to support your claims
- Talking about the state of art in a topic

## Paraphrasing vs. direct quotes

- Paraphrasing is preferred because it allows your ideas and interpretations to shine through
- Paraphrasing can be used to:
  - Summarize long and complex ideas in a short space
  - Get rid of information that might not be relevant
  - Showcase the author's understanding of a topic
- Direct quotes can be used to:
  - Back up delicate or controversial claims
  - Communicate an idea that was expressed perfectly by someone else
  - Provide a passage that will be analysed
  - Provide definitions

## Writing a good essay...

- Try to make an outline of your arguments to make sure they flow in a logical way
- Each paragraph must convey a complete idea, although one paragraph may expand on another's idea

## For next session....

- Write a 2 or 3-paragraph body for the essay to support the introduction you turned in this week
- Deadline: Thursday at midnight
- Remember:
  - Give yourself permission to write badly, but also remember to revise

**Week 3**

# Academic Writing Workshop

## Week 3

- Recap: The body of an essay
- How to revise the body of an essay
- Peer feedback: the body of an essay
- The conclusion of an academic text

# Purpose of the body of an essay

- Support the main idea proposed in the introduction through
  - Analysis
  - Arguments
  - Citations

# Revising the body of an essay

Establishes important definitions: right to educate, involuntary immigrants, etc.

The educational curriculum excludes minorities

One form of exclusion is through the use of language

Allowing children to explore their language is a challenge

Part of the challenge is that linguistic exclusion is institutionalized in the education system

A good education system should be plural

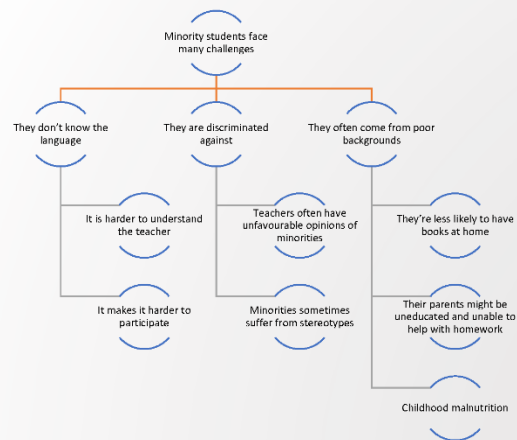
A way to fix this is rethinking student-teacher relationships

Multicultural education should be dialogical and allow students a voice

It is important to break the “silence” of minority groups

The community should be taking into account

## Organising ideas in an essay



## Revision questions

- Do the paragraphs follow logically from one another to build an argument?
  - Does each paragraph have a complete argument?
- Are all the important claims well-sourced?
  - Am I using too many direct quotes? Could some of them be paraphrased?
- Are all the arguments relevant to the main idea of the paper?
- Is my writing clear and concise?
  - Do I use short sentences?
  - Am I using any filler words that I could remove?



## Common mistakes: not having a voice

- However, evaluation now becomes different. Rosales mentions that “the evaluation now has the goal of providing the most information to improve this process, readjusting the objectives, critically revising plans, programs, methods and resources, facilitating the most help and orientation for students” (2004).

## Commonly used filler words and phrases

- Very
- That
- Just
- Only
- Really
- Absolutely
- Basically
- Actually
- Sort of
- Kind of
- A little
- Certainly
- In order to
- Perhaps
- Completely
- Definitely
- The fact that

For more information and examples, visit: <https://wordvice.com/avoid-fillers-powerful-writing/>

## Commonly used filler words

- It is very important that we understand the main causes of discrimination.
  - It is important that we understand the main causes of discrimination.
  - It is often thought that students perform better when they're engaged.
  - It is often thought students perform better when they're engaged.
- Protip: try removing the filler words from your sentence and, if the meaning doesn't change, then you did not really need them.

## Commonly used filler words

- It is very important that we understand the main causes of discrimination.
  - It is important that we understand the main causes of discrimination.
  - It is often thought that students perform better when they're engaged.
  - It is often thought students perform better when they're engaged.
- Protip: try removing ~~the~~ filler words from your sentence and, if the meaning doesn't change, ~~then~~ you did not ~~really~~ need them.

## Workshop activity

- Get together in pairs
- Exchange your essay with another person and give them feedback
  - Use the revision questions as a guide

## The conclusion

## Structure of a conclusion

- If the main idea of the paper is...
  - It is necessary to revise the aims of integration and bilingual education for minorities, considering the ethical implications that underlie these educational efforts.
- Summarize main points presented in the academic essay
  - Until we are able to appreciate involuntary minorities, not as an objectified, silent 'other' but as an active subject, a 'self' with different but valid experiences and perceptions, any attempts at an educational program that can serve their interests would be, at best, incomplete. Failing to acknowledge the creative power of minority languages and cultures by excluding them from the curriculum leaves minority students with an incomplete, contradictory identity that places them at a disadvantage during their interactions with dominant groups.
- Reflect on the consequence, implications or significance of the ideas presented
  - An ethical approach to bilingual education would then attempt to present the dominant and minority cultures as standing on equal ground with regards to the potential creativity of expression in the student. Only then can education empower minority individuals to integrate to mainstream society on their own terms, instead of taking their traditionally assigned role of subordinates.

## Putting it all together...

- Until we are able to appreciate involuntary minorities, not as an objectified, silent 'other' but as an active subject, a 'self' with different but valid experiences and perceptions, any attempts at an educational program that can serve their interests would be, at best, incomplete. Failing to acknowledge the creative power of minority languages and cultures by excluding them from the curriculum leaves minority students with an incomplete, contradictory identity that places them at a disadvantage during their interactions with dominant groups. An ethical approach to bilingual education would then attempt to present the dominant and minority cultures as standing on equal ground with regards to the potential creativity of expression in the student. Only then can education empower minority individuals to integrate to mainstream society on their own terms, instead of taking their traditionally assigned role of subordinates.

## Common mistakes when writing conclusions

- You should not introduce any new information or concepts in the conclusion
- Don't just repeat everything you said in the introduction.
  - In the introduction you **mention** what the arguments are going to be
  - In the conclusion you **paraphrase** what the arguments were
- Open with phrases like "in conclusion" or "to conclude" – they are considered redundant

## For next session....

- Write a 1-paragraph conclusion for the essay to support the essay you turned in this week
- Deadline: Thursday at midnight

## Week 4

# Academic writing workshop

## Week 4

- Recap: The conclusion
- How to revise the conclusion
- Peer feedback: the conclusion
- Writing a strong abstract
- Citations and Reference lists

## Structure of a conclusion

- Summarize main points presented in the academic essay
  - Until we are able to appreciate involuntary minorities, not as an objectified, silent 'other' but as an active subject, a 'self' with different but valid experiences and perceptions, any attempts at an educational program that can serve their interests would be, at best, incomplete. Failing to acknowledge the creative power of minority languages and cultures by excluding them from the curriculum leaves minority students with an incomplete, contradictory identity that places them at a disadvantage during their interactions with dominant groups.
- Reflect on the consequence, implications or significance of the ideas presented
  - An ethical approach to bilingual education would then attempt to present the dominant and minority cultures as standing on equal ground with regards to the potential creativity of expression in the student. Only then can education empower minority individuals to integrate to mainstream society on their own terms, instead of taking their traditionally assigned role of subordinates.

