DOCTORAL SCHOOL OF EDUCATION
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
UNIVERSITY OF SZEGED

NANG KHAM THI

THE ROLE OF FEEDBACK IN THE PROCESSES AND OUTCOMES OF ACADEMIC WRITING IN ENGLISH AS A FOREIGN LANGUAGE AT INTERMEDIATE AND ADVANCED LEVELS

DOCTORAL DISSERTATION

SUPERVISOR:
PROFESSOR MARIANNE NIKOLOV

SZEGED, HUNGARY, 2023

ABSTRACT

Providing feedback on students' texts is one of the essential components of teaching second language writing. However, whether and to what extent students benefit from feedback has been an issue of considerable debate in the literature. While many researchers have stressed its importance, others expressed doubts about its effectiveness. Regardless of these continuing and well-established debates, instructors consider feedback as a worthwhile pedagogical practice for second language learning. Based on this premise, I conducted three experimental studies to investigate the role of written feedback in Myanmar and Hungarian tertiary EFL classrooms. Additionally, I studied syntactic features and language-related error patterns in Hungarian and Myanmar students' writing. This attempt was made to understand how students with different writing proficiency acted upon teacher and automated feedback.

The first study examined the efficacy of feedback on Myanmar students' writing over a 13-week semester and how automated feedback provided by Grammarly could be integrated into writing instruction as an assistance tool for writing teachers. Results from pre-and post-tests demonstrated that students' writing performance improved along the lines of four assessment criteria: task achievement, coherence and cohesion, grammatical range and accuracy, and lexical range and accuracy. Further results from a written feedback analysis revealed that the free version of Grammarly provided feedback on lower-level writing issues such as articles and prepositions, whereas teacher feedback covered both lower-and higher-level writing concerns. These findings suggested a potential for integrating automated feedback into writing instruction.

As limited attention was given to how feedback influences other aspects of writing development beyond accuracy, the second study examined how feedback influences the syntactic complexity of Myanmar students' essays. Results from paired sample t-tests revealed no significant differences in the syntactic complexity of students' writing when the comparison was made between initial and revised texts and between pre-and post-tests. These findings suggested that feedback on students' writing does not lead them to write less structurally complex texts despite not resulting in syntactic complexity gains. The syntactic complexity of students' revised texts varied among high-, mid-, and low-achieving students. These variations could be attributed to proficiency levels, writing prompts, genre differences, and feedback sources.

The rationale for conducting the third study was based on the theoretical orientation that differential success in learners' gaining from feedback largely depended on their engagement

with the feedback rather than the feedback itself. Along these lines of research, I examined Hungarian students' behavioural engagement (i.e., students' uptake or revisions prompted by written feedback) with teacher and automated feedback in an EFL writing course. In addition to the engagement with form-focused feedback examined in the first study, I considered meaning-focused feedback, as feedback in a writing course typically covers both linguistic and rhetorical aspects of writing. The results showed differences in feedback focus (the teacher provided form-and meaning-focused feedback) with unexpected outcomes: students' uptake of feedback resulted in moderate to low levels of engagement with feedback. Participants incorporated more form-focused feedback than meaning-focused feedback into their revisions. These findings contribute to our understanding of students' engagement with writing tasks, levels of trust, and the possible impact of students' language proficiency on their engagement with feedback.

Following the results that Myanmar and Hungarian students responded to feedback on their writing differently, I designed a follow-up study to compare syntactic features of their writing as indices of their English writing proficiency. In addition, I examined language-related errors in their texts to capture the differences in the error patterns in the two groups. Results from paired sample t-tests showed that most syntactic complexity indices distinguished the essays produced by the two groups: length of production units, sentence complexity, and subordination indices. Similarly, statistically significant differences were found in language-related error patterns in their texts: errors were more prevalent in Myanmar students' essays. The implications for research and pedagogical practices in EFL writing classes are discussed with reference to the rationale for each study.

ACKNOWLEDGEMENTS

Writing the acknowledgements section is both exciting and emotional for me as I have never been away from my parents for such a long period of time. Many failures and successes in my academic work that I have experienced during this exciting PhD journey have pushed me to always stick to my goals. Reflecting on my past four years, I would like to extend my sincere gratitude to people who have helped me.

I would like to acknowledge the contribution and support of my supervisor, Marianne Nikolov, who has been always ready to help with all research-related issues. Her constant support, guidance, and encouragements during hard times motivated me to accomplish the PhD dissertation. Expressing gratitude would never be enough, as she has been the one who cared for me the most during these four years.

I would like to express my great appreciation to all professors at the Doctoral School of Education who acquainted me with the required knowledge of educational research. Taking courses and attending weekly seminars have given me the foundation for growing as a novice researcher. I appreciate all instructors' academic support and feedback, and I took their comments as opportunities to learn and grow.

I greatly appreciate the support provided by the University of Szeged, which has allowed me to continue my research. Most importantly, the Klebelsberg Könyvtár (tik) is the place I spent most of my time during the past four years. The quiet atmosphere and picturesque views from the big windows created a perfect environment for my ideas to flourish.

I would also like to express my sincere gratitude to all my friends at the Doctoral School of Education. They have been always there for me whenever I needed help and guidance. The time we spent chatting, laughing, and enjoying one another's company was one of the best times of my life. Thank you for listening to the ups and downs of my personal life and expressing sympathy.

My heartfelt thanks go to those who contributed to my research, including my colleagues in Myanmar (Burma), the teachers at the Department of English Applied Linguistics, University of Pécs, and all the students who have voluntarily participated in my research. I am particularly indebted to Dr Vo Van De for his endless support of my research from the time he acted as a reviewer on my research proposal to this day.

Finally, I would like to express my warmest gratitude to my parents who have listened to my emotional stress and cheered me up when I needed their support. They have done everything in their power to motivate and encourage me to complete this challenging journey.

TABLE OF CONTENTS

LIST OF ACRONYMS

AES	Automated essay scoring
AWE	Automated Writing Evaluation
AI	Artificial Intelligence
BA	Bachelor of Arts
CAF	Complexity, Accuracy, and Fluency
CEFR	Common European Framework of Reference for Languages
CN/C	Complex Nominals per Clause
CN/T	Complex Nominals per T-unit
CR	Correct Revision
CS	Communicative Skills
C/S	Sentence complexity ratio
CP/C	Coordinate phrases per clause
CP/T	Coordinate phrases per T-unit
C/T	Clauses per T-unit
D	Deletion
DC/C	Subordinate Clauses per Clause
DC/T	Dependent clauses per T-unit
DWCF	Dynamic Written Corrective Feedback
EFL	English as a Foreign Language
ELT	English Language Teaching
ENA	Epistemic Network Analysis
ERIC	Educational Resources Information Centre
ESL	English as a Second Language
ESOL	English to Speakers of Other Languages
FYC	First Year Composition
GMAT	Graduate Management Admission Test
GRE	Graduate Record Examinations
IELTS	International English Language Testing System
IR	Incorrect Revision
LAA	Language Analytic Ability
L1	First Language
L2	Second Language

L2SCA	L2 Syntactic Complexity Analyzer
MLC	Mean Length of Clause
MLS	Mean Length of Sentence
MLT	Mean Length of T-unit
NCEL	National Centre for English Language
NR	No Revision
NUCLE	NUS Corpus of Learner English
PE	Proficiency Examination
PEG	Project Essay Grade
PRISMA	Preferred Items for Systematic Reviews and Meta-Analysis
RQ	Research Question
SLA	Second Language Acquisition
SYNLE	Left embeddedness, words before main verb
SYNNP	Number of modifiers per noun phrase
SYNSTRUTt	Sentence syntax similarity
TESOL	Teaching English to Speakers of Other Languages
TOEFL iBT	Internet-based Test of English as a Second Language
T/S	Sentence coordination ratio
VP/T	Verb Phrases per T-unit
WCF	Written Corrective Feedback

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

This chapter presents the background of the research projects and some relevant concepts, such as the role of feedback in developing second language writing and what has been discovered in the literature on written feedback in L2 writing. This is followed by an outline of the current research trends, emphasising the complementation of different sources of feedback in EFL students' writing and student engagement with feedback. A summary of the research contexts and general aims and objectives of four experimental studies conclude the chapter.

1.1 Introduction

Second language (L2) writing is an essential component of students' literacy development in school curricula because it is a catalyst for personal and academic advancement. Providing feedback on their writing is generally considered to be an effective pedagogical practice aimed at improving students' writing. In the field of L2 writing, the importance of providing feedback on students' writing has been highlighted by writing teachers. For example, Lee (2008a, 2008b) examined teachers' feedback practices in writing classrooms and found that they provided feedback on a wide range of language- and content-related issues, even though the majority of feedback was error-focused. Also, students' willingness to receive written comments on their writing (Lee, 2008a) and their positive attitudes toward teacher feedback (McMartin-Miller, 2014; Zacharias, 2007) trigger teachers to continue providing feedback on L2 texts.

Despite these practices and students' preferences to receive feedback on their writing, the facilitative role of feedback in L2 writing research has been debated over the past decades. In particular, the debates about the value of feedback were centred on the claim that written corrective feedback (WCF) is "ineffective or harmful", and it should therefore be abandoned (Truscott, 1996, p. 328). The controversy over the value of WCF resulted in considerable discussion and further inquiries, investigating the effectiveness of feedback on student writing. Many studies examined whether and to what extent students benefit from feedback on their writing (Ferris, 2006; Truscott & Hsu, 2008), and compared the relative effects of different feedback types (e.g., Benson & DeKeyser, 2018; Hartshorn & Evans, 2015; Karim & Nassaji, 2018; Kim et al., 2020; Nicolas-Conesa et al., 2019; Shintani et al., 2014; Zhang, 2021).

Given that providing feedback on students' writing requires a great deal of time and effort on the teacher's part (Zhang, 2017), recent studies in WCF research shifted the emphasis to exploring how automated feedback can be integrated into writing instruction to supplement teacher feedback. Time constraints, large class size, and teachers' workload pose major

challenges that prevent them from giving adequate feedback. Consequently, teachers tend to offer feedback primarily on language-related errors rather than content-related issues in students' writing (Lee, 2009). Thus, to ease teacher feedback burden and to enhance the efficacy of teacher feedback, automated feedback can be used. In line with the favourable evidence of the reliability of automated writing evaluation (AWE) feedback (Li et al., 2015), L2 writing scholars (e.g., Koltovskaia, 2020; Ranalli, 2018) recommended integrating automated feedback into writing instruction to free up teachers' time spent on focusing lower-order concerns (e.g., grammar and mechanics) and pay more attention to higher-order concerns (e.g., content and organization).

In studies examining the effect of WCF, the primary aim was to develop accuracy with little consideration of the fact that an increase in accuracy might come at the cost of syntactic complexity. Specifically, although research on the impact of WCF on accuracy development has demonstrated that feedback on student writing is conducive to their accuracy development, analysing accuracy without regard for other dimensions of writing (e.g., complexity) would be meaningless. For example, Polio (2012a) argued that studies on error correction emphasized the importance of feedback on accuracy development, but a likely tendency is that "attention to accuracy could help students' accuracy but harm the fluency or the complexity" (p. 147). In other words, the attention of L2 writers to accuracy tends to divert their attention from other aspects of writing. Therefore, Polio (2012b) suggested that it would be beneficial for WCF studies to examine how feedback affects other aspects of language development, such as complexity and fluency.

A great deal of research on written feedback has investigated the relative effects of implicit and explicit corrective feedback on students' writing (Karim & Nassaji, 2018) and compared feedback from different sources (Dikli & Bleyle, 2014; M. Yang et al., 2006). The underlying hypothesis of most studies is that feedback enables learners to notice the mismatches between the target language and their interlanguage system. However, the mere provision of feedback does not always result in improvement as several factors, including the degree of student engagement with the feedback likely impact the benefits. For example, Zheng and Yu (2018) stated that if learners were not fully engaged with feedback, they were less likely to benefit from it. Zhang (2020) also argued that compared to a narrow focus on accuracy improvement in L2 student writing, it is more meaningful to examine how learners engage with feedback from different sources to enhance the possible benefits. Bearing this in mind, recent studies in written feedback research have shifted attention from investigating the efficacy of different feedback types to exploring student engagement with feedback from multiple sources.

Based upon a brief description of the research trends and what has been discovered in the literature, the present research explored the role of feedback in EFL students' writing from different theoretical perspectives and empirical insights I gained from the previous studies. Particularly, four research projects with specific research aims were undertaken. The first inquiry aimed to explore the effectiveness of written feedback on EFL students' writing and the potential of integration automated feedback into writing instruction. In the case of exploring the efficacy of written feedback, I focused on how students made use of teacher and automated feedback (i.e., Grammarly) in their revisions and how their writing performance improved in the post-tests at the end of the academic semester. The second investigation examined whether feedback had any unfavourable effects on syntactic complexity of students' writing. In the third research project, the focus of attention shifted from product-oriented writing to process-oriented writing, i.e., from focusing on students' writing performance to their behavioural engagement with feedback. More specifically, I explored how students behaviourally engaged with teacher, Grammarly, and combined feedback during the revision process. The fourth inquiry considered the role of syntactic features and their relationships to L2 writing proficiency. This investigation was an attempt to understand how students with different L2 writing proficiency acted upon feedback differently. I hope the findings to inform several important implications for writing instruction, e.g., using Grammarly as a precursor to teacher feedback, and exploiting automated tools (e.g., L2 syntactic complexity analyzer) to analyse or track students' linguistic developments in their writing.

1.2 Research contexts

This section overviews some contextual information about the two educational contexts, Myanmar and Hungary, in which the empirical studies were conducted. I give a brief introduction to the status of English, English language teaching in these contexts, the nature of teaching writing skills in tertiary-level EFL classrooms, and some pedagogical issues in teaching L2 writing in universities. Due to the military coup in February 2021 (Brown & Hung, 2022), almost all schools and universities in Myanmar experienced long-term closures. For this reason, I was unable to continue my projects in Myanmar and decided to carry on my research in a different educational setting.

Though Myanmar was once a British colony, English was regarded as a foreign language in school settings. After independence obtained in 1948, English ceased to be the official language and lost its importance in schools and colleges (Allott, 1985). However, it regained

its power in 1981 when English was taught as a compulsory subject from the first grade in all primary schools and re-introduced as a medium of instruction for the science stream and economics in secondary schools and post-secondary education starting in 1986. In recent years, learning English has become increasingly popular as it provides more opportunities to study or work abroad (Tin, 2014).

In higher education, English is an obligatory subject in all disciplines in undergraduate courses and taught as a specialized subject in English literature, linguistics, and language teaching courses. For example, first-year undergraduate students enrol in a communicative skills (CS) module regardless of their specializations. These foundation courses aim to develop students' communicative skills while engaging them in communicative activities. Particularly for writing skills, different genres of written texts (e.g., narrative and argumentative essays) are introduced to develop students' writing through model texts and follow-up practice. However, concerns have been raised regarding the development of students' writing (Maung et al., 2022; San & Soe, 2016). Some underlying reasons could be explained by institutional, contextual, and teacher-related factors. For first-year English majors, for instance, the prescribed curriculum, New Language Leader (Cotton et al., 2014) employs an integrated approach in introducing all language skills. Within a three-month semester, course instructors are encouraged to finish teaching all units covered in the course book. Consequently, most writing activities and tasks are given as take-home assignments and students do not receive regular feedback on their writing. Also, other contextual constraints including large class size ($50 < N < 70$) and excess workload render it difficult for teachers to offer effective feedback to individual students. Nonetheless, writing accuracy and rhetorical competence (e.g., organization and flow of ideas in students' writing) are highly demanded on important testing occasions like university entrance examinations (Kirkpatrick & Hlaing, 2013) and other high-stakes language proficiency tests like International English Language Testing System (IELTS) and Internet-based Test of English as a Second Language (TOEFL iBT).

In Hungary, the knowledge of foreign languages is important to communicate with citizens in neighbouring and other European countries, as Hungarian belongs to the Finno-Ugric family of languages, whereas most European countries use Indo-European languages. During 1949-1989, for political reasons, even though Russian was taught at all levels of the school system, Hungarians were not willing to learn it, as it evoked oppressive power (Dörnyei & Csizér, 2002). Therefore, after the political transition in 1990, Russian lost ground to other foreign languages, primarily English and German. Though German and English have been the two dominant foreign languages in the Hungarian educational system since the fall of Soviet

occupation, students' interest in learning English has increased due to its status as the lingua franca of science, business, and higher education (Nikolov & Csapó, 2010).

At the university level, those who major in English Studies are required to complete language development courses (i.e., Reading and Writing Skills, and Listening and Speaking Skills courses) in the first two terms of the Bachelor of Arts (BA) in the English programme. A further explanation of Reading and Writing Skills course is necessary for the purposes of the present research. This course aims to enhance undergraduate students' reading and writing skills and prepare them for the corresponding parts of the language proficiency exam (Horváth, 2001). Regarding writing, various writing strategies including narrowing down a topic, prewriting, and using a logical structure are introduced. Also, students complete regular writing tasks targeting different genres of writing (e.g., description essay, classification essay, narrative essay, and definition essay) which are followed by teacher feedback. At the end of the first academic year, students must sit for the language proficiency exam at C1 level upon their completion of the three obligatory courses (i.e., Listening and Speaking Skills I-II, Reading and Writing Skills I-II, and English Grammar in Use I-II).

1.3 The present dissertation

My studies aim to examine the impact of written feedback on EFL students' writing. While previous research has focused on the role of feedback in improving students' accuracy, two gaps are to be addressed. First, these studies primarily concentrated on how students benefit from feedback targeting language errors. L2 teachers, however, provide feedback on both language and other global issues relating to content and organization (Cheng & Liu, 2022). Therefore, in my research I operationalize feedback as responses to learner output ranging from attempts to rectify errors in writing (e.g., grammatical errors; Kang & Han, 2015) to written commentary on content and rhetorical concerns (Goldstein, 2004). Second, much of the existing research examined the effectiveness of focused feedback, in which the teacher/researcher pre-selected specific linguistic features based on individual students' needs (Lee, 2020). While prior research provided evidence that focused feedback facilitates the development of accuracy in specific error types (Bitchener et al., 2005; Bitchener & Knoch, 2009; Rummel & Bitchener, 2015; Shintani & Ellis, 2013), such feedback practices do not tend to reflect the reality of L2 writing classrooms (Van Beuningen et al., 2012) where teachers provide feedback in a more comprehensive manner without having predetermined types of errors in mind. Particularly, despite all these positive contributions to writing classrooms,

scholars have questioned the ecological validity of focused feedback, as feedback in classroom settings is not limited to a number of selective linguistic features, insisting that more research on comprehensive feedback (i.e., unfocused feedback) should be employed to investigate the far-reaching effects of WCF (Storch, 2010; Van Beuningen, 2010). In other words, teachers correct language- and content-related issues rather than focus on a limited number of error types when responding to students' writing. Moreover, though the efficacy of focused WCF has been widely researched, little attention has been paid to the impact of comprehensive WCF (i.e., feedback on all errors/a wide range of errors) (Li & Vuono, 2019; Zhang & Cheng, 2021). Therefore, I examined the extent to which students benefit from comprehensive feedback provided either by the teacher or Grammarly.

Against this backdrop, I reviewed and synthesized the findings from previous studies which investigated the impact and effectiveness of written feedback on English as a second language (ESL)/EFL students' writing. Particularly, the review attempts to present a critical synthesis of research on four key variables which impact the effectiveness of written feedback: research designs, feedback treatments, writing tasks, and accuracy measures. As a result, the findings of this review throw light on the variations in the number of treatment sessions, duration of intervention, feedback treatments, and written accuracy measures with distinctive advantages and pitfalls. Further divergent issues concern the application of different genres of writing tasks that demand learners' various cognitive and linguistic efforts. These variations make it difficult to compare results across empirical studies. The findings contribute to a better understanding of why and how uniform criteria for selection of writing tasks and accuracy measures can ensure the comparability of studies.

Based on this review, four classroom-based inquiries were designed to investigate the effectiveness of written feedback on EFL students' writing. The first study explored the effectiveness of feedback on students' writing performance as assessed on pre- and post-tests over a 13-week semester. Particularly, this study recruited Myanmar EFL students who were enrolled in a first-year undergraduate course (see Section 4.3.2). As explained in Section 1.2, we attempted to address the contextual issues (e.g., time constraints, large classes, and excess workload) encountered by the teachers at the university in focus. To this end, I examined the focus of teacher and Grammarly feedback to indicate the possibility of integrating Grammarly into writing instruction, as a supplement to teacher feedback. Thus, I investigated how students utilised feedback under three conditions (teacher, Grammarly, and combined) and how they perceived the usefulness of feedback from different sources in their EFL course. The findings have pedagogical implications for the integration of Grammarly into teaching L2 writing.

Considering the emphasis of Grammarly feedback on language-related errors as an advantage, writing teachers could use it as a supplementary tool in their classes on a regular basis or encourage students to use it independently. In this way, teacher feedback burden could be reduced and challenges regarding time constraints and inadequacy of attention paid to individuals in large classes could be addressed to a certain extent.

In the literature, concerns were raised with regard to the unfavourable impact of feedback on students' syntactic complexity (Eckstein & Bell, 2021; Hartshorn et al., 2010) which possibly resulted from their attention to producing accurate texts. In response to this need for research, the second investigation focused on exploring whether and how the provision of feedback affected syntactic complexity in students' texts over a 13-week semester. It is important to note that that the data used in this study came from the same participants as those in the first study (see Section 5.3.1). From a pedagogical perspective, investigating syntactic complexity might help teachers gain a better understanding of which aspects of syntactic complexity could or could not be affected by feedback. Moreover, such awareness might provide some indication of whether feedback leads students to produce structurally less complex writing as a result of attempting to improve their linguistic accuracy.

Echoing previous research on the role of student engagement with feedback in determining how students benefit from it (Zheng & Yu, 2018), the third classroom-based inquiry probed Hungarian EFL learners' engagement with teacher and Grammarly feedback in a BA English Studies programme (see Section 6.3.2). To this end, I studied how students behaviourally engage with form-focused and meaning-focused feedback through analysing their revision operations and feedback uptake. In this case, I also examined the accuracy of Grammarly feedback, as automated feedback tends to be fallible which in turn might affect how students engage with Grammarly feedback. Findings from this study inform writing teachers about how they can integrate feedback from different sources into their practice and the areas where students need more assistance to make effective use of feedback. The high accuracy of Grammarly's flagging and correcting specific language errors (e.g., verb tense, articles, and prepositions) allows it to be used selectively. In addition, teachers will have more opportunities to address higher-level writing concerns if they utilise Grammarly as a precursor to teacher feedback.

To understand how students with different L2 writing proficiency responded to feedback, further investigations were undertaken into syntactic features of students' texts. Particularly, as analysing syntactic features in academic writing have gained importance (Biber et al., 2011; Maamuujav et al., 2021), understanding these features in students' written texts could provide

crucial information regarding their L2 writing proficiency. Therefore, the fourth inquiry used the data from Study I and III and investigated how syntactic features helped distinguish writing proficiency of students in Myanmar and Hungary and their language-related errors in writing to indicate the differences in the error patterns. The results of the study contribute to the understanding of the linguistic features of EFL students' texts, including error patterns in writing in two educational contexts.

1.4 The structure of the dissertation

This dissertation includes eight chapters in line with the aims and objectives of the research described in Section 1.3 (Figure 1.1). Chapter 1 outlines the importance of providing feedback to enhance writing development. Referring to previous review studies, a brief account of the debates about the value of feedback on students' writing is also provided. Next, I discuss current research trends, emphasizing the role of automated feedback and how recent studies have focused their attention on students' engagement with feedback rather than their accuracy development. The chapter concludes with an overview of the research contexts and general aims and objectives of the four experimental studies.

Figure 1.1.

Structure of the Dissertation
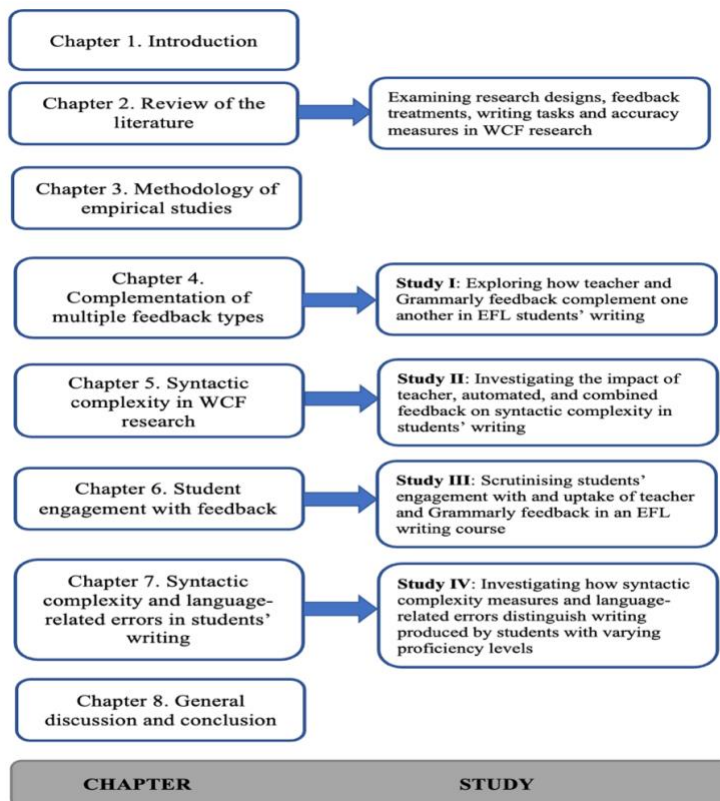
Chapter 2 zooms in on WCF research from various theoretical and empirical perspectives of the role of written feedback in L2 writing while emphasizing the three strands of research (i.e., investigating the effectiveness of teacher and automated feedback, examining the impact of feedback on syntactic complexity, and studying student engagement with feedback) that the present study intends to focus on. The beginning sections of the chapter (see Sections 2.1.1 – 2.1.5) concerns the existing knowledge of what previous research syntheses have established, followed by the details of the current review (e.g., problem statement, and research questions). Findings of the review were published in the *Teaching English as a Second or Foreign Language: The Electronic Journal* (TESL-EJ) (Thi & Nikolov, 2021a). The remaining sections of the chapter discuss the findings from previous studies regarding research strands and gaps in the literature.

Chapter 3 turns to the research design and methodology used in the three experimental studies. Given that these studies are naturalistic classroom-based inquiries, some contextual constraints (e.g., the absence of control groups) are noted. Specifically, the chapter provides an overview of instruments, feedback treatments, and data analysis. In contrast with these three studies, the fourth inquiry examines the differences in syntactic complexity and language-related errors in Myanmar and Hungarian students' texts. Accordingly, no description about this study was made in this chapter for the sake of distinguishing the investigations (Studies I, II, and III) which took place within a specific educational context from the investigation which compared the texts produced by students in the two educational contexts (Study IV, see Chapter 7).

Chapter 4 reports the findings of the first experimental study which explored the potential of integrating Grammarly into writing instruction as a supplementary tool to complement the traditional teacher feedback in an EFL course. It identifies the scope of teacher and Grammarly (Free version) feedback to better understand how these feedback sources could be integrated. Students' successful revisions in response to feedback from multiple sources and their writing improvement on the post-test shed light on the positive impact of feedback on students' writing performance. These findings were triangulated with students' self-assessment questionnaires in which their perspectives on the usefulness of feedback were elicited. Some of these findings were published in the *Asia-Pacific Education Researcher* (Thi & Nikolov, 2021b).

While WCF may contribute to linguistic accuracy, little evidence suggests that it promotes the development of syntactic complexity. In response to the conflicting findings and scarcity of studies investigating how WCF affects syntactic complexity, Chapter 5 details the second experimental study which examined the impact of WCF on syntactic complexity in students'

texts. Initial findings revealed no significant differences between first drafts and revised texts, resulting in minimal variance between comparison pairs. Moreover, no significant differences were found on the pre- and post-tests on all complexity measures. These findings suggest that providing feedback on students' writing does not lead them to write less structurally complex texts. This study was published as a regular article in the *Asian-Pacific Journal of Second and Foreign Language Education* (Thi & Nikolov, 2023).

Chapter 6 comprises the third experimental study: it explored Hungarian EFL students' behavioural engagement with form- and meaning-focused feedback. After identifying the focus of teacher and Grammarly feedback (Paid version), I studied how the students engaged with feedback through analysing the revision operations in their revised texts. The results showed differences in feedback focus (the teacher provided form- and meaning-focused feedback) with unexpected outcomes: students' uptake of feedback resulted in moderate to low levels of engagement with both teacher and Grammarly feedback. The findings of this study were published in *Innovation in Language Learning and Teaching* (Thi et al., 2022). Recognizing the fallibility of automated feedback, the accuracy of Grammarly feedback was also examined in this project.

Chapter 7 presents the results of the fourth inquiry: it examined whether syntactic complexity and language-related errors can help differentiate written texts produced by students at varying proficiency levels. This investigation helped us better understand how students at varying levels of writing proficiency responded to feedback differently. Findings suggested that most complexity measures distinguished the texts produced by Myanmar and Hungarian students. Further examination into students' language-related error patterns also revealed statistically significant differences. This study is to be published in Volume 13, Issue 1 in the *Language Learning in Higher Education* (Thi & Nikolov, in press).

Chapter 8 presents a bird's eye view of my research within a rich field of inquiry in light of the results of the four studies and summarises the implications for research and pedagogical practice. It also discusses the limitations of the studies and outlines directions for future research.

CHAPTER 2. REVIEW OF THE LITERATURE

Three main sections make up this chapter to highlight the main findings in the literature. The first section focuses on the construct of feedback; it provides a comprehensive overview of theoretical and empirical perspectives of the role of written feedback in L2 writing. Then, it discusses the differing views on the efficacy of written feedback. The second part of the chapter analyses previous studies on four key variables of WCF research: research design features, feedback-related features, writing task-related features, and accuracy measures. The final section addresses the research gaps to be filled in light of the knowledge gained from the literature. This section emphasises the effects of teacher and automated feedback on L2 writing, the impact of feedback on syntactic complexity, syntactic features and their relationships to L2 writing proficiency, and student engagement with feedback.

2.1 The construct of feedback

2.1.1 Understanding dimensions of feedback in second language writing

Feedback is regarded as a central concept in language learning; it is viewed as a means to ensure language accuracy and to foster learner motivation (Ellis, 2009a, 2009b). By bridging the gap between students' existing knowledge (i.e., what is understood) and the target language (i.e., what is aimed to be understood), feedback identifies areas for further improvement. Feedback is conceptualized as a consequence of performance, as it is information provided by an agent (i.e., teacher, peer, or self) regarding aspects of one's performance or understanding (Hattie & Timperley, 2007).

Feedback can be either positive or negative. Positive feedback, also known as praise (Zhou, 2022), is any feedback provided to reinforce or affirm that a learner's response to an activity or a task is correct, highlighting the linguistic correctness or the quality of idea development and argumentation. For example, teachers praise the strengths of students' texts and provide affective motivation to subsequent writing. In other words, teachers provide commentary feedback on meaning-related issues (Kepner, 1991) such as drawing their attention to the communicative intent or organization of texts. Generally, positive feedback is provided on the aspects of linguistic performance, task achievement, organization, guiding learners in themes and so on. However, little attention has been paid to the role of positive feedback in second language acquisition (SLA) research (Hyland & Hyland, 2001). The underlying reasons might be related to the fact that too many positive comments (e.g., praise) may confuse and mislead

students and that criticism seems to be more effective than praise in facilitating student improvement (Ferris et al., 1997; Hyland & Hyland, 2001; Zhou, 2022).

Negative feedback, also known as corrective feedback or error correction is defined as "responses to a learner's non-target-like L2 production" (Li, 2010, p. 309). These responses refer to "either an indication that an error has been committed or the provision of the correct target language form or the metalinguistic information about the nature of the error or any combination of these" (Ellis et al., 2006, p. 340). Even though the nature and functions of positive and negative feedback are different, both play essential roles in helping students improve their overall writing performance. It is possible for a teacher to provide both positive and negative feedback when responding to students' texts. In this regard, indication of language errors (i.e., negative feedback) and providing encouragement and positive comments on form and content (i.e., positive feedback) will result in writing development and positive attitudes towards feedback.

As there has been an upsurge of interest in determining the role of negative feedback and corrective feedback in the literature, the term has been defined by different scholars. Russell and Spada (2006) define WCF as "any feedback provided to a learner, from any source that contains evidence of learner errors of language form" (p. 134). Similarly, Loewen (2012) views it as "information provided to students about the ill-formedness of their L2 production" (p. 24). Ellis (2009) stresses that negative feedback is corrective in nature, as it signals that a student's written text contains linguistically deviant elements.