## Revising a conclusion

- Have I wrapped up all the main arguments in my essay?
- Am I making it clear what the contribution of my text was?
  - ... by mentioning future pathways for the topic?
  - ... by explaining the implications of the ideas presented?
  - ... by reflecting on the significance of my arguments?
- Is my conclusion directly related to the main idea of my paper?
  - .... Or did I lose track of what I was saying at some point?

This study has focused on how two individual students engaged with teacher and AWE feedback on their writing during a university semester and offers some insights into the complex processes of learner engagement with writing tasks. Such engagement is a dynamic process whereby behavioral, affective, and cognitive responses are simultaneously at work. Both students displayed behavioral engagement with two types of feedback on their writing although the depth of this varied. The highly engaged learner tended to participate actively and reflect more deeply with both teacher and AWE feedback on her writing, while the moderately engaged learner was less motivated and showed less willingness to use the feedback he received from both sources. More highly engaged learners thus tend to spend more time working with feedback, show more positive attitudes toward it, and employ more revising strategies, demonstrating that behavioral, affective, and cognitive engagement dynamically interact with each other. There are only two participants in this study and so caution should be exercised when generalizing the findings to a wider student population. While the different styles of their engagement led us to see how each responded to teacher and AWE feedback differently, for example, both demonstrated some engagement with the feedback they received. It is uncertain how other, perhaps less-engaged or disaffected students might respond. Thus, future research is needed which explores the engagement patterns of a more varied range of learners.

We hope to have shown that student engagement with feedback is a key construct in L2 writing research that merits more attention. Both the source of the feedback and individual learner factors can shape the process so that language proficiency and knowledge of learning strategies can play a pivotal role, not only in helping students to understand and interpret feedback, but also in evaluating and monitoring the revision process. Affective factors such as reactions to red ink or a sympathetic tone, and student learning beliefs also have a profound influence on how they engage with the two types of feedback. These results suggest that teachers can play a more active role in helping second language students to develop positive learning beliefs and attitudes to writing and revising in English. In addition to adopting a reassuring tone and seeking to develop students' general proficiency, explicit instruction on the use of revision operations and cognitive strategies can be crucial in improving student engagement and therefore enhancing their writing skills.

Student motivation and self-regulation both have important roles to play in college student learning and achievement. The four generalizations offered in this chapter serve as good first principles for understanding how student motivation can facilitate or constrain self-regulated learning and achievement in the college classroom. Students who feel efficacious about their ability to learn and to do the work are more likely to be engaged and to do better. Likewise, students who are focused on learning, mastery, and self-improvement are more likely to be involved in learning and perform better. Finally, a third facilitating factor of engagement and achievement is task value with students who think the material is interesting, important, and useful more likely to be engaged and learning. A constraining factor on engagement and learning is test anxiety with higher levels of test anxiety interfering or impeding cognitive engagement, learning, and achievement.

These generalizations seem to apply to all groups of students, but there is a clear need for more research on how different personal characteristics may moderate or delimit how these four principles can be generalized. Finally, classroom context factors can certainly influence student motivation and cognition. Moreover, the classroom context factors discussed here are inherently open to manipulation and change, offering hope to faculty members who want to make improvements in their classrooms and in the nature of their instruction to facilitate student motivation and learning. Much research remains to be done, but the general model offered here should provide a conceptual framework for future research as well as practice.



# Abstracts

## What is an abstract?

- It is a short summary of your paper, typically between 200-500 words
- Contains all the important parts of the academic text
  - The purpose of the research (main idea)
  - The methods used
  - Brief overview of the findings and conclusions

The analysis of data collected from the interaction of users with educational and information technology has attracted much attention as a promising approach for advancing our understanding of the learning process. This promise motivated the emergence of the new research field, learning analytics, and its closely related discipline, educational data mining. This paper first introduces the field of learning analytics and outlines the lessons learned from well-known case studies in the research literature. The paper then identifies the critical topics that require immediate research attention for learning analytics to make a sustainable impact on the research and practice of learning and teaching. The paper concludes by discussing a growing set of issues that if unaddressed, could impede the future maturation of the field. The paper stresses that learning analytics are about learning. As such, the computational aspects of learning analytics must be well integrated within the existing educational research.

Research on feedback in second language writing has grown enormously in the past 20 years and has expanded to include studies comparing human raters and automated writing evaluation (AWE) programmes. However, we know little about the ways students engage with these different sources of feedback or their relative impact on writing over time. This naturalistic case study addresses this gap, looking at how two Chinese students of English engage with both teacher and AWE feedback on their writing over a 16-week semester. Drawing on student texts, teacher feedback, AWE feedback, and student interviews, we identify the strengths and weaknesses of both types of feedback and show how engagement is a crucial mediating variable in the use students make of feedback and the impact it has on their writing development. We argue that engagement is a key factor in the success of formative assessment in teaching contexts where multiple drafting is employed. Our results show that different sources of formative assessment have great potential in facilitating student involvement in writing tasks and we highlight some of these pedagogical implications for promoting student engagement with teacher and AWE feedback.

## Things to avoid

- Too much background information
- References to other literature
- References to images, tables, figures, etc.
- Abbreviations that are not explained

**Week 5**

# Academic writing workshop

## Week 5

- Recap: the abstract
- How to revise the abstract
- Peer feedback: the abstract
- Citation and referencing
- General writing tips and tricks

## What is an abstract?

- It is a short summary of your paper, typically between 200-500 words
- Contains all the important parts of the academic text
  - The purpose of the research (main idea)
  - The methods used
  - Brief overview of the findings and conclusions

The analysis of data collected from the interaction of users with educational and information technology has attracted much attention as a promising approach for advancing our understanding of the learning process. This promise motivated the emergence of the new research field, learning analytics, and its closely related discipline, educational data mining. This paper first introduces the field of learning analytics and outlines the lessons learned from well-known case studies in the research literature. The paper then identifies the critical topics that require immediate research attention for learning analytics to make a sustainable impact on the research and practice of learning and teaching. The paper concludes by discussing a growing set of issues that if unaddressed, could impede the future maturation of the field. The paper stresses that learning analytics are about learning. As such, the computational aspects of learning analytics must be well integrated within the existing educational research.

## Questions to ask when revising an abstract

- Have I included the three main parts?
- If it were a published paper, does it include the information I need to decide whether to read it?
- Does it make my paper look interesting?