2.1.2 Theoretical foundations of the use of written feedback

The role of feedback in L2 writing was not noticeable in the early 1900s, as errors were to be avoided, and students were given opportunities to practice the right models. Moreover, writing in classrooms was regarded as a means to practice grammatical structures rather than a method for communicative purposes (Ferris & Hedgcock, 2005). However, when behaviourism (Schunk, 2012) became the leading psychological viewpoint in education, the role of feedback was seen as significant, as the concept of reinforcement tended to be related to that of feedback; feedback acts as a reinforcer to impact performance. In line with the theoretical perspectives of behaviourism, language learning is viewed as a process of forming associations between stimuli and responses in which reinforcement (e.g., tangible rewards and informative feedback) is a necessary condition to elicit the desired response from a learner who is presented with a target stimulus. Therefore, the proponents of behaviourism make use of strategies to establish

the stimulus-response associations including the use of instructional cues, practice, and reinforcement. They asserted that errors should not be tolerated, as they can be part of habit-forming and will hinder target-like habit in language learning (Bitchener & Ferris, 2012). Instead, they laid emphasis on the role of positive feedback with target behaviours or habits by means of rewards or other positive responses from the environment.

The three theoretical perspectives, namely Noticing Hypothesis (Schmidt, 2010), the Skill Acquisition Theory (DeKeyser, 2007), and Interaction Hypothesis (Long, 1981) support the role of written feedback in improving students' writing. First, in terms of Noticing Hypothesis, Schmidt (2010) stated that noticing an error plays a major part in language learning and through noticing or conscious attention, students will become aware of their gaps in their language proficiency. It is only through conscious attention that input can be converted into intake for language learning. Specifically, Adams (2003) suggested that noticing the gap, a major aspect of noticing, occurs when L2 writers receive WCF and notice the difference between the target language and their written output. In other words, provision of WCF functions as a noticing facilitator, as it draws students' conscious attention to the form and content of their written output to notice the gaps between their interlanguage (i.e., a natural language produced by L2 learners) and the target language.

Another theoretical perspective that reinforces the role of WCF is the Skill Acquisition Theory (DeKeyser, 2007) which views language learning as equivalent to learning other complex cognitive skills. Different schools of psychology provide the foundations for Skill Acquisition Theory, ranging from behaviourism to cognitivism to connectionism (Dekeyser & Criado, 2013). The majority of WCF research (Frear & Chiu, 2015; Hartshorn et al., 2010; Hartshorn & Evans, 2015) has been grounded in the Skill Acquisition Theory, insisting that accuracy is a matter of practice and thus, an effective balance between explicit instruction and extensive practice plays a crucial role for overall accuracy gains. In consonance with this theory, DeKeyser (2007) indicated three cognitive stages through which students progress in language learning: (1) acquisition of declarative knowledge, (2) proceduralization, and (3) automatization.

In the first stage of acquiring declarative knowledge, students attain new factual knowledge as formal rules or exemplar-based analogies. Then, it is converted into procedural knowledge when students are engaged in activities to practice their skills and subskills based on their declarative knowledge. In the final stage, procedural knowledge is fine-tuned and automatized, allowing for skills to be efficiently performed by changing the scope of the application of rules (Leeman, 2010). A close examination of the theory revealed that declarative knowledge is

required for the development of procedural knowledge. Developing declarative knowledge requires explicit rules, numerous examples, and extensive practice, which leads to greater automatization. The two key tenets of Skill Acquisition Theory are that accuracy is a function of practice and that explicit instruction and extensive practice are preconditions to converting declarative to procedural knowledge. This approach provides a theoretical basis for some L2 practices (e.g., providing feedback).

In the case of WCF in L2 writing, declarative knowledge (what one knows) can be regarded as the knowledge that is learnt during controlled practice and procedural knowledge (what one can do) as the ability to put into practice before it becomes automatized. Precisely, declarative knowledge can be learned through WCF which is provided either by teachers, peers, or computer-mediated tools and through controlled practice (i.e., revision of texts following feedback). It is noteworthy that I do not distinguish between learning and acquisition, as WCF can either be provided explicitly or implicitly and whether students notice the provision of feedback depends on a number of factors such as their linguistic knowledge about the target language, their engagement with the feedback, their attitudes towards error correction, and their beliefs about the roles of errors in developing their writing skills. In the stage of proceduralization, fine-tuning, and automatization, learners can rely on their declarative knowledge they have acquired or learnt through frequent practice and the provision of feedback and apply their knowledge while giving greater attention on the areas that they need to develop further.

The other theoretical underpinning which stressed the importance of WCF is the Interaction Hypothesis (Long, 1981). Studies were couched in the Interaction Hypothesis, especially oral feedback studies, which examine how interactional communication between teacher (i.e., feedback provider) and students leads to language development. The hypothesis stresses that interaction and negotiation of meaning occur when there is a communication breakdown between interlocutors and such interaction enhances L2 learning, as learners can get access to comprehensible input (Krashen, 1985). Additionally, through interaction, teachers can draw learners' attention to linguistic forms and provide feedback on their language. Though Long's (1981) theory was primarily intended for oral feedback, it is also applicable to written feedback as the two types of input, positive and negative evidence proposed by the theory, are of equal relevance for both oral and written feedback. In the process of L2 learning, positive evidence provides information to learners on what is grammatically correct which leads to improving learners' motivation, whereas negative evidence in the form of WCF deals with the issues of what is grammatically incorrect. Though positive and negative evidence in oral feedback occur

simultaneously during the interaction process, they are typically delayed in written feedback and occur in the form of teacher commentary (Ferris et al., 1997) on learners' written output, aiming to provide feedback on different aspects of writing. Recent empirical studies (Frear & Chiu, 2015; Saricaoglu, 2019) were framed along the Interaction Hypothesis, suggesting that WCF provides opportunities for the provision of negative evidence.

To sum up, providing WCF could be pedagogically and theoretically facilitative of language learning. From a pedagogical perspective, providing written error correction is the most time-consuming practice of writing teachers. However, they believe in the facilitative role of their feedback provision and so do students who admit that they want their teachers to correct as many errors as possible regardless of whether they are related to linguistic features or content and organization of their written outputs (Lee, 2004). From the theoretical perspectives, both the Noticing Hypothesis and the Skill Acquisition Theory support the tenet that written feedback facilitates development. Within the framework of Noticing Hypothesis (Schmidt, 2010), error correction helps learners see differences between their interlanguage and the target language norms. Likewise, Skills Acquisition Theory predicts a facilitative effect for written error correction. It suggests that explicit instruction (i.e., providing written feedback) and extensive practice are crucial phases in the process of converting learners' declarative knowledge into procedural knowledge which may lead to automatization. Furthermore, the Interaction Hypothesis supports the idea that providing feedback is of essential value, and thus teachers should draw learners' attention to form in meaningful interactions. Even though meaningful interaction occurs primarily in oral feedback, it can also be pertinent to written feedback as two concepts of positive and negative evidence can be applied to both feedback modes.

2.1.3 Taxonomy of feedback

Feedback can be generally differentiated in terms of "its directness which ranges from direct (e.g., writing the correct form above the incorrect form) to indirect (e.g., using editing symbols to signal an error)" (Storch & Wigglesworth, 2010, p. 304). More specifically, Ellis (2009) divides WCF into six broad categories (Table 2.1), which constitute three major categories of WCF that have been used in many experimental studies. These categories are direct feedback, indirect feedback, and metalinguistic feedback.

Direct feedback is given through the provision of the correct form to the student. In this case, a number of variations such as crossing out an unnecessary word, phrase, or morpheme,

inserting a missing word or morpheme, and writing the correct form above or near to the incorrect form can be provided (Ellis, 2009b). The advantage of direct feedback is that it provides students with explicit guidance about how to correct their errors. Therefore, Ferris and Roberts (2001) suggested that direct feedback probably works well with students at low proficiency levels. However, the drawback of direct feedback is that it requires minimal processing on the part of the learner which may not contribute to long-term learning. In the case of indirect feedback, feedback is provided implicitly: an indication that the student has made an error, but direct correction is not offered. An indication of an error can take many forms such as underlying the errors or using cursors to show omissions in the student's text or placing a cross in the margin next to the line containing the error (Ellis, 2009b). Accordingly, students need to engage in deeper processing to figure out how the error should be corrected. For this reason, Ferris and Roberts (2001) suggested that it possibly encourages students to reflect on linguistic forms which can lead to long-term learning. Among the three main categories of WCF, metalinguistic feedback is the most explicit feedback type, as it involves providing students with some form of explicit comments about the nature of the errors (Ellis, 2009b) through the use of error codes or the provision of metalinguistic explanations of students' errors.

It is worth noting that Ferris (1999) distinguished between treatable errors and untreatable errors. Treatable errors "occur in a patterned, rule-governed way" and students can fix them when reviewing the rules governing these errors independently; examples of treatable errors include "verb form errors, subject-verb disagreement, missing articles, run-on sentences and comma splices" (Ferris, 1999, p. 6). Ferris (2006) suggested that employing indirect feedback is an effective way of dealing with treatable errors in which students' attention is drawn towards such errors, and they figure out how to correct the errors after reviewing the related grammatical rules. On the other hand, untreatable errors are idiosyncratic in terms of language rules, as no fixed rules can be assigned to correct these errors. These errors constitute "a wide variety of lexical errors and problems with sentence structure, word choice, including missing words, unnecessary words, and word order problems" (Ferris, 1999, p. 6). Ferris (2010) claimed that these untreatable errors include some that are crucial for communication and thus it is inadequate to simply underline these errors. Rather, direct feedback should be provided not to generate cognitive overload in dealing with untreatable errors. In other words, it is predicted that students at the lower proficiency levels might find it difficult to figure out how they could deal with untreatable errors which are related to sentence structure and word choice

if teachers provide indirect corrections, like indication of such errors. Therefore, direct correction should be used to provide input for learning correct forms of untreatable errors.

Another major treatment variable that has been considered in WCF studies is the focus of feedback: focused versus unfocused. The focused/unfocused distinction pertains to the number of linguistic structures that are targeted. Focused feedback (Ellis, 2009a) refers to feedback that targets a limited number of linguistic structures, whereas unfocused feedback (Ellis, 2009a) targets errors related to multiple structures. In the literature, many studies explored the effectiveness of focused feedback (Sheen, 2007; Suzuki et al., 2019) whereas few studies were designed to compare the effectiveness of focused and unfocused feedback (Ellis et al., 2008; Sheen et al., 2009). Precisely, the bulk of WCF studies was carried out to investigate the differential effectiveness of feedback types (i.e., direct, indirect, and metalinguistic feedback) and the extent to which these feedback types help improve writing accuracy and quality of writing, considering treatment variables (e.g., focus of feedback, type of feedback, and mode of feedback).

Another trend in WCF research literature explores the effectiveness of electronic or automated feedback, integrating the AWE tools such as *Criterion*® (https://criterion.ets.org/), *Writing To Learn* (https://www.pearsonassessments.com/), and *My Access* (http://www.vantagelearning.com/) on students' writing. Since the provision of individual feedback to learners' writing takes time and effort, incorporating AWE tools in the process of feedback provision saves time and teachers can focus on other global language concerns including meaning, idea development, and argumentation. One of the main advantages of automated feedback is their capacity to generate immediate feedback which can identify learners' strengths and weaknesses in terms of linguistic aspects of writing. Making use of this noticeable strength of automated feedback, the amount of time that teachers use for feedback provision can be shortened. Along the line of electronic feedback, many studies (Dikli, 2011; Dikli & Bleyle, 2014; El Ebyary & Windeatt, 2010; Kellogg et al., 2010; Li et al., 2015; Liao, 2016; Liu et al., 2017; Luo & Liu, 2017; Saricaoglu, 2019; Stevenson & Phakiti, 2014, 2019; Wang et al., 2020; Wilson et al., 2017; Wilson & Czik, 2016; Zhang, 2020) examined the effect of automated feedback on the quality of students' writing and how it helps students reduce language errors during the revision process. Among these previous research, a few studies (Dikli, 2011; Dikli & Bleyle, 2014) compared the nature of automated feedback with that of teacher written feedback. Findings from these studies suggested that automated feedback helps improve students' texts in the process of revision, although some limitations linger on inability to respond to texts' argumentation and communicative intent.

Turning to another strand of research, reformulation includes a native speaker rewriting the student's text in a way that preserves as many of the writer's ideas as possible while expressing them in their own words so that a written piece sounds as native as possible (Ellis, 2009a). Previous studies (Kim & Bowles, 2019; Sachs & Polio, 2007) compared the effects of direct correction and reformulation on students' revisions of their texts and suggested the potential of using reformulation as feedback in assisting students with sentential and paragraph-level errors rather than surface level linguistic errors. Kim and Bowles (2019) found that learners processed sentential and paragraph-level errors more deeply than surface-level linguistic errors when reformulation was provided. These findings led the authors to suggest that reformulation would be effective in dealing with higher order stylistic and organisational errors.

Table 2.1

Ellis's (2009a) Typology of Written Corrective Feedback (p.98)

Type of feedback	Description
1. Direct feedback	The teacher provides the learner with the correct form.
2. Indirect feedback	The teacher indicates that an error exists but does not provide the correction.
a. Indicating + locating the error	This takes the form of underlining and use of cursors to show omission in the learner's text.
b. Indication only	This takes the form of an indication in the margin that an error or errors have taken place in a line of text.
3. Metalinguistic feedback	The teacher provides some metalinguistic clue as to the nature of the error.
a. Use of error code	Teacher writes codes in the margin (e.g., ww = wrong word; art = article).
b. Brief grammatical descriptions	Teacher numbers errors in text and writes a grammatical description for each numbered error at the bottom of the text.
4. The focus of the feedback	This concerns whether the teacher attempts to correct all (or most) of the students' errors or selects one or two specific types of errors to correct. This distinction can be applied to each of the above options.
a. Unfocused feedback	Unfocused feedback is extensive.
b. Focused feedback	Focused feedback is intensive.
5. Electronic feedback	The teacher indicates an error and provides a hyperlink to a concordance file that contains examples of correct usage.

6. Reformulation	This consists of a native speaker's reworking of the students' entire text to make the language seem as native-like as possible while keeping the content of the original intact.

2.1.4 Theoretical frameworks for investigating written corrective feedback

WCF research has been driven by practical and pedagogical concerns rather than by theoretical considerations. Ferris (1999), for instance, noted that WCF researchers are interested in probing whether WCF helps L2 learners improve the overall accuracy of their writing. Despite having a substantial body of WCF research, studies differ in terms of the linguistic features, research designs, and the type of feedback. Therefore, a framework for identifying the variables that WCF research has entailed was developed by Ellis (2010) (Figure 2.1). The framework could be called a componential framework, as it constitutes different variables that should be considered in framing WCF studies. Understanding the components included in the framework can help feedback researchers consider factors they need to consider in their studies.

Figure 2.1.

Ellis's (2010) Componential Framework for Investigating Corrective Feedback (p.337)
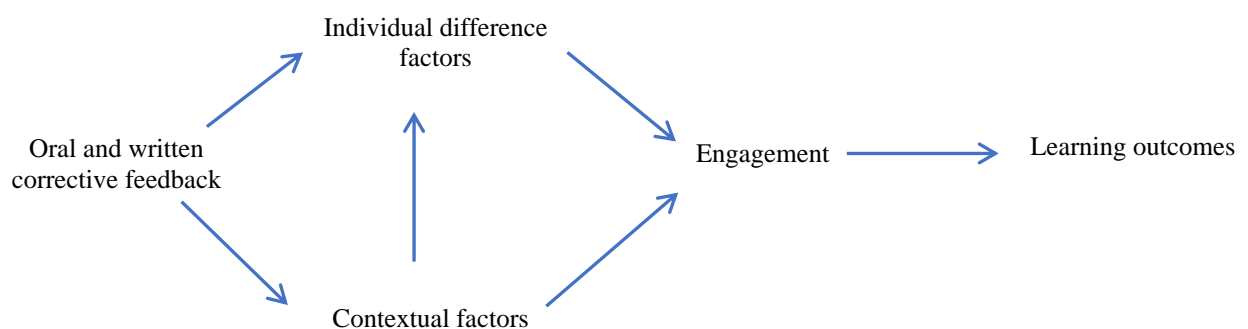
The framework comprises a number of factors that can impact the effects of feedback. Regardless of whether the study pursues the impact of oral or written feedback on learning outcomes, the roles of individual difference factors and other contextual factors need to be addressed as they may mediate the efficacy of feedback. WCF research targets a variety of individual differences, including motivation, language aptitude, working memory, learner beliefs, and attitudes towards written feedback for academic writing. Though the early research neglected these individual factors, recent studies (e.g., Benson & DeKeyser, 2018; Can & Walker, 2011; Goldstein, 2006; Li & Roshan, 2019; Li & Li, 2012; Orts & Salazar, 2016; Rummel & Bitchener, 2015) have paid attention to them and explored how these factors impact the efficacy of WCF. Also, contextual factors include macro factors (i.e., the setting in which learning takes place) and micro factors (i.e., the nature of the activity learners participate in

when they receive feedback). For example, the extent of learners' motivation might be lower or higher depending on both the general setting and their attitudes toward the specific activity in which they are engaged in. Therefore, Ellis (2010) claimed that "the individual difference factors interact with contextual factors to mediate between the feedback that learners receive and their engagement with the WCF and thereby influence learning outcomes" (p. 339). Moreover, student engagement with feedback (i.e., how learners respond to feedback they receive) has a significant effect on the effectiveness of feedback. In this regard, Ellis (2010, p. 342) suggested three perspectives that engagement can be examined from: "a cognitive perspective (where the focus is on how learners attend to the WCF they receive), a behavioural perspective (where the focus is on whether and in what way learners revise their written texts), and an affective perspective (where the focus is on how learners respond attitudinally to the WCF)".