## Citing and referencing (APA format)

## Types of citations - paraphrasing

- Author and year in parenthesis
  - The way we use languages and the ideas we have about them are socially situated and are inseparable from questions of power within societies (Blackledge & Pavlenko, 2001).
- Only year in parenthesis
  - Edwards (2010) alongside others suggests that the endangerment of these languages is brought about by social, economic, political and cultural reasons.
- Ideas found in more than one author
  - Many authors have studied how implicit and explicit policies of exclusion in education lead to a negative impact in student achievement and sense of self-worth (Freeman, 2006; Holmes, 2004; Nasser & Nasser, 2008).
- Author A quoted in paper B:
  - Some philosophers of education have outright declared that no person or institution has the right to educate (Tamir, qtd. in Rosenow, 1993)

## Types of citations – direct quote

- All information after the quote
  - As it is only through the creation of a “confident, Gaelic-speaking identity” that language revitalisation will succeed or fail (Oliver, 2006, p. 160)
- Split author and year/page number
  - Under these circumstances, Bizzell’s rhetoric of “liberatory teachers to an oppressive system” (1991, p. 58) strikes as condescending and patronizing
- Split author/year and page number
  - However, it has been suggested by advocates of language revitalisation efforts like Hornberger (2009) that it “tilts the use of the indigenous language away from discrimination and oppression and toward emancipation, self-determination and empowerment” (p. 289).

## When should I reference?

- Things that are not common knowledge (depends on your field)
- Original research made by someone else
- Describing claims and opinions made by others
- Some tips:
  - If you can picture your teacher asking: “says who?” then it should probably be referenced
  - When in doubt, reference everything you did not think up yourself
  - When in super doubt, find someone who agrees with you and cite them

## Research and reading advice

- Always make notes on EVERYTHING you read and find useful
  - OneNote
  - Index Cards
  - Research notebook
  - EverNote
  - Microsoft Word
  - Microsoft Excel



## Research and reading advice

- Useful things to write in the notes:
  - Reference information
  - Article keywords
  - One or two main ideas
- Other things you might include:
  - Phrases you might be interested in citing
  - A brief summary of the entire article/paper

## Reference lists

- There are several programs that can help you build reference lists and manage your sources:
  - Microsoft Word has an in-built function (in the References tab)
  - Mendeley (with Word plug-in)
  - Endnote (there is a discount for UoE students)

## Citing: Further resources

- Purdue's Online Writing Center:
  - [https://owl.purdue.edu/owl/purdue\\_owl.html](https://owl.purdue.edu/owl/purdue_owl.html)
- MIT: What is common knowledge?
  - <https://integrity.mit.edu/handbook/citing-your-sources/what-common-knowledge>

## Tips and tricks: subject-verb agreement

- Nouns are made plural by adding “-s” at the end, but verbs are made plural by taking away the “-s”
  - Teachers feel less stress when they are not being micromanaged.
  - A teacher feels less stress when they are not being micromanaged.
  - The teacher feels less stress when she is not being micromanaged.
    - Referring to a particular teacher

## Tips and tricks: subject-verb agreement

- Nouns are made plural by adding “-s” at the end, but verbs are made plural by taking away the “-s”
  - **Teachers feel** less stress when they are not being micromanaged.
    - Referring to all teachers
  - **A teacher feels** less stress when they are not being micromanaged.
    - Referring to a hypothetical teacher
  - **The teacher feels** less stress when she is not being micromanaged.
    - Referring to a particular teacher

## Tips and tricks: incomplete sentences

- Writing a dependent clause that is linked to nothing:
  - **While it is true that students tend to pay more attention when teachers try to make their classes more engaging.**
  - **Because teachers that work in public schools tend to take on more responsibilities and be less well paid than their private school counterparts.**

## Tips and tricks: incomplete sentences

- Writing a dependent clause that is linked to nothing:
  - While it is true that students tend to pay more attention when teachers try to make their classes more engaging, it is often the case that teachers don't have enough training to create truly engaging content.
  - Because teachers that work in public schools tend to take on more responsibilities and be less well paid than their private school counterparts, their levels of stress tend to be higher.

## Tips and tricks: incomplete sentences

- Try to avoid starting sentences with connecting words like...
  - Because
  - In order to
  - Especially
  - From
  - Although
  - Whereas
  - Since
  - Etc.

## Further information and resources

- General writing advice:
  - [https://owl.purdue.edu/owl/general\\_writing/index.html](https://owl.purdue.edu/owl/general_writing/index.html)
- Subject-verb agreement:
  - [https://owl.purdue.edu/owl/general\\_writing/grammar/subject\\_verb\\_agreement.html](https://owl.purdue.edu/owl/general_writing/grammar/subject_verb_agreement.html)
- How to use articles:
  - [https://owl.purdue.edu/owl/general\\_writing/grammar/using\\_articles.html](https://owl.purdue.edu/owl/general_writing/grammar/using_articles.html)
- 10 common mistakes in student writing
  - <https://latrobe.libguides.com/language/common-mistakes>
- 10 common mistakes in ESL writing
  - [https://www.scribendi.com/advice/the\\_10\\_most\\_common\\_esl\\_mistakes.en.html](https://www.scribendi.com/advice/the_10_most_common_esl_mistakes.en.html)

## Revising tips:

- Change your font and print out your work to revise it. The difference in presentation will help you spot things you'd otherwise ignore
- Read your own work out loud. It forces you to think about sentence length, word choice and flow.
- If you're not sure how to revise something, make a note for yourself and come back to it later.

Always remember...

- Give yourself permission to write poorly, but revise thoroughly
- Don't overwork yourself. Sometimes the best course of action is to take a break and let your brain rest
- If you can't figure something out, do something that keeps your body busy but doesn't require a lot of brainpower like jogging or knitting

Good luck!

## **Appendix 2 – Info sheet and consent form for Study 1**

### **Information sheet and consent form**

#### **The use of Automated Writing Evaluation programs in Higher Education**

**Researcher: Ana Isabel Hibert Santana**

**Moray House School of Education**

**This Informed Consent Form has two parts:**

- **Information sheet (to share information about the study with you)**
- **Consent form (for you to sign if you agree to take part in this study)**

## **Part I: Information sheet**

I am Ana Hibert, a PhD student at the Moray House School of Education, and I am researching ways in which Automated Writing Evaluation (AWE) programs can help students become more self-sufficient English writers and improve the quality of their texts.

AWE programs are designed to give students automated and immediate feedback on their writing through the use of natural language processing techniques. These programs can give students feedback on not only grammar and spelling, but style, organization and sometimes even content.

AWE programs are being increasingly used in higher education to help English as a second language (ESL) students improve their writing skills and help them succeed in English language programs.

### **Purpose of the research**

This research focuses on studying how ESL students engage with AWE programs and use them to improve their written texts, revision strategies and writing skills.

### **Participant selection**

You are being invited to participate in this study because you are an international ESL student enrolled at a postgraduate program in the University of Edinburgh.