In addition to the factors that WCF studies should take into considerations, other fundamental issues, instructional procedures, and methodological questions are indicated in Ferris's (2003) framework. The framework is a three-part model developed after making adaptions from questions and issues posed by previous research (e.g., Ferris, 1999; Polio, 1997) (Table 2.2). The framework is intended to be used as a checklist not only for assessing the validity of the findings of WCF studies but also for considering the variables that should be considered when designing future WCF studies.

Table 2.2

Ferris's (2003) Framework for Analysing and Designing Error Correction Studies (p. 45)

Part I: Basic Parameters
  ° Subject (students and teachers) characteristics: SL/FL, language majors or non-majors, L2 proficiency, background in writing (process vs. product), formal grammar knowledge
  ° Sample size (including the size of treatment groups into which subjects were divided)
  ° Duration of instructional treatment and /or data collection
Part II: Instructional Phenomenon
  ° Type of writing considered (e.g., free-writes or journal entries vs. multiple draft compositions; in-class vs. out-of-class)
  ° Larger instructional context: Were students given grammar instruction or resources for processing error feedback? Did they track their progress, and were they given increasing responsibility for self-editing?
  ° The nature of error feedback: Who provided it? What linguistic issues were addressed? What mechanisms (direct/indirect feedback, codes, etc.) were used for giving feedback?

Part III: Research design
- ° Was an appropriate quantitative design employed (control group, pre-test/post-test, accurate statistics, confounding variables accounted for)?
- ° Were multiple raters or coders used, were inter-rater reliabilities calculated and reported, and was it clear to what those reliability coefficients referred?

Unlike Ellis's (2010) componential framework which considers learners' individual difference factors and contextual factors separately, Evans et al. (2010) suggested three contextual variables in WCF research: learner, situation, and instructional methodology (p. 449). They proposed that these variables may either facilitate or hinder writing accuracy, claiming that any research on WCF must account for the potential effects of these contextual variables before drawing any conclusions about the efficacy of WCF. Learner variables entail everything that the student brings to the learning experience. Major learner differences that impact the success of WCF include learners' writing ability in L1 and L2, motivation, proficiency levels, attitudes, and perceptions. Situational variables consist of the teacher, the physical environment, and the learning atmosphere which "shapes the learning context beyond what can be attributed specifically to the learner or to the instructional methodology" (p. 450). These situational variables can be closely interrelated to the learner variables and influence one another.

The third component of the framework concerns the features of the specific design of instructions, especially what is taught and how it is taught. In the case of WCF, the factors such as what errors to correct (language or content), how to correct (direct or indirect), and how often these errors are corrected should be identified before conducting WCF studies. In terms of instructional methodology, Evans et al. (2010) developed an instructional strategy, also known as dynamic WCF, to help improve L2 learners' writing performance and it is conceptualized in skill acquisition theory (DeKeyser, 2007). The two principal characteristics of this methodology are that "feedback reflects what the individual learner needs most as demonstrated by what the learner produces", and "tasks and feedback are manageable, meaningful, timely, and constant for both the learner and teacher" (p. 452). Based on these guidelines, learners are asked to write for ten minutes every class session to ensure that tasks and feedback are manageable. Moreover, the focus of some earlier sessions of a class is on meaning-focused feedback, introducing learners with the basics of how to organize a coherent text, how to develop a main topic, and provide adequate support. During the whole process of writing, learners are encouraged to keep a tally of errors by type and revise their texts until their texts are linguistically accurate.

2.1.5 Differing views on the efficacy of written corrective feedback

The debate over WCF has continued since Truscott (1996, 2001, 2004, 2007) made a provocative statement: grammar correction in L2 writing classes should be abandoned due to the ineffectiveness of the correction (Truscott, 1996). Based on the findings of previous studies (e.g., Kepner, 1991; Knoblauch & Brannon, 1981; Polio et al., 1998) which reported little facilitation of feedback on students' writing, he argued that grammar correction is not useful for either theoretical or practical reasons. As for the theoretical reasons, he argued that learning is not simply the transfer of information from the teacher to the learner, as many language teachers believed; therefore, providing feedback to grammatical structures without any understanding of the complex process of interlanguage development is ineffective (Truscott, 1996). In this respect, he stressed the teachability/learnability hypothesis of Pienemann (1989), i.e., the hypothesis that certain aspects of grammar should be taught in the order in which they are learnable. Truscott (2010) argued that learning involves a complex and unconscious system in students' minds; therefore, language teachers should not make intuitive judgments about what should happen when WCF is provided, taking into account of students' readiness to process the targeted linguistic structures. On the other hand, for practical considerations, he doubted whether L2 language teachers are able to provide consistent and systematic feedback on students' errors. Additionally, he argued against grammar correction because it takes time and energy to do additional writing practice. Instead of grammar correction, he suggested that extensive experience with the target language – experience in reading and writing should be provided to students in writing classes to improve writing.

Ferris (1999, 2004, 2010) evaluated Truscott's (1996) arguments as a rebuttal to his controversial statements. In her argumentation, she underlined a substantial amount of research evidence that proved the effectiveness of error correction: she pointed out some studies that confirmed the positive effects of WCF (e.g., Ashwell, 2000; Ellis, 1998; Ferris & Roberts, 2001). Also, she noted that well-designed research should be implemented before drawing any conclusions about the (in)effectiveness of WCF. Moreover, she argued that the problems in early WCF research were related to research design: in some research, no control groups or no pre-tests were used before the treatments (Ferris, 2010), whereas some studies failed to define which types of students' errors received WCF. In contrast to Truscott's (2007) claim, findings of other WCF studies (e.g., Ashwell, 2000; Chandler, 2003; Ferris, 2006; Ferris & Roberts, 2001) suggested that WCF can facilitate L2 development and help students improve the accuracy of their writing, at least for the particular linguistic features under considerations.

2.2 Literature review

Correcting learner errors through written feedback has long been part of the pedagogical traditions that writing teachers practice and is of widespread interest to L2 writing research. Due to the multifaceted nature of the construct of writing, feedback on students' writing includes "a wide variety of responses and may contain information regarding the accuracy, communicative success, or content of learner utterances or discourse" (Leeman, 2010, p. 112). Pedagogically, feedback links assessment to teaching and learning. It reflects information about learners' actual performance and guidance on future learning goals. In light of this premise, exploring the effectiveness of feedback has received significant attention in the past few decades.

As a field of inquiry, research on written feedback has been examined in the light of theoretical and methodological perspectives from L2 writing or SLA (Ferris, 2010). While WCF within the field of L2 writing is perceived as a means of helping learners improve the overall quality of their texts, SLA-oriented perspectives view WCF as a potential contributor to their language development. Informed by these perspectives, several studies have explored the effectiveness of WCF either on the overall accuracy and quality of texts or on some pre-selected linguistic features. Particularly, the majority of studies examined the role of WCF in writing classrooms (Ferris, 2006; Truscott & Hsu, 2008) through comparing the relative effectiveness of explicit and implicit feedback types (Bitchener, 2008; Ellis et al., 2008). These studies (e.g., Bitchener, 2008; Bitchener & Knoch, 2010b; Ferris, 2014) provided evidence that WCF helps improve accuracy development in students' texts.

The mushrooming of empirical studies has led to numerous meta-analyses and review papers synthesizing the findings. For example, Truscott (2007) looked at the impact of error correction in twelve studies and concluded that correction had a harmful effect on students' ability to write accurately. Another meta-analysis conducted by Biber et al. (2011) examined 25 studies published between 1982 and 2007 and suggested that feedback resulted in accuracy gains in writing development. Kang and Han (2015) later identified 21 primary studies (from 1980 to 2013) and concluded that WCF led to greater grammatical accuracy in L2 writing. More recently, Sia and Cheung (2017) conducted a qualitative synthesis of 68 empirical studies published in journals from 2006 to 2016 and shed light on the role of individual differences in the effectiveness of WCF.

Karim and Nassaji (2019) presented a critical synthesis of research on WCF and its effects on L2 learning over the past four decades. The authors addressed major issues in WCF research

including the effects of different types of feedback and those of focused and unfocused feedback. Moreover, they shed light on a practical limitation of studies which focused on a single or limited number of preselected errors and urged future studies to examine the long-term effect of unfocused feedback. Li and Vuono (2019) reviewed 25 years of research on WCF in *System* and synthesised the findings of selected studies regarding the effects of WCF in facilitating learning gains and the factors that constrain its effectiveness. The most recent meta-analysis was conducted by Lim and Renandya (2020) and it included 33 studies and five Ph.D. and Master's dissertations published between 2001 and 2019. The results of these reviews and meta-analyses suggested that WCF has the potential to improve grammatical accuracy and is conducive to writing development, except the findings by Truscott (2007).

Unlike other meta analyses, Liu and Brown (2015) conducted a methodological synthesis of research on the efficacy of corrective feedback. Their findings revealed some methodological limitations, including inadequate reporting of research context, methodology, and statistical analysis, and the use of mixed feedback types, making it impossible to separate the effectiveness of individual feedback. Also, the authors noted that these methodological variations rendered it difficult to compare the findings across studies.

With these issues in mind, this review adds to the existing literature by investigating the four key variables in WCF research: research designs, feedback treatments, writing tasks, and accuracy measures. The selection of these four variables were informed by the previous studies which suggested that these factors tend to mitigate the effect of written feedback. For example, the extent to which students benefit from written feedback is likely to depend on feedback-related factors (e.g., the scope of feedback, the type of feedback, and the source of feedback) (Kim et al., 2020; Yang et al., 2006), writing-task-related factors (e.g., timed/untimed writing, and genres of writing) (Riazantseva, 2012; Shi, 2021), and how accuracy is measured as an indicator of writing development (Liu & Brown, 2015). Moreover, the lack of control groups and variations in research design make it difficult to compare results across studies. Therefore, the present review aims to illuminate the importance of achieving consistency in some of these aspects to facilitate the comparison of research findings. Accordingly, four research questions were addressed:

RQ1. What differences are present in WCF studies in terms of research design features (e.g., treatment session, duration of the intervention, and inclusion of comparison group)?

RQ2. What differences are present in WCF studies in terms of feedback-related features (e.g., feedback type, feedback focus, and feedback mode)?

RQ3. What differences are present in WCF studies in terms of writing task-related features (e.g., genres, provision of prompts, and length of writing)?

RQ4. What are the common outcome accuracy measures used in WCF studies and what are the advantages and pitfalls of these measures?

The next section provides a brief outline of the targeted variables before establishing a set of inclusion and exclusion criteria for the selection of studies. The selection criteria resulted in the identification of 42 empirical studies published in peer-reviewed journals between 2001 and 2021. The section concludes with a discussion of the main properties of the selected studies and the key findings of the review.

2.2.1 Variables in focus: Research designs, feedback treatments, writing tasks, and linguistic accuracy

2.2.1.1 Research designs in written corrective feedback

Even though the bulk of feedback research was designed to investigate the efficacy of WCF, research designs in such studies vary from one another. As Ferris (2004) stated, WCF studies are incomparable due to inconsistencies in design, as they varied on every research parameter: subject characteristics, size of samples and treatment groups, duration of feedback treatments, types of text considered, types of feedback given, who provided the feedback, and how accuracy was measured. Moreover, Liu and Brown's (2015) methodological synthesis also revealed a number of methodological limitations and inconsistencies in the literature including the research designs. Particularly, the results indicated issues such as lack of control groups (Chandler, 2003; Evans et al., 2010; Fazio, 2001), focus on revised texts rather than new pieces of writing (Ashwell, 2000; Ferris & Roberts, 2001), and incomparability due to inconsistent treatments and accuracy measures (Ellis et al., 2008; Van Beuningen et al., 2008). Specifically, even when comparison groups are included in the studies, how these groups are operationalized differs from one study to another: the control groups in some studies receive traditional grammar instruction (Kurzer, 2018) or do self-correction (Bonilla Lopez et al., 2018), whereas those from other studies receive general feedback on content (e.g., Benson & DeKeyser, 2018).

Looking at the studies that explore the effectiveness of WCF, two trends of research can be categorized: (i) studies that focus on the role of WCF during the revision process (i.e., studies that considered the improved accuracy of texts by comparing the initial version and the revised version), and (ii) studies that investigate the effectiveness of WCF by comparing the accuracy of the first text with that of the new text. In this regard, Truscott and Hsu (2008) contended that

the first trend of research does not constitute evidence of learning, as two versions of the same text are compared to measure the accuracy development relying on the WCF provided. Therefore, it is questionable whether WCF results in the acquisition of the corrected forms. Based on their claim, current WCF studies investigate the effects of different strategies of WCF on linguistic accuracy of students' texts: they include a new writing task which is supposed to be used in an immediate post-test and delayed post-tests after the revision process.

2.2.1.2 Feedback treatments in written corrective feedback

In L2 writing, scholars and teachers have stressed the importance of written feedback in developing students' writing abilities. Based on the dichotomy between feedback on form and content, written feedback could be classified into corrective and non-corrective feedback (Luo & Liu, 2017). Corrective feedback promotes the learning of the target language by providing negative evidence and non-corrective feedback scaffolds writing in aspects of content, organization, linguistic performance, and format. In other words, corrective feedback focuses on developing students' accuracy, whereas non-corrective feedback provides commentary to rhetorical and content issues (Goldstein, 2006, 2004). Compared to research investigating the influence of written commentary feedback on students' revision and future texts, there has been increased interest to determine the role of different types of WCF in L2 writing. As suggested by Storch and Wigglesworth (2010), corrective feedback can be differentiated based on "its directness which ranges from direct (e.g., writing the correct form above the incorrect form) to indirect (e.g., using editing symbols to signal an error)" (p. 304). With reference to empirical studies on WCF, Ellis (2009b) identified three major strategies for providing feedback: direct, indirect, and metalinguistic feedback. Direct feedback is given through the provision of the correct target language form. Indirect feedback is provided implicitly, by an indication that an error has been committed. Metalinguistic feedback provides students with some form of explicit explanations about errors (Ellis, 2009b). Accordingly, empirical studies have investigated the facilitative role of feedback either by comparing specific feedback strategies across no-feedback conditions (e.g., Kurzer, 2018; Truscott & Hsu, 2008) or by comparing the relative effectiveness of two or more feedback strategies (e.g., Mirzaii & Aliabadi, 2013; Riazantseva, 2012).

Further distinctions in feedback treatments concern the scope of feedback (comprehensive versus focused) – i.e. "the amount of WCF teachers should give to students – whether to respond to all written errors or to respond to them in a selective or focused manner" (Mao &

Lee, 2020, p. 1). Feedback scope is a pedagogically important issue in WCF research, as it has direct relevance to the work of teachers. According to Mao and Lee (2020), teachers can be better informed about the extent to which they should respond to errors in students' writing when they apply research findings into their feedback practice. Earlier studies (e.g., Han & Hyland, 2015; Lee, 2004, 2005) investigated the nature of teachers' written feedback practices in classrooms and found that both teachers and students preferred comprehensive error correction. However, studies in WCF research have put emphasis on the utilization of focused feedback (e.g., Benson & DeKeyser, 2018; Stefanou & Révész, 2015), suggesting that responding to errors in a focused manner is more beneficial than responding to all errors in an unfocused manner. Other recent studies have used a comprehensive approach (e.g., Bonilla Lopez et. al., 2017; Van Beuningen et. al., 2012) and offered feedback on diverse error types rather than on errors of a single type.