### **Procedure**

This study will begin with a brief questionnaire about your study habits and writing strategies. For the main part of this pilot, you will be asked to attend four 30-minute sessions, during which you will be given a chance to revise a text of your choice (it can be your dissertation or any other coursework, the only requirement is that the text be in English). During each of those sessions, you will be asked to participate in a think-aloud protocol, which means that you will be asked to talk



about what you're doing as you use the AWE program to revise your text. Your voice and activity on the screen will be recorded during each session for further analysis. At the end of the study, you will be asked to participate in a brief interview about your experiences using the program.

### **Confidentiality**

I will not be sharing information about you to anyone. While your voice will be recorded in both the interview and think-aloud protocols, your face will not be.

The data I collect from you will be anonymized and kept private: all the information will have a number on it instead of your name, and only I will know what that number is, the key will be stored in an encrypted file. The videos will be securely stored in an encrypted device, and all transcripts will have any identifying information about you, your writing or your program scrubbed from them.

Furthermore, I will keep no files belonging to you. The computer session used during the recordings will be wiped clean of all files, history and information, so nobody else will have access to your work, identity or any other information related to you.

### **Sharing the results**

Part of the transcripts may be shared with my PhD supervisors, but as stated above, all identifying information will be scrubbed from them.

The results will be used as part of my doctoral thesis, and I may publish a paper based on this research. If the research is published, anonymity will be respected as stated in the previous section.

If you have any questions about this project, either now or in the future, please contact me at

**Part II: Certificate of Consent**

I have read the foregoing information, or it has been read to me. I have had the opportunity to ask questions about it and any questions I have been asked have been answered to my satisfaction. I consent voluntarily to be a participant in this study

**Print Name of Participant** \_\_\_\_\_

**Signature of Participant** \_\_\_\_\_

**Date** \_\_\_\_\_

Day/month/year

**Statement by the researcher/person taking consent**

I have accurately read out the information sheet to the potential participant, and to the best of my ability made sure that the participant understands what the think-aloud protocols and interviews entail.

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I

confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

**A copy of this ICF has been provided to the participant.**

**Print Name of Researcher** \_\_\_\_\_

**Signature of Researcher** \_\_\_\_\_

**Date** \_\_\_\_\_

Day/month/year

## Appendix 3 – Code for cleaning up the text database for Study 1

14/07/2021

Data cleanup

In [1]:

```
import pandas as pd
import matplotlib
import pymysql as mysql
from sqlalchemy import create_engine
from bs4 import BeautifulSoup
```

In [2]:

```
## Initialize connection, Apache and MySql need to be running
db_connection_str = 'mysql+pymysql://root:hamsterpasta@localhost/wwwgrammaraidedu_datacollection'
db_connection = create_engine(db_connection_str)
```

### User data

In [3]:

```
user_data = pd.read_sql('SELECT * FROM user', con=db_connection)
user_data.head()
```

Out[3]:

	id	email	password	gender
0	1	prueba@probando.com	2b12\$HEg3TCox7ldgfa2oMnccmOpHeJD/SYUBkpnNpjX...	
1	2	alexia.revueltas@gmail.com	2b12\$4N.aDbqTfbqY2vt8wr4pPub6rLmOUEMAG4Cub2r...	
2	3	s1844319@ed.ac.uk	2b12\$W.1zGEllU.8ITyzThIG90uDKhU9HxxplrVYFzYf...	
3	4	l.li-65@sms.ed.ac.uk	2b12\$ANo/NP9Gtu9VS3jbDWkQJ.cyoZDSXCRkejcc.Mz...	
4	5	874297075@qq.com	2b12\$j3nZJeUsU9WfyGPqgWEpHeT.LG0fEkqbC4NsPu4...	

In [4]:

```
## Remove first 15 users because there's not data on them and/or they are tests
clean_users = user_data[user_data.id > 15]
clean_users.head()
```

Out[4]:

	id	email	password	gender
15	16	1929121@ed.ac.uk	2b12\$pMbx/i3skpcz4uCsFYmcM.tnXl.1DybyqLhJGVVA...	2
16	17	594546738@qq.com	2b12\$A3zfSrQm7lxYHJeXHcGaJOfnYdeadX751RPScym...	2
17	18	xueqin0809@gmail.com	2b12\$EEEEeHL/1Bpxq69HKzgmYe/WYK25jMVMwBfsxRW...	2
18	19	s1902567@ed.ac.uk	2b12\$2aCuTiiLKlx2yZWB6OEZtuwF9KoDihkULdb8esA...	2
19	20	e328308294@163.com	2b12\$TniWpiFzHykxom28AVo23uFI3KZkoH7CWcyp/Uv...	2

In [5]:

```
gender = pd.crosstab(index=clean_users.gender, columns="count")
gender
```

Out[5]:

gender	col_0	count
1	1	6
2	2	64

In [6]:

```
l1 = pd.crosstab(index=clean_users.language, columns="count")
l1
```

Out[6]:

language	col_0	count
CHINESE	1	1
Chinese	49	49
Chinese Mandarin	2	2
German	1	1
Greek	1	1
Madarian	1	1
Mandarin	4	4
Spanish	1	1
chinese	7	7
mandarin	3	3

In [7]:

```
years_english = clean_users["english"].mean()
years_english
```

Out[7]:

11.857142857142858

## Import texts

In [8]:

```
texts = pd.read_sql('SELECT * FROM texts', con=db_connection)
texts.head()
```

Out[8]:

	id	title	date_posted	content	user_id
0	1	some new text	2019-05-02 12:09:52	<p>Some text! More text. Updated text. New upd...	1
1	2	Anita lava la tina	2019-05-02 12:20:57	<p>Como sé yo, Anita lava la tina, pero no est...	2
2	3	Seriousnessess	2019-05-02 12:23:56	<p><strong>I'm wondering if this works a bit&n...	2
3	4	Testing	2019-05-02 12:37:16	<p>One day Gregoir Samsa found himself transfo...	1
4	5	email	2019-05-03 14:28:29	<p>Write your text h</p>\r\n<p>Dear Jasmine, <...	3

Cleaning out the texts from users 1-13

In [9]:

```
clean_texts = texts.iloc[14:197]
len(clean_texts)
```

Out[9]:

183

## Import text versions

In [10]:

```
text_versions = pd.read_sql('SELECT * FROM text_versions', con=db_connection)
text_versions.head()
```

Out[10]:

	id	content	fecha	user_id	text_id
0	1	<p>some text!</p>	2019-05-02 12:09:52	1	1
1	2	<p>Como sé yo, Anita lava la tina, pero no est...	2019-05-02 12:20:57	2	2
2	3	<p><strong>I'm wondering if this works a bit b...	2019-05-02 12:23:56	2	3
3	4	<p>This is a test text. Alliteration!</p>	2019-05-02 12:37:16	1	4
4	5	<p>This is a test text. Alliteration!</p>	2019-05-02 12:37:23	1	4