2.2.1.3 Writing tasks in second language writing

As Ellis (2009c) posited, the primary focus of a task in language learning should be on meaning with a clearly defined outcome and learners should rely on their linguistic or non-linguistic resources to complete a task. Tasks can either be unfocused or focused (p. 223) based on a distinction whether a task requires learners to use language in general or use specific linguistic features. In L2 writing, both unfocused (e.g., essay) or focused writing tasks (e.g., grammatical structure) are used to assess learners' L2 writing proficiency (Ellis, 2010). In the case of unfocused writing tasks, understanding how task demands impact variations in the quality and quantity of L2 writing plays a part in eliciting specific levels of L2 performance, as "tasks provide a context for negotiating and comprehending the meaning of language provided in task input" and "tasks provide opportunities for uptake of (implicit or explicit) corrective feedback on a participant's production" (Robinson, 2011, p. 4). Adopting Robinson's cognition hypothesis, Kuiken and Vedder (2008), for instance, examined the effect of a writing prompt on complexity and accuracy. They concluded that texts written in response to more cognitively demanding tasks turned out to be more accurate, with lower error ratios per T-unit, than cognitively less demanding tasks. These findings informed L2 researchers how interactions between the genres and cognitive demands of the writing tasks impact students' writing accuracy. In WCF research, how cognitive demands of writing tasks affect learners' accuracy has received relatively scant attention. For example, Riazantseva (2012) examined the effects of WCF along three outcome measures of writing performance: in-class essays, in-class

summaries, and at-home summaries which differed in terms of cognitive and linguistic demands. The findings suggested that these outcome measures produced different estimates of L2 writing accuracy.

Task-related factors including task types and task complexity are supposed to impact the score reliability of students' writing in both high-stakes and classroom assessment contexts (Liu & Huang, 2020). In terms of genres of writing tasks, empirical investigations (e.g., Kuiken & Vedder, 2008; Polio & Yoon, 2018; Yoon & Polio, 2017) provided evidence that different genres have different communicative or functional requirements that may result in different language use. For example, Yoon and Polio (2017) posited that more complex language can occur in argumentative essays because they have higher reasoning demands than narrative texts. Similarly, Polio and Yoon (2018) found that the functional requirements for narrative and argumentative writing are different; thus, the two genres require different language. These findings help deepen the understanding about the impact of diverse writing tasks on learners' linguistic performance (e.g., accuracy and complexity).

2.2.1.4 Linguistic accuracy as the outcome measure in written corrective feedback

Linguistic accuracy which has been defined as "the ability to be free from errors while using language to communicate in either writing or speech" (Wolfe-Quintero et al., 1998, p. 33) is a relevant construct for research in L2 writing assessment and pedagogy. As noted by Ferris (2006), "accuracy in writing matters to academic and professional audiences" (p. 81). Polio and Shea (2014) enumerated five reasons for measuring accuracy in L2 writing research: to investigate (i) the effects of WCF, (ii) the effects of planning, (iii) the effect of task complexity, (iv) the difference between individual and collaborative writing, and (v) change over time.

Specifically to WCF research, attention given to linguistic accuracy in writing classes has been a reason of measuring accuracy gains and both L2 writing teachers and students have agreed that written accuracy is expected in academic writing. Lee (2008), for instance, found that 94.1% of teacher feedback focused on form, 3.8% on content, 0.4% on organization, and 1.7% on other aspects when investigating the feedback practices of teachers in Hong Kong secondary English classes. In a similar vein, the review by Liu and Brown (2015) reported that 36% of empirical studies provided feedback solely on grammar, 18% included feedback focusing on both grammatical and lexical errors, and 27% provided extensive feedback, whereas other studies did not specify the focus of feedback. Other possible reasons for providing feedback on language-related errors rather than on content-related issues may

concern their beliefs (what teachers assume students can deal with) and students' level of proficiency. They may also be impacted by contextual factors including time constraints, teacher workload, and large class sizes (Mao & Crosthwaite, 2019).

In the WCF literature, students' accuracy has been the key dependent measure used to assess the effects of feedback. As Nicolas-Conesa et al. (2019) stated, a distinction was made between feedback for accuracy and feedback for acquisition with reference to a dichotomy between uptake (i.e., errors successfully corrected in rewritten texts) and retention (i.e., reduction in error-making over time). In other words, whereas feedback for accuracy concerns how the provision of feedback helps improve learners' accuracy shortly after processing it, feedback for acquisition favours "long-term language learning by involving students in feedback processing, detection of errors, self-reflection on errors, and new output" (p. 849). Though earlier studies (e.g., Ferris & Roberts, 2001) measured accuracy gains by comparing the accuracy of students' first drafts and revised texts, Truscott and Hsu (2008) claimed that accuracy gains in learners' rewritten texts failed to provide evidence that feedback provision is beneficial for acquisition. Bearing this claim in mind, recent empirical studies (e.g., Sheen et. al., 2009; Van Beuningen et al., 2012) included new pieces of writing in their research designs and compared outcome accuracy developments in both revised and new texts.

2.2.2 Method of the literature review

This section overviews how data searches were carried out along five criteria for inclusion and discusses the data analysis procedure. I followed the guidelines of the Preferred Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Moher et al., 2009) to ensure that the review is systematic. First, inclusion/exclusion criteria were established, and relevant studies were identified through electronic and hand searches. Then, a coding scheme was developed drawing on a framework for analysing error correction studies (Ferris, 2003). Lastly, detailed analyses were conducted followed by synthesizing and interpreting the findings.

Peer-reviewed journal articles discussing the effects of WCF on L2 written accuracy development were retrieved from the electronic databases of the Educational Resources Information Centre (ERIC), ScienceDirect, Scopus, and from the Google academic search engine (http://scholar.google.com) using key terms including *written corrective feedback*, *comprehensive corrective feedback*, and *feedback in second or foreign language writing*. In addition to electronic searches, a hand search in the key journals on L2 writing was conducted to ensure that all relevant empirical studies were identified. Journals included *Journal of*

Second Language Writing, Language Teaching Research, System, Assessing Writing, TESOL Quarterly, Modern Language Journal, Language Learning, and *Language Teaching.*

The literature search covered studies published from 2000 to 2021 (January) and the initial search in the databases resulted in a pool of thousands of journal articles, book chapters, books, and review articles. Due to a bulk of empirical studies examining the effects of WCF on linguistic accuracy, I included primary studies along these six criteria: (i) WCF must be the focus of the study, (ii) it must explicitly describe methodological considerations, (iii) it must consider text samples that include the production of either revised or new texts, (iv) it must utilize unfocused writing tasks in which students are allowed to produce language with meaningful communication (Ellis, 2009c; Norris & Ortega, 2000), (v) WCF must be provided by a teacher and/or a researcher, and (vi) it must be written in English.

A study was excluded if it (i) focused mainly on learners' beliefs and engagement with WCF (e.g., Han, 2017; Han & Hyland, 2015), (ii) considered the effectiveness of peer feedback or automated feedback (e.g., Luo & Liu, 2017), (iii) focused exclusively on learners' perceptions and how individual differences mediate the effectiveness of WCF (e.g., Park et al., 2016) without assessing accuracy gains, and (iv) concerned how learners process written feedback (e.g., Kim & Bowles, 2019). The database search delivered more than 20,000 references and hand searches added 50 studies. After removing duplicates and articles that did not satisfy the inclusion criteria, the review yielded a sample of 42 studies which were then judged for their quality and relevance.

Data analysis was conducted in three iterative phases. First, the 42 studies were carefully read to identify their theoretical perspectives. Second, different aspects of feedback treatments, writing tasks, and accuracy measures were reviewed to ensure that they were not left out in the coding scheme. Third, a draft coding scheme was developed to organise all relevant information related to each study. The development of the coding scheme was guided by a framework for analysing error correction studies (Ferris, 2003). The coding scheme comprises four categories: (i) research designs, (ii) feedback treatments, (iii) writing tasks, and (iv) measures for linguistic accuracy (Table 2.3). These were further divided into 18 sub-categories after identifying contributing variables in the dataset. After the coding scheme was established, the selected studies were categorised in the scheme.

Table 2.3

Coding Scheme of Four Categories and 18 Sub-categories

Categories	Variables
Research designs	Research methods: single method (quantitative), mixed methods (quantitative + qualitative)
	Research designs: pre-test post-test design, pre-test-immediate post-test-delayed post-test design, comparison among drafts, comparison between two classes, random assignment, pre-test-treatment-edition
	Treatment session: open-ended, not applicable, not stated
	Duration of the intervention: open-ended, not stated
	Inclusion of comparison group: yes, no
Feedback treatments	Feedback type: direct feedback, indirect feedback, comparison of direct and indirect feedback, comparison of direct and metalinguistic feedback, comparison of different types of indirect feedback, etc.
	Feedback focus: article, preposition, past tense, hypothetical condition, multiple sentence level issues (organization, paragraphing, cohesion, relevance), a wide range of grammatical errors, and a wide range of grammatical and lexical errors, verb tense choice
	Feedback mode: teacher, researcher, teacher-researcher, not stated
	Feedback scope: focused feedback, unfocused feedback, comparison of focused and unfocused feedback, not stated
Writing tasks	Genre: narrative, argumentative, autobiography, picture-composition, dictogloss, etc.
	Time: timed writing, untimed writing, both timed and untimed writing, not stated
	Provision of prompts: yes (pictures, pictures plus prompts, reading text, listening task), no, not applicable, not stated
	No. of writing during the intervention: open-ended
	Length of writing: open-ended
	Revision: yes, no
	New writing tasks: yes, no
Measures of linguistic accuracy	Measures: error ratio, error-free clause ratio, error-free T-unit ratio, holistic scoring, obligatory occasion analysis, etc.
	Reliability: inter-, intra-, both inter-and-intra, no

2.2.3 Surface properties of the 42 selected empirical studies

Appendix A presents the general characteristics of the selected studies. A closer examination of the findings revealed that a large proportion of studies (93%) targeted adult learners enrolled in general English classes or in academic writing classes. Seven percent targeted teenagers between 11 and 17. Concerning the study contexts, the results posit that some studies (Ashwell,

2000; Kurzer, 2018; Shintani & Ellis, 2013; Truscott & Hsu, 2008) were conducted in academic writing classes whereas the others (Ellis et al., 2008; Shintani et al., 2014; Shintani & Ellis, 2015) took place in general English classes where instruction is equally distributed among all the language skills rather than solely on developing the writing skills. What should be noted is that the difference between general English classes and academic writing classes might determine variations in the amount of instruction on writing skills and that of feedback learners receive during the courses. For instance, the study by Shintani and Ellis (2013) recruited participants from academic writing classes in an intensive English language programme in the US, whereas Ellis et al.'s (2008) study recruited the students from general English classes in an EFL context. In this case, the possible differences in the amount of input, attention paid to improve their L2 writing, and different study contexts (i.e., ESL/EFL) ought to be some key variables which could influence the effectiveness of feedback even if no explicit attention was paid to the targeted linguistic features during the intervention.

In relation to participants' L2 proficiency level, approximately half of the studies (45%) recruited learners at high and low intermediate level of L2 proficiency. Possibly, the underlying reason is that students at the intermediate level tend to make a range of grammatical and lexical errors in writing, compared to elementary and advanced learners. Other researchers (e.g., Stefanou & Révész, 2015) explicitly posited that recruiting participants with intermediate proficiency increased the comparability of their research to previous studies which targeted learners at intermediate levels. Some studies (e.g., Bonilla Lopez et al., 2017) recruited students with low and high proficiency levels and found that WCF effectively enhanced students' grammatical accuracy regardless of their L2 proficiency.

Sample sizes ranged from 27 to 325 with single or multiple language backgrounds. In particular, 59% of the dataset recruited participants from multiple language groups, whereas 36% came from single groups; 5% did not report the students' language background. Most of the 42 studies (88%) were conducted in educational contexts where English was taught either as a second or a foreign language. A further imbalance was found with respect to the studies' language contexts. Though WCF research was conducted worldwide, 18 studies were done in US contexts and 24 inquires in EFL contexts such as Japan, Korea, Laos, and Spain. Furthermore, 69% were done in university settings, with considerably fewer studies (9.5%) (e.g., Fazio, 2001; Van Beuningen et al., 2008; Van Beuningen, De Jong, & Kuiken, 2012) in other educational settings including primary and high schools. Therefore, most studies examined how university students respond to WCF and provided less evidence on how younger learners act upon teacher feedback.

2.2.4 RQ1: Research design features in written corrective feedback studies

Concerning the research design features in WCF studies, the review examines five sub-categories. The overall findings revealed considerable divergence in these variables: research methods, research designs, the number of treatment sessions, duration of interventions, and the inclusion of a comparison group posed serious challenges to synthesizing results across many empirical studies. Table 2.4 indicates that studies examining the effectiveness of WCF reported a wide range of variations in research design features. As for the methods used, single quantitative methods ($n = 30$) were far more typically used than mixed methods ($n = 12$), including additional instruments like interview sessions, and retrospective self-reports for the sake of data triangulation. Regarding research designs, a pre-test-post-test-delayed post-test experimental design is the most predominant research design in WCF research: it accounted for 45% of the selected studies. This research trend paves the way for examining the potential long-term effectiveness of feedback provision. However, the number of studies without including any delayed post-tests were 43% of the total studies with a single treatment session. Only 19% of the studies assigned the participants in experimental groups randomly, as most studies were conducted in intact classes where random assignments do not seem feasible.

Regarding feedback treatments, significant differences were observed in 15 of the 42 studies (36%) involving a single feedback treatment session. Nevertheless, the studies with two to three feedback treatments represented 14% of the 42 studies. These variations in the number of feedback treatments within the range of a single treatment to twelve treatments determine a considerable impact on the extent to which WCF is beneficial. It is a fact that WCF studies with one-shot designs are methodologically easier to implement. Yet, their contributions to language learning may be limited, as the time span during which feedback is provided is short to draw a conclusion about the effectiveness of WCF. Furthermore, the present review reveals that recent WCF studies ($n = 6$) made use of a dynamic WCF approach (e.g., Evans et al., 2010; Hartshorn & Evans, 2015; Kurzer, 2018): they included regular treatment sessions (Section 2.1.4). The main difference between WCF and dynamic WCF studies is that the studies that employ dynamic WCF approach (e.g., Hartshorn et al., 2010) need to ensure that the feedback is *"meaningful, timely, constant, and manageable"* for both teachers and students (p.87), whereas these aspects are rarely considered in traditional WCF studies in particular.

Another issue arising from research design features concerns the inconsistency in the duration of the intervention within the range of three to 30 weeks. Table 2.4 illustrates the distribution of the studies based on the duration of each study. The wide range of variations in

the timespan in which WCF studies were conducted might also be a key indicator of differential effectiveness of WCF studies. Moreover, further differences emerge when examining the inclusion of a comparison group and these differences relate to the different nature of operationalizing the control group. It is a common practice for WCF research to include a comparison group in research design, as the results pinpoint that 86% of the studies included comparison groups. Despite an absence of control groups in earlier studies on WCF, the present review indicates that the majority of studies include comparison groups, indicating a positive trend. The studies without control groups (e.g., Chandler, 2003; Evans et al., 2010; Fazio, 2001; Ferris, 2006) accounted for 14 % of the database. It is worth noting that the divergence in the way how comparison groups are operationalized might mediate the efficacy of WCF provided and render the interpretation of research findings difficult. Taking ethical considerations into account, comparison groups in some studies receive traditional grammar instruction (e.g., Kurzer, 2018) whereas those in other studies receive general feedback on content (e.g., Benson & DeKeyser, 2018). However, other studies included control groups without any feedback (e.g., Adams, 2003; Shintani & Ellis, 2013). These variations in how comparison groups are operationalized make the findings across WCF studies incomparable.