Cleaning away all the test text and texts by people with no demographic information (users 1 - 13)

In [11]:

```
clean_versions = text_versions[text_versions.user_id > 13]
len(clean_versions)
```

Out[11]:

401

Seeing how many texts have more than one version and are therefore useful for the analysis

In [12]:

```
frequency = pd.crosstab(index=clean_versions.text_id, columns="count")
useless = frequency[frequency["count"] == 1].index
useful_index = frequency.drop(useless)
len(useful_index)
```

Out[12]:

75

Frequency count for how many versions there are of texts

In [13]:

```
useful_index["count"].value_counts()
```

Out[13]:

```
2    22
3    17
4    12
6     7
7     6
5     4
8     2
15    1
14    1
13    1
11    1
9     1
```

Name: count, dtype: int64

Get the texts that are useful

In [14]:

```
# Turn useful_index to a list so it can be extracted from the dataset
useful_ids = useful_index.index.tolist()

# use the list to extract only the texts that have more than one version
clean_dataset = clean_versions[clean_versions.text_id.isin(useful_ids)]
clean_dataset.head()
```

Out[14]:

	id	content	fecha	user_id	text_id
59	60	<p><strong>What do </strong><strong>Supportive...	2019-10-20 15:10:54	22	28
60	61	<p><strong>What do&nbsp;</strong><strong>Suppo...	2019-10-20 15:19:22	22	28
71	72	<p>Educational assessment is considered as an ...	2019-10-23 22:44:28	55	38
72	73	<p>We consider educational assessment as an es...	2019-10-23 23:10:19	55	38
73	74	<p>We consider educational assessment as an es...	2019-10-23 23:11:05	55	38

In [15]:

```
len(clean_dataset)
```

Out[15]:

325

## Clean texts of html



In [16]:

```
clean_dataset
```

Out[16]:

	id	content	fecha	user_id	text_id
59	60	<p><strong>What do </strong><strong>Supportive...	2019-10-20 15:10:54	22	28
60	61	<p><strong>What do&nbsp;</strong><strong>Suppo...	2019-10-20 15:19:22	22	28
71	72	<p>Educational assessment is considered as an ...	2019-10-23 22:44:28	55	38
72	73	<p>We consider educational assessment as an es...	2019-10-23 23:10:19	55	38
73	74	<p>We consider educational assessment as an es...	2019-10-23 23:11:05	55	38
...	...	...	...	...	...
505	506	<p>Research studies have identified multiple t...	2020-08-15 01:35:58	63	197
506	507	<p>However, resilience is not permanent (Reyes...	2020-08-15 01:42:31	63	197
507	508	<p>By contrast, protective factors that associ...	2020-08-15 01:45:49	63	197
508	509	<p>It is crucial for individuals to build resi...	2020-08-15 01:48:15	63	197
509	510	<p>In addition, a large number of studies have...	2020-08-15 01:50:38	63	197

325 rows × 5 columns

In [ ]:

```
for index, row in clean_dataset.iterrows():
    print(row['content'])
```

In [18]:

```
small_subset = clean_dataset.head()
small_subset
```

Out[18]:

	id	content	fecha	user_id	text_id
59	60	<p><strong>What do </strong><strong>Supportive...	2019-10-20 15:10:54	22	28
60	61	<p><strong>What do&nbsp;</strong><strong>Suppo...	2019-10-20 15:19:22	22	28
71	72	<p>Educational assessment is considered as an ...	2019-10-23 22:44:28	55	38
72	73	<p>We consider educational assessment as an es...	2019-10-23 23:10:19	55	38
73	74	<p>We consider educational assessment as an es...	2019-10-23 23:11:05	55	38

In [19]:

```
# cleans the html out of the content
for i, row in clean_dataset.iterrows():
    soup = BeautifulSoup(row['content'], 'lxml')
    clean_dataset.at[i, 'content'] = soup.text
```

In [20]:

```
clean_dataset['wordcount'] = clean_dataset.content.str.count(' ') + 1
clean_dataset.head()
```

<ipython-input-20-f2a01eccd294>:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
clean_dataset['wordcount'] = clean_dataset.content.str.count(' ') + 1
```

Out[20]:

	id	content	fecha	user_id	text_id	wordcount
59	60	What do Supportive Design Theory and Attention...	2019-10-20 15:10:54	22	28	351
60	61	What do Supportive Design Theory and Attention...	2019-10-20 15:19:22	22	28	340
71	72	Educational assessment is considered as an essen...	2019-10-23 22:44:28	55	38	104
72	73	We consider educational assessment as an essen...	2019-10-23 23:10:19	55	38	102
73	74	We consider educational assessment as an essen...	2019-10-23 23:11:05	55	38	100

In [21]:

```
## Remove texts with less than 100 words
final_dataset = clean_dataset[clean_dataset.wordcount >= 100]
len(final_dataset)
```

Out[21]:

298

In [22]:

```
## Check how many text have 100 words or more
frequency = pd.crosstab(index=final_dataset.text_id, columns="count")
tooshort = frequency[frequency["count"] == 1].index
good_index = frequency.drop(tooshort)
len(good_index)
```

Out[22]:

70

In [23]:

```
## write every cell in the content column into an individual txt file
##file = 'D:\OneDrive - University of Edinburgh\dissertation\data collection sept2019-m
ay2020\data\individual_texts\{}.txt'
##for i, row in final_dataset.iterrows():
    ## with open(file.format(str(row['text_id']) + '-' + str(row['id'])), 'w') as f:
        ## f.write(str(row['content']))
```

In [24]:

```
final_dataset["wordcount"].max()
```

Out[24]:

4539

In [25]:

```
final_dataset.to_csv(r'final_dataset.csv', index=False)
```

## Import user actions

In [26]:

```
user_actions = pd.read_sql('SELECT * FROM user_actions', con=db_connection)
user_actions.head()
```

Out[26]:

	id	fecha	action	user_id	text_id
0	2	2019-05-02 12:09:37	1	1	NaN
1	3	2019-05-02 12:09:52	4	1	NaN
2	4	2019-05-02 12:19:31	1	2	NaN
3	5	2019-05-02 12:20:57	4	2	NaN
4	6	2019-05-02 12:23:10	1	2	NaN

Get rid of all test users and whatnot

In [27]:

```
clean_actions = user_actions[user_actions.user_id > 13]
clean_actions.head()
```

Out[27]:

	id	fecha	action	user_id	text_id
74	76	2019-06-13 08:18:51	1	14	NaN
130	134	2019-10-17 10:34:37	1	16	NaN
131	135	2019-10-17 11:23:17	1	17	NaN
132	136	2019-10-17 13:24:50	1	18	NaN
133	137	2019-10-17 13:25:42	1	18	NaN

In [28]:

```
clean_actions.to_csv(r'all_user_actions.csv', index=False)
```

Get only new, summary and save actions

In [29]:

```
subset_text_actions = clean_actions[clean_actions['action'].isin(['2', '3', '4', '6'])
]
subset_text_actions.head()
```

Out[29]:

	id	fecha	action	user_id	text_id
145	149	2019-10-20 11:21:46	4	30	NaN
149	153	2019-10-20 15:10:54	4	22	NaN
150	154	2019-10-20 15:15:12	6	22	28.0
151	155	2019-10-20 15:19:22	3	22	28.0
152	156	2019-10-20 19:02:40	4	26	NaN

Check which users are in the final dataset

In [30]:

```
test_versions = text_versions[text_versions.user_id > 13]
len(test_versions)
```

Out[30]:

401

In [31]:

```
# Get a list of all the texts that only have 1 version
frequency = pd.crosstab(index=test_versions.text_id, columns="count")
useless = frequency[frequency["count"] > 1].index
useful = frequency.drop(useless)

# Turn useful_index to a list so it can be extracted from the dataset
useful_userid = useful.index.tolist()

# use the list to extract only the texts that have more than one version
useful_actions = test_versions[test_versions.text_id.isin(useful_userid)]
useful_actions.head()
```

Out[31]:

	id	content	fecha	user_id	text_id
58	59	<p>Write your text here</p>	2019-10-20 11:21:46	30	27
61	62	<p>I think you have a critical thinking. Your ...	2019-10-20 19:02:40	26	29
62	63	<p><strong>The Incident and Aftermath</strong>...	2019-10-22 19:46:03	37	30
67	68	<p>NBA's Chinese market grows swiftly into the...	2019-10-23 18:51:10	49	34
68	69	<p>Over the past few decades, researchers have...	2019-10-23 19:11:50	34	35

In [32]:

```
# Get list of user ids for texts that have more than 1 version  
version_user_ids = pd.crosstab(index=useful_actions.user_id, columns="count")  
version_user_ids
```

Out[32]:

col_0	count
user_id	
16	3
17	3
18	3
19	1
21	9
26	2
30	1
31	1
32	11
34	2
35	3
36	1
37	3
43	1
45	2
46	1
49	1
50	2
53	1
55	2
57	1
62	3
63	11
65	1
66	1
71	1
74	4
79	1

In [33]:

```
# wtf is happening with users 32 and 63
user_32 = subset_text_actions[subset_text_actions.user_id == 32]
user_32
```

Out[33]:

	id	fecha	action	user_id	text_id
234	238	2019-10-24 09:20:26	4	32	NaN
294	298	2019-10-25 15:17:47	6	32	49.0
295	299	2019-10-25 15:17:53	2	32	49.0
296	300	2019-10-25 15:23:50	6	32	49.0
297	301	2019-10-25 15:24:02	2	32	49.0
...	...	...	...	...	...
1065	1069	2020-04-28 18:09:50	3	32	190.0
1093	1097	2020-08-02 00:29:16	4	32	NaN
1094	1098	2020-08-02 00:29:16	6	32	196.0
1095	1099	2020-08-02 00:51:48	6	32	196.0
1096	1100	2020-08-02 00:51:52	6	32	196.0

84 rows × 5 columns

In [34]:

```
# subset_text_actions.to_csv(r'user_actions.csv', index=False)
```

How many users actually made changes to their texts?

In [35]:

```
### Take all the users that are in the final dataset
active_users_tab = pd.crosstab(index=final_dataset.user_id, columns="count")

# Turn active_users_tab to a list so it can be extracted from the dataset
useful_active_users = active_users_tab.index.tolist()

# use the list to extract only the texts that have more than one version
active_users = clean_actions[clean_actions.user_id.isin(useful_active_users)]
active_users.head()
```

Out[35]:

	id	fecha	action	user_id	text_id
131	135	2019-10-17 11:23:17	1	17	NaN
137	141	2019-10-18 10:38:47	1	22	NaN
140	144	2019-10-19 19:39:29	1	26	NaN
143	147	2019-10-20 09:28:20	1	29	NaN
146	150	2019-10-20 13:42:56	1	32	NaN

In [36]:

```
# Just checking they're all there  
active_users_check = pd.crosstab(index=final_dataset.user_id, columns="count")  
len(active_users_check)
```

Out[36]:

28

In [37]:

```
active_users.to_csv(r'active_users.csv', index=False)
```



## **Appendix 4 – Invitation letter, info sheet and consent form for Study 2**

### **Invitation letter**

Hi!

My name is Ana Hibert, and I'm a PhD student researching ways in which technology can help students become more proficient English writers. I am currently recruiting people for a four-week research project in the use of a special type of software that helps students revise their academic writing.

### **Who am I looking for?**

I am looking for postgraduate students whose first language is not English.

### **What will the study involve?**

Four sessions in which you sit down at a computer and receive feedback from the computer to revise the writing in your dissertation. You will be asked to talk about what you're doing as you revise, and the screen will be recorded. No personal information will be obtained, and your face will not be recorded. You will also be asked to do two short interviews: one at the beginning and one at the end of the project.

### **What's in it for you?**

You will have the opportunity to revise your dissertation, or any other piece of writing you want! This will help you have a more polished product and will help you improve your English writing skills.

### **When will the sessions be?**

The sessions should be once a week during the month of July. Ideally, we'll find dates and times that work for you!

If have any questions, or you're interested in participating in this study and receiving help in proofreading and revising your dissertation (or any other piece of writing you need to polish!), please contact me at

**Information sheet and consent form**

**Information sheet and consent form**

**The use of Automated Writing Evaluation programs in Higher Education**

**Researcher: Ana Isabel Hibert Santana**

**Moray House School of Education**

**This Informed Consent Form has two parts:**

- **Information sheet (to share information about the study with you)**
- **Consent form (for you to sign if you agree to take part in this study)**

## **Part I: Information sheet**

I am Ana Hibert, a PhD student at the Moray House School of Education, and I am researching ways in which Automated Writing Evaluation (AWE) programs can help students become more self-sufficient English writers and improve the quality of their texts.

AWE programs are designed to give students automated and immediate feedback on their writing through the use of natural language processing techniques. These programs can give students feedback on not only grammar and spelling, but style, organization and sometimes even content.

AWE programs are being increasingly used in higher education to help English as a second language (ESL) students improve their writing skills and help them succeed in English language programs.

### **Purpose of the research**

This research focuses on studying how ESL students engage with AWE programs and use them to improve their written texts, revision strategies and writing skills.

### **Participant selection**

You are being invited to participate in this study because you are an international ESL student enrolled at a postgraduate program in the University of Edinburgh.

### **Procedure**

This study will begin with a brief questionnaire about your study habits and writing strategies. For the main part of this pilot, you will be asked to attend four 30-minute sessions, during which you will be given a chance to revise a text of your choice (it can be your dissertation or any other coursework, the only requirement is that the text be in English). During each of those sessions, you will be asked to participate in a think-aloud protocol, which means that you will be asked to talk

about what you're doing as you use the AWE program to revise your text. Your voice and activity on the screen will be recorded during each session for further analysis. At the end of the study, you will be asked to participate in a brief interview about your experiences using the program.

### **Confidentiality**

I will not be sharing information about you to anyone. While your voice will be recorded in both the interview and think-aloud protocols, your face will not be.

The data I collect from you will be anonymized and kept private: all the information will have a number on it instead of your name, and only I will know what that number is, they key will be stored in an encrypted file. The videos will be securely stored in an encrypted device, and all transcripts will have any identifying information about you, your writing or your program scrubbed from them.

Furthermore, I will keep no files belonging to you. The computer session used during the recordings will be wiped clean of all files, history and information, so nobody else will have access to your work, identity or any other information related to you.

### **Sharing the results**

Part of the transcripts may be shared with my PhD supervisors, but as stated above, all identifying information will be scrubbed from them.

The results will be used as part of my doctoral thesis, and I may publish a paper based on this research. If the research is published, anonymity will be respected as stated in the previous section.

If you have any questions about this project, either now or in the future, please contact me at

[ana.hibert@ed.ac.uk](mailto:ana.hibert@ed.ac.uk)

**Part II: Certificate of Consent**

I have read the foregoing information, or it has been read to me. I have had the opportunity to ask questions about it and any questions I have been asked have been answered to my satisfaction. I consent voluntarily to be a participant in this study

**Print Name of Participant** \_\_\_\_\_

**Signature of Participant** \_\_\_\_\_

**Date** \_\_\_\_\_

Day/month/year

**Statement by the researcher/person taking consent**

I have accurately read out the information sheet to the potential participant, and to the best of my ability made sure that the participant understands what the think-aloud protocols and interviews entail.

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I

confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

**A copy of this ICF has been provided to the participant.**

**Print Name of Researcher** \_\_\_\_\_

**Signature of Researcher** \_\_\_\_\_

**Date** \_\_\_\_\_

Day/month/year

## Appendix 5 – Semi-structured interview questions

1. What are your overall impressions of ProWritingAid?
  - Quality of feedback
  - Ease of use
  - Relevance of feedback
  - Helpfulness
2. Which features did you find most useful?
  - For planning revisions
  - For revising
  - For noticing common mistakes
  - For reflecting on your writing
3. Which features did you find the least useful?
  - Limitations of the program
  - Sources of frustration
4. Would you continue using the program in the future?
  - Why/why not
  - Future uses they might envision

## Appendix 6 – Sample transcript for study 2

Timespan	Content
0:09.0 - 0:39.5	Today I'm going to look over the Methodology. It's a bit short, but still... I feel like I wrote it in a hurry. Let's see. Style... grammar... I think style, because I think something...
0:39.5 - 0:48.3	I want to see how consistent it is in relation to my other chapter, so I want to see how it goes.
0:48.2 - 1:18.9	"In order" ... Delete... (reading). Yes. I use a lot of "in order to" and it's always correcting me, telling me to get rid of it. I think it's going to make me delete more here because there's a lot of words, and my teacher did recommend me to lower the count.
1:18.8 - 1:25.6	"In order"... yeah, I'm also going to delete this one.
1:25.5 - 2:05.3	"To" ... To.. to... I have a lot of "in order to". It's much more... It flows better. And then I save two words per sentence. Delete, delete, delete. (reading) Yes.
2:05.3 - 3:33.7	"That" ... Let's see "that". (reading) Yeah. I'll delete that. (reading), (rehearsing). If I delete... I'm going to change (xxx). Another one. (reading). (rehearsing) I'm going to shorten it. Yeah. And another one. (reading). Yeah, and here I can also, "to corroborate".
3:33.6 - 3:57.4	"Possible", let's see. (reading). No, I can't get rid of that possible. I feel like it's okay, so I'm going to leave it there.
3:57.3 - 4:24.8	"As well as" (reading). "As a guide". Okay. (rehearsing). Yeah. There is an "and" close by. Okay.
4:24.0 - 4:57.9	(reading) Mmm. (reading). (rehearsing). No, here I feel like I'm not going to change it. So, next one...
4:57.8 - 5:28.4	(reading) You could also remove this... it makes sense there, but... I feel like it's not right. I'll ignore it.
5:28.4 - 5:39.5	"Participant". Take part... no. Can't I go back? "take part", "take part"... no. Ignore.
5:39.4 - 6:09.7	"Lastly"... To "last". Hmm. (reading). No, "Finally". And the other "lastly" was here. Mmm. I'll leave it. Ignore.
6:09.7 - 6:42.7	"Whereas". Let's see. (reading). Whereas... while... yeah. And the other one? (reading). No, I'll leave that one there, to have a bit more variety.
6:42.7 - 7:28.0	"Objective". (reading). R: I did click record, didn't I?



	<p>Yes.  R: Oh, yes. Whew. Sorry.  (reading). Aim. Yeah, I've used "objective" too much. School's aims. And the other one? (reading). Aim, yes.</p>
7:28.0 - 7:37.0	(reading). Yeah, this way I get rid of a redundnace.
7:36.8 - 7:51.8	(reading) "In addition"... Oh. Yeah, I'll remove it.
7:51.7 - 8:10.2	"Additionally, the" .... (reading) "they were pro... provided in the pilot sessions". (reading). Yeah, a lot of "additionally" and "also".
8:12.0 - 9:05.8	(reading). (rehearsing). Yes. Yes. (reading). Oh, yeah, the... Hmmm... Maybe if I delete this. (rehearsing). Yeah.
9:05.7 - 9:36.8	(reading). Yeah. (rehearsing). For the transition, because I think it's missing a word there to... for the transition.
9:36.8 - 9:48.9	(reading). Divided? No, that doesn't sound right.
9:48.8 - 10:37.3	"a period of" ... (reading) "in a period of one month" .... "in one month" ... yeah, it was too redundant, and unnecessarily long. (reading) Yeah. Yes. I don't know why I tend to make my writing so long.
10:37.2 - 10:55.0	(reading) Not "taking part". (reading). Evaluate? Decide. (It was xxx) but in the end I had to consider it.
10:54.9 - 11:02.4	"Methodology"... ah, no, that's the chapter. That's the name of the chapter.
11:02.4 - 11:45.2	"Additionally, their" ... (reading). (rehearsing). Yeah, that's fine. (re-reading). Mhmm. Yeah, I'll leave it like that, otherwise it's going to be too long.
11:45.1 - 12:08.2	"Currently". (reading) Mmm... I can take that out. (re-reading). "work with high school levels". Yeah.
12:08.1 - 12:29.0	"Considered" ... (reading) (rehearsing). Yes.
12:29.0 - 13:02.4	"In light of" ... mmm... "in light of" ... "considering" ... (rehearsing). Yeah, I'll make it less long. (re-reading). And that's much better. The other one was too long. It's easier to understand as well. (re-reading)
13:09.2 - 13:56.3	(reading). No, I do need to keep a connector here. I'll just leave it, otherwise I feel you'd lose coherence.
13:56.3 - 16:24.4	"on the other hand" ... (reading) Well, I deleted "they" there, because I have two clauses and I'm making a reference to Dennise and Natalia. It would be too redundant, and maybe not too grammatical. But "on the other hand"... I'm contrasting here. Maybe if I get rid of it... (re-reading). And I might have to cut this, because otherwise I'd have two adjective clauses, and it's a bit disruptive, I think. Maybe I can add at the end. (re-reading). (rehearsing). Yeah. That's well.

	Cause I had two one after the other and there was too much information before I said what I actually wanted to say. This is also important, but I think I could leave it for the end as something a bit more peripheric and focus on the other things. Yeah.
16:24.4 - 16:44.7	"is not allowed to" Hmm... (reading). No, I'll leave it like that because I want to emphasise that I couldn't because I didn't get permission, not because I didn't want to or because I didn't try.
16:44.6 - 17:15.4	"in nature" ... (reading) Mm... I think I could erase it. It's more for... I don't know. It sounds prettier, but this way it's more direct, and that's what they're looking for, for me to get to the point.
17:15.5 - 17:34.3	(reading) (rehearsning). Oh, I hadn't edited that sentence. (reading). (rehearsing) ... yeah, that's better. I need to cut a lot.
17:34.2 - 18:18.1	(reading). What if I write "it also" ... (re-reading). Hmm, I'm speaking in present, I should stay like that. "makes up"? No.
18:18.1 - 19:07.8	Phrasal verbs. (reading). (rehearsing). I should change this. I think I want to emphasise that what I wrote here it was really a suggestions, but they had the freedom... (re-reading). Yeah, it's fine. Much more direct.
19:07.8 - 19:25.4	Okay. (reading). (rehearsing). Yeah, much better that way. (re-reading). Yeah.
19:25.4 - 19:40.8	"consisting of" ... Comprising? (reading). (rehearsing). Yeah.
19:40.8 - 20:12.8	"employed" ... no, employed. (reading). Mmm. Yeah. Better this way. I don't know if it's more direct, but it's more simple. They told me to make this chapter really direct with everything I've done.
20:12.7 - 21:57.6	"at the same time" ... (reading). No, I need to keep it, because they were doing it at the same time. I'll leave that there. But this sentence... (reading) So I need to see what they "used to strengthen" ... What did I mean to say here? (re-reading). (rehearsing). I'd better separate this. "Their comments were used"... because otherwise, what I wanted to say, of course, that I used their comments to strengthen the analysis, which I'd already done, but before I think it implied that I used the observations, that I'd used my interpretation but that's not what I...
21:57.5 - 22:15.3	"as well"... (reading) (re-hearsing). Yeah.
22:15.3 - 22:19.6	I've still got a lot to do. "In order to" ... "to" ... I'd already...
22:19.6 - 22:51.6	"the scope of" ... (reading). (re-reading). (rehearsing). Nice.
22:51.5 - 22:55.5	"Obtaining" ... (reading). Very informal.
22:55.5 - 23:14.1	"At any time" ... (reading). No, I'm going to leave taht there.

23:14.0 - 23:25.1	(reading). (rehearsing). yeah.
23:25.1 - 23:49.0	"located" ... (reading). (rehearsing). Yeah, that's better.
23:49.0 - 25:30.4	"You have started 3 sentences in a row with the same word". Ugh. The, the, the. (reading). How can I change "the"? (reading). (rehearsing). Yeah, I'll accept that one. (rehearsing). Yeah. (re-reading). (xxx) (re-reading). (rehearsing). (re-reading). (reading). Yeah, because I was repeating myself there. I should've changed it before.
25:30.3 - 25:39.4	"The researcher"... (rehearsing).
25:39.4 - 26:09.0	"were used as" ... (reading). "Served" sounds better than "were used", and I get rid of the passive voice. (re-reading).
26:09.0 - 26:50.2	(reading). (rehearsing). because I think I'd mixed two sentences here and then I didn't fix them. (re-reading). (rehearsing). Yeah, this is much more direct and you can understand what I'm trying to say.
26:48.0 - 28:07.9	(reading). (rehearsing). (reading). No, this is very repetitive. But I'm going to add "journals" up here. (reading). (rehearsing). (reading). I'm just going to leave that there because I can't think how to change it. (reading). (rehearsing). Yeah, great.
28:07.9 - 28:11.0	"were able to"... "could"... yes.
28:11.0 - 28:27.9	"limited number of"... (reading) (rehearsing).
28:27.8 - 28:38.2	"the purpose of" ... (reading) (rehearsing). Hmm.
28:38.2 - 29:02.3	"willing to" ... (reading). Mmm (re-reading). (rehearsing). Okay.
29:02.3 - 30:06.7	(reading). (rehearsing). (reading). (rehearsing). Yeah, this is a better sentence. (reading). Yeah. (reading). Attempt... try... no. (reading). No, I'm going to leave that there.
30:06.6 - 30:14.0	There's a lot of passive verbs. I'm going to see the hidden verbs.
30:13.9 - 30:31.7	(reading) (rehearsing).
30:31.6 - 30:46.2	"the application of" ... (rehearsing).
30:46.2 - 30:58.0	And "a reduction of" ... Ah, but that's a quote, I can't change that.
30:57.9 - 31:13.5	No style improvements found, no long subordinate clauses, two adverbs found. Use adverbs sparingly in your writing.
31:13.4 - 31:25.1	(reading). (rehearsing).
31:25.1 - 31:55.0	67 passive verbs. 3 repeated sentence starts. Hmmm. (reading). But I already changed that.

31:54.9 - 32:16.1	Mmm. Sentence. R: It has been half an hour. Do you want to keep on going another 15, or...? Yeah. R: Yeah? Okay. I'll let you know in 15.
32:16.0 - 34:01.3	Two thousand seven hundred words. Average length, 11 to 18. Mmmm. 22 long sentences. 45 words! Mmm. (reading). Yeah, I'm going to divide it here. (rehearsing). 23... 35... (reading). I'll just... (rehearsing). Yeah.
34:01.2 - 35:24.9	I'm going to... Yeah. 27, that is fine. 24, yeah, yeah. 37. This one is too long. (reading). Maybe I can... (rehearsing). No. If I write it... (rehearsing)... No, but I chose them because of that. (re-reading). (rehearsing). Yeah, that's better.
35:24.9 - 36:45.5	39... (reading). (rehearsing). Yeah, better. (re-reading). No, this is poorly writting. (re-reading). (rehearsing). Yeah. (re-reading). I'm repeating myself. (re-reading). (rehearsing).

## Appendix 7 – Code for cleaning up and analysing student actions