Table 2.4

Research Design Features (n = 42)

Variables	Levels	No. of studies	%
Research methods	Single method (quantitative)	30	71
	Mixed methods (quantitative + qualitative)	12	29
Research design	Pre-test-post-test design	18	43
	Pre-test-immediate post-test-delayed post-test design	19	45
	Comparison among drafts	3	7
	Comparison between two classes	1	2
	Random assignment	8	19
	Pre-test-treatment-edition	1	2
Treatment sessions	One	15	36
	Two	6	14
	Three	6	14
	Four	1	2
	Five	2	5
	Seven	1	2

	Twelve	2	5
	Many treatments	6	14
	Not applicable	1	2
	Not stated	2	5
Duration of intervention	2 weeks	1	2
	3 weeks	2	5
	4 weeks	4	10
	5 weeks	2	5
	6 weeks	5	12
	7 weeks	1	2
	8 weeks	5	12
	9 weeks	2	5
	10 weeks	5	12
	12 weeks	2	5
	13 weeks	1	2
	15 weeks	4	10
	16 weeks	1	2
	30 weeks	1	2
	5 months	1	2
	10 months	2	5
	Not stated	3	7
Inclusion of a comparison group	Yes	36	86
	No	6	14

2.2.5 RQ2: Feedback-related features in written corrective feedback studies

As shown in Table 2.5, three feedback types were identified in the 42 studies: (i) direct feedback, (ii) indirect feedback, and (iii) metalinguistic explanation. These studies investigated how WCF functions through comparing feedback with no feedback conditions or by comparing different feedback strategies. The most frequently applied design compared the differential effectiveness of direct and metalinguistic feedback (29%), followed by indirect feedback (21%), the comparison of direct and indirect feedback (14%) and different direct feedback types (14%). As feedback types are regarded crucial in influencing the effects of WCF, controversies relating to this factor need special attention.

Table 2.5

Feedback-related Features

Variables	Levels	No. of studies	%
Feedback type	Comparison of direct and metalinguistic	12	29
	explanation	9	21
	Indirect feedback	6	14
	Comparison of direct and indirect feedback	6	14
	Comparison of different direct feedback types	4	10
	Comparison of direct, indirect, and		
	metalinguistic explanation	2	5
	Direct feedback	2	5
	Comparison of different indirect feedback types	1	2
	Indirect feedback followed by direct feedback		
Feedback focus	A wide range of grammatical features	15	32
	A wide range of grammatical and lexical errors	9	19
	Multiple sentence level issues (organization,		
	paragraphing, cohesion, relevance)	9	19
	Article		
	Past tense	7	15
	Hypothetical condition	3	7
	Preposition	2	4
	Verb tense choice	1	2
		1	2
Feedback mode	Researcher	18	43
	Teacher	13	31
	Teacher-Researcher	8	19
	Not stated	3	7
Feedback scope	Comprehensive feedback	21	50
	Focused feedback	17	40
	Comparison of focused and unfocused feedback	3	7
	Not stated	1	2

I found that combining feedback strategies complicates the operation of a single feedback type. For example, direct feedback alone can be turned into many feedback types such as direct focused feedback, direct unfocused feedback, direct feedback with metalinguistic explanation, direct feedback with revision, direct feedback followed by individual conference. These variations render it difficult to compare and generalize the findings across studies. Discrepancies in feedback types still exist even when a similar feedback type is provided. For example, though studies employed metalinguistic feedback in a similar vein, the students in

Shintani and Ellis's (2013) study received handouts with the rule of the targeted linguistic structures, whereas participants in Benson and DeKeyser's (2018) study had their errors marked and received consistently worded metalinguistic comments in the form of brief grammar rules on their Microsoft Word documents. The trivial differences in how similar feedback is provided can contribute to the effectiveness of WCF to some extent.

With respect to feedback focus, 32% of the dataset (e.g., Benson & DeKeyser, 2018; Bitchener et al., 2005; Van Beuningen et al., 2008) provided feedback on linguistic aspects of writing such as verb tense, verb form, articles, singular-plural, and subject-verb agreement. However, the linguistic features targeted in the studies were not well-balanced: seven studies focused solely on two functions of English article systems. I found that 19% provided feedback on grammatical and lexical errors (e.g., Mawlawi Diab, 2015; Riazantseva, 2012) or on form and multiple sentence-level issues such as organization, paragraphing, cohesion, and relevance (e.g., Ashwell, 2000; Chandler, 2003; Evans et al., 2010; Fazio, 2001; Ferris, 2006). Despite the prevalent emphasis on how WCF helps improve linguistic accuracy of L2 texts, few studies (e.g., Hartshorn & Evans, 2015; Hartshorn et al., 2010; Van Beuningen et al., 2012) examined how WCF influenced other aspects of writing (e.g., writing fluency, writing complexity, lexical diversity, and rhetorical appropriateness) in addition to measuring accuracy.

In connection with feedback source, researchers were the predominant source of feedback (e.g., Benson & DeKeyser, 2018; Bonilla Lopez et al., 2017, 2018; Sheen et al., 2009; Stefanou & Révész, 2015) due to logistical and methodological reasons. In particular, 43% of the studies stated that researchers provided feedback to ensure the consistency and to avoid influencing the results, whereas teachers provided feedback in 31% of the studies (e.g., Evans et al., 2010; Ferris, 2006; Rummel & Bitchener, 2015; Vyatkina, 2010). In the latter cases, there are additional variations: either classroom teachers or other teachers who did not teach the target classes provided WCF. In Fazio's (2001) study, for instance, a francophone elementary teacher who was not one of the classroom teachers offered feedback on students' texts to minimize variability in feedback quality due to teacher effects and to strengthen the research design. In contrast, in other studies (e.g., Ferris, 2006; Ferris & Roberts, 2001; Hartshorn & Evans, 2015; Vyatkina, 2010), multiple teachers gave feedback in their intact classes. Overall, 19% of the studies claimed that the instructor was one of the researchers. Though no previous studies have considered how the source of feedback (i.e., either from a researcher or a teacher) influences the efficacy of feedback, I would reason that it is an important variable: students' motivation and their engagement with feedback may be higher if they receive it from their teachers, whereas they may attend to feedback less if they know it was provided by a researcher.

Moreover, the quality and quantity of teachers' feedback provided may be more tuned to their students' needs. Thus, Liu and Brown (2015) suggested that training should be provided when teachers provide feedback in a study to better control the variations resulting from the source of feedback.

In terms of feedback scope, half of the studies (e.g., Ashwell, 2000; Chandler, 2003; Truscott & Hsu, 2008; Van Beuningen et al., 2008) used comprehensive feedback, whereas 40% (e.g., Mawlawi Diab, 2015; Shintani & Ellis, 2013, 2015; Shintani et al., 2014) applied focused feedback. Here, the pros and cons of focused and unfocused approaches deserve explicit discussion, as both have distinctive advantages and disadvantages. The focused approach has the advantage of yielding a greater effect of WCF within a time frame. For instance, Zhang (2021) noted that focused feedback is most effective as it allowed students to pay attention to only one or a limited number of linguistic types at a time. As a result of less attentional strain on students, Sheen (2007) also emphasised that students are more likely to process focused feedback and make effective use of it. However, this approach reduces ecological validity, as it does not seem to represent feedback practices used in authentic classroom contexts (Van Beuningen et al., 2012). Drawbacks include the difficulty in selecting target linguistic features based on individual students' needs (Lee et al., 2021) and if a single error type is considered, it is unnecessary to use direct testing of writing. Instead, grammar exercises focusing on specific language features can be utilized for this purpose. Moreover, the absence of obligatory occasions of the targeted error types in students' subsequent writing might impact the process of measuring accuracy gains.

Unlike focused feedback, comprehensive feedback tends to be more compatible with classroom practices due to unlimited foci on error categories regardless of whether they are related to form- or meaning-focused aspects of writing. However, some issues in comprehensive feedback studies resulted from a wide range of error types that they targeted. For example, as comprehensive feedback targets many aspects of writing, challenges relating to how to deal with almost all aspects of texts and how to offer consistent feedback need to be considered. Due to its unlimited feedback scope, it is more time-consuming compared to the focused approach.

The review revealed a new trend of research: 7% of the studies (e.g., Ellis et al., 2008; Frear & Chiu, 2015; Sheen et al., 2009) compared the impact of focused and unfocused feedback. The key question of whether focused or unfocused feedback leads to higher accuracy gains is still an open one in need of future investigations, as the findings revealed conflicting results. For example, Ellis et al. (2008) demonstrated that the direct unfocused group improved in terms

of accuracy compared to the direct focused group initially, but the focused group continued to improve in the long run. A study conducted by Sheen et al. (2009) also found that unfocused feedback is of limited pedagogical value when compared to direct focused feedback. However, this was not the case in Frear and Chiu's (2015) inquiry which claimed that both focused and unfocused WCF groups outperformed the control groups on immediate and delayed post-tests. Also, students' proficiency levels might constrain the effects of focused/unfocused feedback. Li and Vuono (2019) reasoned that learners at different levels benefit differently from different types of feedback. For example, it is possible that focused feedback might be more effective for low-proficiency learners and unfocused feedback for advanced learners. Therefore, future studies should investigate how proficiency affects focused/unfocused feedback.

2.2.6 RQ3: Writing task-related features in written corrective feedback studies

As for the different genres of writing tasks used in WCF research, I found that among the range of task types, narrative writing (29%) and paragraph writing with a variety of genres (14%) were the most predominant (Table 2.6). Despite using unfocused writing tasks in the studies, inconsistencies emerged from diverse linguistic and cognitive efforts that different genres of texts demand.

Table 2.6

Writing Task-related Features (n = 42)

Variables	Levels	No. of studies	%
Genres	Narrative writing	12	29
	Paragraph writing (during treatment)	6	14
	plus opinion-led essays (pre- and post-test)	6	14
	Picture description	3	7
	Opinion essay	3	7
	Email writing	2	5
	Autobiographical essay	2	5
	Dictogloss	2	5
	Argumentative essay	2	5
	Letter (e.g., informal, job application)	1	2
	Journal writing	1	2
	Persuasive essay	1	2
	Essay and summary	1	2
	Not stated		

Provision of prompts	Yes (pictures/prompts/reading text/ listening task)	22	52
	No	12	29
	Not stated	7	17
	Not applicable	1	2
Time constraints	Timed writing	33	79
	Not stated	5	12
	Untimed writing	3	7
	Both timed and untimed writing	1	2
Length of writing	30 minutes	11	26
	10 minutes (treatments) + 30 minutes (pre-and post-test)	6	14
	Not stated	6	14
	20 minutes	4	10
	50 minutes	3	7
	No limits	2	5
	70 words	1	2
	100 words	1	2
	500 words	1	2
	5 pages	1	2
	8 pages	1	2
	12 minutes	1	2
	15 to 20 minutes	1	2
	25 minutes	1	2
	45 minutes	1	2
	1 hour (timed) * untimed writing included	1	2
No. of texts	Three	11	26
	Four	9	21
	Five	6	14
	Many short texts	6	14
	Two	5	12
	Not stated	3	7
	One	2	5
Revision	Yes	34	81
	No	8	19
New writing task	Yes	39	93
	No	3	7

Even when similar genres are used, noticeable differences relate to writing prompts. Concerning this aspect, two distinctions can be made: whether the prompts are provided and

how they are operationalized. Overall, 22 studies (52%) offered writing prompts (either word prompts, picture prompts, or both) to assist students with unfamiliar vocabulary items (e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Truscott & Hsu, 2008; Van Beuningen et al., 2008), whereas the other twelve studies (29%) did not make any claims about the provision of writing prompts.

The other distinction concerns how writing prompts were operationalized: 18 studies offered four to six picture series (e.g., Benson & DeKeyser, 2018; Karim & Nassaji, 2018; Shintani & Ellis, 2013) to trigger L2 writers' ideas. The writing tasks in Benson and DeKeyser (2018) and Sheen et al. (2009) comprised word prompts in addition to picture series. A closer examination of writing tasks highlighted that some studies (e.g., Shintani & Ellis, 2013; Vyatkina, 2010) allowed the use of online/electronic dictionaries during the writing process, whereas in two studies (Karim & Nassaji, 2018; Lee & Yoon, 2020) students were not allowed to use any reference materials or to discuss the pictures with other members of the group during the writing sessions. These trivial variances relating to the nature of writing tasks must have impacted outcomes.

The third variable relating to writing tasks is time constraints. Most studies (79%) were timed and 12% did not report information about time constraints. The length of production measures can be used as an index of L2 writing proficiency under time constraints (Wolfe-Quintero et al., 1998). Writing produced within 30 minutes can provide insight into the level of fluency of L2 writers and the length of writing is also closely associated with the number of errors that a student may make which would impact the accuracy of written texts. For example, in Chandler's (2003) study, when calculations of error rate on the first and fifth writing assignments were made, text length was controlled by adjusting the measure of errors per 100 words, as the assignments did not yield texts of the same length. As for the length of time, I found different methods of limitations (i.e., word limit, page limit, or time limit). In all means of limitations, a wide range of differences were observed, making it impossible to compare findings across studies. Precisely, 26% of the selected studies limited the writing tasks up to 30 minutes, whereas 14% included writing tasks which lasted ten minutes during the treatment sessions and 30 minutes in pre- and post-tests.

In line with the Skills Acquisition Theory (DeKeyser, 2007), a balance between explicit instruction and extensive practice is among requisite conditions for linguistic accuracy gains. Therefore, the number of writing tasks (i.e., the amount of practice) that L2 writers are asked to complete during the intervention process is another key factor which most probably mediates the effects of WCF. The present review revealed that about a quarter of the studies (26%)

included three written tasks on three testing occasions (i.e., pre-test, post-test, and delayed post-test), whereas 21% asked the participants to write four texts during the intervention period. The dynamic WCF studies (n = 6) (e.g., Hartshorn & Evans, 2015; Kurzer, 2018) invited participants to attempt many short texts to ensure that writing practice is extensive and manageable during the whole treatment process.

Two other critical issues in WCF concern whether participants are required to revise their writing following the feedback and whether new writing tasks are used as indicators of improved linguistic accuracy. The inclusion of the revision process ensures that their attention has been drawn to a single or multiple aspects of writing that they need to improve. Most studies (81%) reported on students' revision as a mandatory step following WCF. However, critical debates persist regarding the value of revision studies, as these studies have failed to demonstrate that WCF can be carried over to new texts (see Ferris, 2010; Truscott, 2007, for detailed discussions). Due to criticism, claiming whether improvement in accuracy in students' revised texts is an indication of learning, 93% of studies included new writing tasks and compared the accuracy gains between students' initial and new texts.

2.2.7 RQ4 (Part I): Measures of linguistic accuracy in written corrective feedback studies

The effectiveness of WCF is primarily measured by assessing the improvement of linguistic accuracy of learners' texts using outcome measures such as error ratio, obligatory occasion analysis, and holistic scoring (Table 2.7). Four of the 42 included studies included more than one measures (e.g., error ratio and holistic scoring), resulting in 46 total studies (Ashwell, 2000; Chandler, 2003; Evans et al., 2010; Vyatkina, 2010). Among the diversity of accuracy measures, error ratio (Chandler, 2003) was the most frequently used in 37% of the 42 studies (e.g., Bonilla Lopez et. al., 2017; Mawlawi Diab, 2015; Riazantseva, 2012; Sheen et. al., 2009; Van Beuningen et. al., 2008, 2012). Riazantseva's (2012), for instance, claimed that error ratio is one of the accuracy measures used in earlier studies to correlate with measures of language proficiency in which accuracy was measured as "a ratio of the total number of errors to the total numbers of words in the sample" (p. 425).

Depending on the scope of feedback (i.e., focused versus unfocused), the calculation of the percentage of errors varied. For example, in Mawlawi Diab's (2015) focused feedback study, the percentage of pronoun agreement errors in students' writing samples were calculated using the formula: (the number of pronoun agreement error/ the total number of words per essay) ×

100, whereas in an unfocused feedback study (e.g., Nicolas-Conesa et. al., 2019), error ratio percentage for the writing tasks was calculated using the formula: (total number of errors/ total number of words) × 100. Other studies (e.g., Nicolas-Conesa et al., 2019; Truscott & Hsu, 2008) compared error rates in students' initial essays and their subsequent revisions or in their initial and final texts to examine the effect of error feedback.

Table 2.7

Measures of Linguistic Accuracy

Type	Measures	No. of studies	%
1	Error ratio	17	37
2	Obligatory occasion analysis	14	30.4
3	Holistic ratings	4	8.7
4	Error-free clause ratio	3	6.5
5	Other measures (e.g., error counts, error reduction, error-free T-unit ratio, Jacobs scale, suppliance in obligatory content analysis etc.)	8	17.4

Another trend in the reviewed studies showed that more than 30% performed an obligatory occasion analysis to assess the accuracy gains of the targeted linguistic features (e.g., Benson & DeKeyser, 2018; Bitchener, 2008; Bitchener et. al., 2005; Ellis et. al., 2008; Sheen, 2007; Shintani & Ellis, 2013). Bitchener et al. (2005), for instance, compared the efficacy of different feedback types (i.e., direct written feedback and student-researcher 5-minute individual conference and direct written feedback only) on ESL student writing, targeting three types of errors: prepositions, past simple tense, and indefinite article. Accuracy performance was calculated as the percentage of correct usage of each targeted linguistic feature. All obligatory occasions of the target forms in each script were identified and each occasion was then inspected to determine whether they were correct or incorrect, e.g., three correct uses of the targeted linguistic form from the ten obligatory occasions mean a 30% accuracy rate. Other studies calculated the accuracy gains in a similar vein, although the targeted linguistic forms and the feedback types varied across them.

The third widely-used measure in assessing the overall writing quality of students' texts (8.7%; e.g., Chandler, 2003; Evans et. al., 2010; Vyatkina, 2010) was holistic rating. These studies made use of holistic evaluations as a secondary measurement to assess the overall writing quality of students' texts. Several studies have highlighted the importance of utilizing two accuracy measures, as assessing writing accuracy seems difficult, considering both

linguistic accuracy and content. For example, Evans et al. (2010) made use of error-free clause to total clause ratio and holistic scoring to assess students' accuracy improvement. The teacher assigned a score using a holistic scoring rubric which accounted 75 percent for linguistic accuracy and 25 percent for the quality of content development. The authors concluded that similar improvement patterns were found between these accuracy measures. Along the same lines, Vyatkina (2010) investigated the extent to which different feedback types (i.e., direct, coded, and un-coded feedback) benefited students' accuracy of 15 specific error types including lexical choice, noun-related errors, and verb-related errors by comparing their error rate changes between the rough draft and the final draft of three essays. Holistic evaluations were used to assess linguistic accuracy and other dimensions of writing such as the content, relevance, creativity, and complexity. The results of Vyatkina (2010) showed that all groups improved their accuracy in their revisions and no significant differences were found in overall error rate changes among the groups.

The importance of reported reliability estimates of the accuracy measures is worth pinpointing: the lack of reliable measures can be one of the reasons contributing to controversial findings in WCF research. Over half of the studies (59.5%) examined inter-rater reliability scores on the assignment of errors to the targeted categories, and only a few studies (16.6%) failed to report any reliability scores on error identification (Table 2.8). The rest of the inquiries measured intra- and inter-rater reliability scores and reported high reliability estimates. Further results showed that rigorous studies (e.g., Ellis et al., 2008; Hartshorn & Evans, 2012; Sheen, 2007; Sheen et al., 2009) provided important insights into detailed scoring guidelines to assessing free-production writing tests in line with recent calls for replicable research (Polio & Shea, 2014).

Table 2.8

Studies Reporting Inter- and Intra-rater Reliability

Studies with no inter-/intra-rater reliability	Ashwell 2000; Fazio 2001; Chandler 2003 (sample 2); Van Beuningen et al. 2008; Rummel and Bitchener 2015; Stefanou and Révész 2015; Kim, Choi et al. 2020
Studies reporting inter-rater reliability	Ferris and Roberts 2001; Bitchener et al. 2005; Ferris 2006; Sheen 2007; Bitchener 2008; Bitchener and Knoch 2008; Bitchener and Knoch 2009; Sheen et al. 2009; Bitchener and Knoch 2010a; Bitchener and Knoch 2010b; Evans et al. 2010; Hartshorn et al. 2010; Vyatkina 2010; Evans et al. 2011; Hartshorn and Evans 2012; Mirzaii and Aliabadi 2013;

	Shintani and Ellis 2013; Shintani et al. 2014; Hartshorn and Evans 2015; Mawlawi Diab 2015; Shintani and Ellis 2015; Bonilla Lopez et al. 2017; Benson and DeKeyser 2018; Kurzer 2018; Lee and Yoon 2020
Studies reporting intra-rater reliability	Ellis et al. 2008; Truscott and Hsu 2008
Studies reporting inter- and intra-rater reliability	Chandler 2003 (sample 1); Riazantseva 2012; Van Beuningen et al. 2012; Frear and Chiu 2015; Bonilla Lopez et al. 2018; Karim and Nassaji 2018; Nicolas-Conesa et al. 2019; Zhang 2021

2.2.8 RQ4 (Part II): Advantages and pitfalls of frequently used measures of linguistic accuracy

A key consideration in selecting the appropriate measure of linguistic accuracy depends on its discriminating power. As one of the commonly used measures of linguistic accuracy, error ratio has distinctive strengths that other measures tend to lack. For example, it can be used in studies regardless of the scope of feedback. The calculation of errors in a sample can be justified based on the targeted error types (either a limited or a broad coverage of error types). As indicated earlier, error ratio has been utilized as the accuracy measure in both focused- (e.g., Fazio, 2001; Mawlawi Diab, 2015) and unfocused-feedback studies (e.g., Ferris, 2006; Van Beuningen et al., 2012). Unlike other studies which considered either focused or unfocused feedback, Sheen et al. (2009) examined the extent to which focused and unfocused feedback facilitated students' accuracy by utilizing the error ratio as the accuracy measure.

A key limitation of error ratio arises from the discriminating power between the severities of errors, although it is useful for quantifying error distribution in a written sample (Riazantseva, 2012). It is unlikely to be an issue in studies targeting linguistic forms; however, it can be a problem in studies targeting a broad number of error categories. Taking the limitations into account, WCF researchers should define what is considered an error and what is not and offer detailed scoring guidelines for raters to follow. Furthermore, the average length of student sample texts should be considered in cases where text length is not controlled. Van Beuningen et al. (2012), for instance, noted that due to the relatively short texts (i.e., around 120 words), a 10-word ratio was used rather than the common 100-word ratio.

Obligatory occasion analysis is well-suited for focused feedback studies (e.g, Bitchener & Knoch, 2010a, 2010b; Shintani & Ellis, 2013), as it affords greater discriminating power than

the error ratio. However, it seems unrealistic and inefficient to identify the correct and incorrect obligatory occasions of each linguistic feature in unfocused feedback studies (Bonilla Lopez et al., 2018; Karim & Nassaji, 2018). Hartshorn and Evans (2012), for instance, argued that this method "may not be possible to identify all of the obligatory occasions for every linguistic feature; nor is it appropriate for writing samples that include no obligatory occasions for a particular linguistic feature" (p. 232) without accounting for lexical errors. Further limitations were indicated by Sheen et al. (2009): they pinpointed problems in using obligatory occasion analysis for the selected grammatical features (i.e., articles, copula 'be', regular past tense, irregular past tense, and preposition), although they shared little information on what the problems were.

The other widely-employed measure, holistic scoring, tends to be practical and ecologically valid in that learners can be evaluated on such measures by writing teachers. Evans et al. (2010), for instance, used holistic scores and noted that they reflect the fullest dimensions of writing, as they consider both linguistic accuracy and the content and "it is a fairly efficient measure for a teacher who must evaluate multiple paragraphs in a timely manner" (p. 457). However, scoring is limited in that raters may find it difficult to distinguish accuracy from other global issues such as text length and content, and thus results may not be reliable.

In what follows, I provided narrative overviews of previous studies pertaining to the effectiveness of feedback from multiple sources (see Section 2.3), the impact of feedback on syntactic complexity of students' writing (see Section 2.4), and student engagement with teacher and automated feedback (see Section 2.5). In the closing discussion, insights were drawn from these studies and potential research gaps were identified.

2.3 Effectiveness of feedback from multiple sources: Teacher and automated feedback

2.3.1 Efficacy of teacher feedback in second language writing

As of the 1990s, the issue of effectiveness of WCF in L2 teaching has been debated; language teachers and researchers have questioned whether WCF should be provided to improve students' overall writing development. To respond to this intense debate, empirical studies have been designed to investigate the effectiveness of WCF in both ESL and EFL contexts. With the growing realization of the contribution of WCF to improved accuracy of students' writing, researchers are engrossed in undertaking research to explore the efficacy of different types of WCF. Most studies compared the extent of improved accuracy of different types of WCF over an extended period. For example, the following studies investigated the positive effects of

WCF in ESL and EFL contexts: Bitchener, 2008; Bitchener and Knoch, 2009, 2010a, 2010b; Sheen, 2007; Shintani and Ellis, 2013. In these studies, the types of feedback emphasized are direct and metalinguistic feedback and the findings demonstrated that the treatment groups outperformed the control groups both on immediate and delayed post-tests. Furthermore, qualitative studies on WCF studies have been undertaken to understand the nature of feedback practices in classrooms, students' responses and preferences of the feedback types, and the beliefs and perspectives of teachers with regard to the provision of WCF (e.g., Elwood & Bode, 2014; Ferris, 2014; Mao & Crosthwaite, 2019; Montgomery & Baker, 2007).

When examining the effectiveness of direct and indirect feedback, studies resulted in somewhat inconsistent and conflicting findings. Some studies (e.g., Bitchener & Knoch, 2010a; Van Beuningen et al., 2008, 2012) suggested that direct feedback is more effective than indirect feedback. Looking at Bitchener and Knoch's (2010a) study, the findings revealed clear evidence of the greater effectiveness of direct feedback over indirect feedback both in immediate post-test and delayed pieces of writing of advanced L2 writers, whereas the accuracy gains of indirect feedback could not be retained in the delayed post-test. Moreover, their findings corroborated those of recent studies (e.g., Ellis et al., 2008; Sheen, 2007): direct feedback benefited the students more than indirect feedback. In contrast, other studies (e.g., Ferris, 2006) provided evidence that indirect feedback is more effective, as it provides students with a more robust learning experience through meaningful engagement. In other words, it offers opportunities to students to figure out what the underlying issues of the errors that are indicated without providing them with the correct forms. Other studies (e.g., Ferris & Roberts, 2001), however, have observed no major differences between direct and indirect feedback. In addition, whether direct feedback is more beneficial than indirect feedback depends on students' L2 level of proficiency and how far beyond their proficiency the linguistic item is also matters.

Current empirical studies investigated the effects of focused WCF which has targeted a single or some limited numbers of linguistic features. Some studies in this trend (e.g., Benson & DeKeyser, 2018; Bitchener et al., 2005; Bitchener & Knoch, 2009; Kim et al., 2020; Rummel & Bitchener, 2015; Shintani & Ellis, 2013) examined the influence of two or more error types, all of which have been form-focused in some ways. Moreover, this line of research targets linguistic accuracy, measuring the improved accuracy of the targeted linguistic structures in students' writing samples before and after the treatments. In this case, some research compared the improved accuracy of the initial texts to the revised texts whereas others compared the improved accuracy of the initial texts with that of new writing tasks.

A plethora of studies have found significant accuracy improvements in experimental groups compared to control groups, with the exception of some studies which were reviewed by Truscott (2007) (see Truscott, 2007 for details). Despite Truscott's (2007) reviewed studies, current focused studies on WCF revealed the positive effects of the feedback in general. Notably, most focused studies solely examined language issues rather than content-related features such as organization, cohesion, coherence, richness and appropriacy of vocabulary, and task achievement. A limited number of linguistic features (e.g., English article system, tenses, and prepositions) targeted in focused WCF studies focus primarily on the targeted items and measure accuracy gains of these items specifically. However, teachers' WCF practice in instructional settings is not on single targeted linguistic features as writing constitutes many aspects to be considered. Regarding teacher's feedback practice, the most frequent feedback employed by teachers appears to be unfocused (Karim & Nassaji, 2018, 2019; Lee, 2004). For example, Lee (2004) revealed that 67% of EFL teachers (n = 58) in Hong Kong secondary schools provided feedback on all errors made by their students and both teachers and students preferred unfocused and comprehensive error feedback. Also, Karim and Nassaji (2018, 2019) questioned the ecological validity of focused feedback, as teachers often provide feedback on diverse errors they observe rather than on errors of a single or a limited number of errors.

Regarding the measurement of the impact of teacher feedback on students' revision, previous studies have looked into either students' revision operations (Ferris, 2006; Han & Hyland, 2015) or revision accuracy (Karim & Nassaji, 2018). Ferris (2006) classified students' revision operations into three categories: error corrected, incorrect change, and no change, whereas others (e.g., Karim & Nassaji, 2018; Van Beuningen et al., 2012) calculated accuracy improvement in students' revised texts through the measurement of error ratio. Long-term effectiveness of written feedback was verified by several studies (e.g., Karim & Nassaji, 2018; Rummel & Bitchener, 2015), which employed a range of accuracy measures. Despite these differences in measurement tools, most studies reported a positive influence of feedback on students' revisions and new texts.

Although significant positive impact of teacher feedback has been found on students' writing, it takes teachers considerable time and effort (Ferris, 2007; Zhang, 2017) to find ways to tackle students' writing issues at word, sentence, and text levels. Time constraints, large class size, and teachers' workload pose major challenges that prevent them from giving adequate feedback. Consequently, teachers tend to offer feedback primarily on language-related errors rather than on content-related issues in students' writing (Lee, 2009). Thus, to

ease teacher feedback burden and to enhance the efficacy of teacher feedback, the role of automated feedback has come on the foreground.

2.3.2 Efficacy of automated feedback in second language writing

Due to the substantial numbers of test-takers, Automated Essay Scoring (AES) systems like Project Essay Grade (PEG) (Page, 1994), Intelligent Essay Assessor (Hearst, 2000), and Electronic Essay Rater (E Rater) (Burstein et al., 2001) have become the primary engines in high-stakes examinations (e.g., GRE, TOEFL, IELTS, and GMAT) to analyse and assess text features at lexical, syntactic, semantic, and discourse levels. Deane (2013) provided evidence for the effectiveness of AES systems, demonstrating strong correlations with overall performance, as they can measure text production skills through an automated process to identify text features, assign a score, and generate feedback. In line with the online writing evaluation services such as Criterion Online English Evaluation Service (https://criterion.ets.org/), and the Pearson Test of English (https://www.pearsonpte.com/), writing teachers have applied automated feedback in tandem with instructor feedback in classrooms to provide instant and consistent feedback, employing AWE tools like *Criterion®*, *Writing To Learn*, and *My Access!*. With the advancement of educational technologies and computer-mediated language learning in the 21st century, the integration of computer-generated automated feedback in writing instruction has gained popularity due to its consistency, ease of scoring, instant feedback, and multiple drafting opportunities (Stevenson & Phakiti, 2014). In this regard, exploring the role and effectiveness of automated feedback has become a key field. Therefore, this section reviews the findings of the current literature on automated feedback to examine the role of automated feedback with an emphasis on the extent to which it helps L2 writers improve their accuracy.

Appendix B presents a summary of the selected studies ($n = 11$) on automated feedback. Several studies (El Ebyary & Windeatt, 2010; Kellogg et al., 2010; Li et al., 2015) proved that automated feedback leads to increased revision and helps improve accuracy from a first draft to a final draft. Li et al. (2015) examined the impact of AWE feedback on writing accuracy with the focus on how *Criterion* affected writing performance. When measuring the change of accuracy after calculating the average numbers of errors identified by *Criterion*, they found that *Criterion* led to increased revisions and helped improve linguistic accuracy from a first draft to a final draft. The findings of the study corroborated with studies conducted by El Ebyary and Windeatt (2010) and Kellogg et al. (2010). However, Dikli and Bleyle (2014)

claimed that AES systems promote revisions mainly on the surface features of an essay in terms of language issues (i.e., grammar, usage, mechanics, word choice, sentence structures, and convention) rather than other aspects of writing such as communicative intent, idea development, and rhetorical purpose. Hyland and Hyland (2006) also noted that the nature of automated feedback is limited, as it emphasizes surface features of writing, such as grammatical correctness and language use.

Concerning the big question of whether AWE feedback improves writing quality of students' texts, it is difficult to draw conclusions about the efficacy of automated feedback due to the paucity of research, controversial results, heterogeneity of participants, and other methodological issues (Stevenson & Phakiti, 2014). For example, Wilson et al. (2017) proved the effectiveness of automated feedback, whereas Saricaoglu (2019) and Wilson and Czik (2016) found no significant improvements between pre- and post-tests. Also, these studies were different from one another in terms of contexts and designs, and research foci. In particular, Saricaoglu (2019) explored the impact of automated feedback on the improvement of ESL learners' written causal explanations in an academic writing class. In contrast, Wilson and Czik (2016) examined the improvement of writing quality by comparing the text samples of the two groups of students (i.e., one group received both automated feedback and instructor feedback and the other received only teacher feedback).

Seven out of eleven studies (64%) failed to specify the proficiency levels of the participants. This might be an important issue if these studies targeted lower proficiency level students, as the language of the feedback generated by AWE tools (e.g., Criterion and MY Access!) is likely to be advanced for lower-proficiency students. As a result, studies targeting lower-proficiency learners might reveal the ineffectiveness of automated feedback. For example, Aluthman (2016) asserted that feedback from *Criterion* was not likely to be beneficial for low-level language learners, as students found it difficult to comprehend complex-error feedback. Moreover, the comments operated by automated feedback are either prompt-specific or generic. For example, written feedback global organization tends to take the form of generic suggestions or reminders (e.g., "Does your text have an introduction?"), whereas feedback on language use frequently takes the form of indirect corrective feedback (e.g., "This verb cannot be used in passive voice").

Concerns over the use of AWE feedback still linger in terms of scoring, complexity of AWE feedback, the amount of feedback, and the failure in reflecting social, contextual, and multimodal aspects of writing (Stevenson & Phakiti, 2019). In the case of scoring, AWE programmes might assign high scores to texts that have been deliberately illogical; using

complex sentences with sophisticated ideas due to the nature of automated feedback generated by the system (i.e., failure to identify the semantic aspects of writing). Accordingly, students might focus on how to get higher scores and paying attention to grammatical correctness without considering other aspects of writing such as communicative intent and the content domains of the writing. Moreover, AWE feedback is also subject to criticism due to the fact that AES systems are not capable of interpreting meaning, inferring communicative intent, evaluating factual correctness, and quality of argumentation, or taking the writing process into account.

Despite these pitfalls of automated feedback, studies document that it lowers teachers' feedback burden and allows them to be selective in the feedback they provide (Grimes & Warschauer, 2010). More precisely, the integration of automated feedback into writing instruction is expected to reduce teacher's feedback workload and allow them to better use their time and focus more on content-related issues. Therefore, Stevenson and Phakiti (2014, 2019) called for more research which examines how automated feedback can be integrated into classroom contexts to support writing instruction. This is what I discuss in the next section.

2.3.3 Studies comparing the effectiveness of teacher and automated feedback

In studies investigating the impact of automated feedback, little attention has been paid to the comparison of the nature of teacher and automated feedback (Dikli, 2011; Dikli & Bleyle, 2014; Qassemzadeh & Soleimani, 2016). Dikli and Bleyle (2014), for instance, investigated the use of *Criterion* in a college ESL writing class and compared instructor and *Criterion* feedback across feedback categories of grammar, usage, and mechanics. They found large discrepancies between these two feedback types and highlighted that the teacher provided more and better-quality feedback compared to *Criterion* feedback. Specifically, the study pointed out that the teacher identified more errors (570 compared to 94) than *Criterion* did. Another study by Qassemzadeh and Soleimani (2016) suggested that both teacher and Grammarly feedback positively influenced students' learning of passive structures. However, the current trend of studies which compared automated feedback and teacher feedback, is still limited. Therefore, relatively little is known about the efficacy of automated feedback in comparison to teacher feedback. Hence, new research is needed to scrutinize the applicability of automated feedback in writing instruction to highlight how AWE tools can be integrated with teacher feedback to produce instant and consistent feedback on students' pieces of writing in L2 classrooms. Such studies contribute to the understanding of how teachers can make effective

use of automated feedback when responding to students' writing and how they can go about selecting errors that automated feedback fails to respond to.

2.4 Written corrective feedback and syntactic complexity

2.4.1 Syntactic complexity as a complex construct

L2 writing scholars commonly agree that complexity, accuracy, and fluency (CAF) measures best capture students' language development (e.g., Barrot & Gabinete, 2021; Housen et al., 2012; Skehan, 2009). As Barrot and Gabinete (2021) posited, complexity is characterized as "the ability to produce more advanced language", accuracy as "the ability to avoid errors in performances", and fluency as "the ability to produce written words and other structural units in a given time" (pp. 1-2). These traits of language development are assessed to investigate the effects of instruction and individual differences (Housen et al., 2012). Although I outline the role of CAF measures in assessing students' writing development, it is beyond the scope of my study to describe how each measure is investigated in detail. I have opted to focus on one of these measures: syntactic complexity.

Syntactic complexity focuses on the sophistication of syntactic features that an L2 learner produces and the range or variety of those features (Ortega, 2003). Therefore, the assessment of syntactic complexity requires manual analysis to calculate the production units including phrases, clauses, and sentences. Though earlier studies (see Ortega, 2003) employed a limited number of syntactic complexity measures (i.e., only two to five), the use of online computational tools render the evaluation of syntactic complexity possible through overcoming the constraints of a labour-intensive nature of manual analysis (Petchprasert, 2021). As a result, recent studies have utilized automated tools to evaluate the syntactic complexity of students' writing including Coh-Metrix (McNamara et al., 2014) and L2SCA (Lu, 2010).

2.4.2 Syntactic complexity and writing proficiency

The role of syntactic complexity in academic writing is obvious, as it has been one of the important measures of L2 writing proficiency studied for decades. As Wolfe-Quintero et al. (1998) suggested, syntactic complexity is defined not in terms of how many production units (e.g., clauses, T-units, or sentences) are present in writing, but in terms of how varied and sophisticated these production units are. The main purposes of employing syntactic complexity measures in L2 writing research include (i) evaluating the effects of a pedagogical intervention on the development of grammar, and writing ability, (ii) investigating task-related variations

in L2 writing, and (iii) assessing differences in L2 texts written by learners across proficiency levels and over time (for a review, see Crossley, 2020; Ortega, 2003).

The relationship between syntactic complexity and language proficiency has been examined extensively (Lu, 2011; Ortega, 2003; Wolfe-Quintero et al., 1998). Research on L2 writing suggests that, despite differences in studies, indices of complexity increase as students become more proficient in the target language (Barrot & Agdeppa, 2021; Crossley 2020; Lu 2010, 2011; Ortega 2003; Wolfe-Quintero et al., 1998). In other words, they tend to produce more complex syntactic structures with longer and more varied sentences. Barrot and Agdeppa (2021) revealed an interaction between language proficiency and syntactic complexity measures such as length of production unit indices, degree of phrasal sophistication indices, and weighted clause ratio. Other studies examined changes in learners' syntactic complexity over time (e.g., Barrot and Gabinete 2019; Bulté and Housen 2014; Yoon and Polio 2017) and reported developments characterized by measures of syntactic complexity. Bulté and Housen (2014), for instance, found a significant increase in the length of linguistic units at all levels of syntactic organization (e.g., phrase, clause, sentence, and T-unit) over the course of a semester-long academic English language programme. Overall, these empirical studies have stated that syntactic complexity is an objective index of L2 writing proficiency.

At the syntactic level, complexity has been operationalized through indices that measure the construct at the phrase, clause, or sentence level such as the length of phrases, clauses, T-units, and sentences. In a research synthesis of college-level L2 writing, Ortega (2003) found that three measures tapping length of production (mean length of sentence, mean length of T-unit, mean length of clause), one measure reflecting amount of coordination (mean number of T-units per sentence), and two measures gauging amount of subordination (mean number of clauses per T-unit and mean number of dependent clauses per clause) were the most frequently used syntactic complexity measures in the literature.

2.4.3 Written corrective feedback and its impact on syntactic complexity

Limited studies in WCF research examined whether the provision of WCF influences syntactic complexity in students' writing (Eckstein et al., 2020; Eckstein & Bell, 2021; Hartshorn & Evans, 2015; Van Beuningen et al., 2012; Xu & Zhang, 2021) (Table 2.9). Generally, findings from such studies are inconclusive: some studies (Fazilatfar et al., 2014; Hamano-Bunce, 2022; Van Beuningen et al., 2012) found that WCF supports the development of syntactic complexity and does not make students produce structures that were linguistically simplified, whereas

others (e.g., Eckstein & Bell, 2021; Hartshorn et al., 2010) stressed an adverse effect on writing complexity.

As Van Beuningen et al. (2012) found, students who received feedback demonstrated higher syntactic complexity than those from the control group. Along the same lines, Fazilatfar et al. (2014) also indicated significant complexity gains in the experimental group when comparing their first and final compositions. These findings were later reinforced by Li et al. (2020) in which students' syntactic competence improved on some syntactic complexity measures following the feedback from an automatic writing evaluation tool. The same goes for Hamano-bunce (2022) who found out that comprehensive WCF contributes to the development of the subordination dimension of syntactic complexity, as results showed a significant increase in the syntactic complexity of revisions.

Other studies, however, demonstrated that writing complexity was largely unaffected by the provision of feedback (Evans et al., 2011; Hartshorn & Evans, 2015; Xu & Zhang, 2021; Zhang & Cheng, 2021). For example, the results from Zhang and Cheng (2021) revealed that comprehensive WCF has no effects on syntactic complexity of students' writing. These results corresponded to those of Evans et al. (2011), as comparing the complexity of the treatment and control groups did not show any significant differences. Building on these results, Hartshorn and Evans (2015) conducted a 30-week study and examined the effects of feedback on complexity. Similar outcomes were reported and thus the authors postulated that a gain in one aspect of writing (accuracy) is offset by a loss in another (complexity). In contrast to these studies, two other inquiries (Eckstein & Bell, 2021; Hartshorn et al., 2010) claimed that learners' writing complexity was negatively affected by dynamic WCF. For example, Eckstein and Bell (2021) found a significant reduction in syntactic complexity in students' texts with dynamic WCF over time compared to a control group. Overall findings from these studies shed light on the fact that L2 writers might produce structurally less complex writing to improve their linguistic accuracy.

Table 2.9

Summary of Empirical Studies on the Impact of Feedback on Syntactic Complexity of

Students' Writing

Authors	Participants & context	Source of feedback	Complexity measure(s)	Key findings
Hartshorn et al. (2010)	47 ESL students; US	Teacher	Mean length of T-unit	Dynamic WCF had slight unfavourable effect on writing complexity
Evans et al. (2011)	30 undergraduate students; US	Teacher	Mean length of T-unit	Negligible effect of dynamic WCF on syntactic complexity
Van Beuningen et al. (2012)	268 students from secondary education; Dutch	Teacher	Number of subordinate clauses per clause	Positive effect of comprehensive error correction on students' structural complexity
Fazilatfar et al. (2014)	30 advanced students from an English institute; Yazd	Teacher	Mean length of sentence and a dependent clause ratio from L2SCA	Unfocused WCF led to gains in syntactic complexity
Hartshorn and Evans (2015)	27 learners; US	Teacher	Mean length of T-unit Clause to T-unit ratio	No meaningful difference between the control and treatment groups in terms of two complexity indices
Eckstein et al. (2020)	23 international graduate students; US	Teacher	Mean length of sentence, mean length of T-unit, and complex nominals per clause from L2SCA	Timely feedback supports syntactic complexity development
Li et al. (2020)	66 non-English major freshmen; China	AWE (Pigai)	All indices from L2SCA	Students' syntactic competences developed in some aspects of syntactic complexity
Eckstein and Bell (2021)	63 international FYC students; US	Teacher	Mean length of T-unit, ratio per T-unit of clauses, coordinate phrases and complex nominals from L2SCA	Students with dynamic WCF demonstrated significantly less syntactic complexity over time

Xu and Zhang (2021)	65 sophomores; China	AWE (Pigai)	Number of clauses per T-unit from L2SCA	Syntactic complexity remained unchanged
Zhang and Cheng (2021)	72 English major sophomores; China	Teacher	Mean length of T-unit and Ratio of clauses per T-unit from L2SCA	Comprehensive WCF showed no effects on syntactic complexity
Hamano-bunce (2022)	42 adult participants from an English course; Scotland	Teacher	Number of coordinate phrases per T-unit, Number of complex T-units per T-unit, and Number of complex nominals per T-unit from L2SCA	Comprehensive WCF results in a significant increase in syntactic complexity (subordination)

Note. L2SCA = L2 syntactic complexity analyzer; FYC = first-year composition; AWE = automated writing evaluation.

2.5 Student engagement with written feedback

In the following section, I offer a brief discussion of how student engagement is conceptualised in educational research. This is followed by the conceptualisation of student engagement with feedback in L2 writing and the three dimensions of student engagement, namely behavioural, cognitive, and affective engagement with feedback. Section 2.5.2 deals with previous research on student engagement with feedback and the insights drawn from these studies. Section 2.5.3 further narrows the scope and reviews research on students' behavioural engagement with teacher and automated feedback which lays a foundation for designing an empirical study (Chapter 6). This section ends with a discussion of the implications arising from the review study, as well as potential research gaps that will be addressed in subsequent chapters.

2.5.1 Conceptualising student engagement with feedback in second language writing

In educational research, student engagement is conceptualised as "student participation, involvement, commitment, effort, time on task or motivation" (Dunne & Owen, 2013, p. xv) which could be measured through observable (e.g., completing assignments), and unobservable behaviours (e.g., commitment), and emotions. These concepts are also reflected in the student engagement research reviewed by Fredricks et al. (2004) in which researchers recommend studying engagement as a multifaceted construct with three dimensions: behavioural, cognitive, and emotional engagement. Behavioural engagement encompasses the idea of

participation in academic tasks, and participation in school activities and cognitive engagement emphasizes the idea of investment; it involves willingness to exert an effort and cognitive investment in learning. The third dimension, emotional engagement, is concerned with students' observable reactions in the classroom and at school (e.g., interest, boredom, and anxiety) which influence their willingness to do the work (see Fredricks et al., 2004).

In L2 writing research, student engagement with feedback is concerned with what students think, do, and feel when they receive feedback. Ellis (2010) conceptualized student engagement with feedback as the ways in which students respond to WCF; this is determined by students' revision operations in response to the feedback and the strategies they use to revise their work (behavioural engagement), their cognitive investment in processing WCF (cognitive engagement), and their attitudinal reactions to WCF (affective engagement). Cognitive engagement incorporates the depth of processing WCF (Storch & Wigglesworth, 2010), and the use of cognitive (how and to what extent their writing should be revised), and meta-cognitive (how students monitor and regulate their mental effort to process WCF) strategies are considered to be the indications of their cognitive engagement with the feedback. Taken together, as cognition and emotion influence human behaviours (Pessoa, 2008), the fusion of the behavioural, cognitive, and affective dimensions of student engagement has potential to provide a rich characterization of students' responses to WCF.

Despite the general agreement on how to examine student engagement with feedback, how it was operationalized in research studies differed, as different feedback sources tend to elicit different engagement styles (Zhang & Hyland, 2018). For example, behavioural engagement with teacher feedback is generally evaluated by revision operations by comparing students' original texts and revised texts in response to WCF (Han & Hyland, 2015; F. Hyland, 2003) and strategies they use when revising their work (Ferris et al., 2013). However, that was not the case with AWE feedback in which students could attempt to submit their revised texts multiple times. Therefore, studies (Zhang, 2017) examining the student engagement with AWE feedback measured behavioural engagement by means of the number of submissions and the time spent on revisions.

Other major concerns in engagement research relate to difficulties in assessing cognition (Pintrich et al., 2000). Although cognition is not readily observable, most research relied on observation and student self-reports to discern how students exert mental effort to use their prior linguistic and metalinguistic knowledge in understanding feedback. However, as students work, it is difficult to distinguish whether they spend mental effort in how and to what extent their texts should be revised, or they try to get their work done as quickly as possible.

Regardless of these difficulties, research on cognitive engagement with feedback (Han & Hyland, 2015; Storch & Wigglesworth, 2010; Zhang, 2017) examined how students utilize cognitive and meta-cognitive strategies when processing WCF via semi-structured interviews and retrospective verbal reports.

Further inconsistencies stem from the measurement of the depth of processing WCF (either at the level of noticing or understanding) as an indication of students' cognitive engagement. Specifically, Storch and Wigglesworth (2010) measured the depth of processing feedback through analysing the language-related episodes in students' pair talks as evidence of meta-awareness of feedback. However, the authors reasoned that coding of level of engagement is a largely inferential process, and thus the amount of verbalization evident in language-related episodes may not necessarily reflect depth of cognitive processing. On the other hand, Han and Hyland (2015) evaluated the level of students' cognitive engagement by means of awareness (e.g., noticing and understanding) and the use of cognitive and meta-cognitive strategies (e.g., reasoning, and planning for cognition). These differences in how students' cognitive engagement is assessed lead to difficulties in comparing the results across studies.

2.5.2 Previous studies on student engagement with feedback

Student engagement with feedback has been an under-researched area in L2 writing, although student engagement studies in education research demonstrated a positive association with achievement-related outcomes (Fredricks et al., 2004). Taking a multi-case study approach, the majority of studies investigated the nature of student engagement with the teacher, peer, or AWE feedback on students' writing (e.g., Ranalli, 2021; Zhang, 2017; Zhang & Hyland, 2018, 2022). Other studies examined how individual factors such as learners' beliefs, language proficiency, and feedback literacy mediated their engagement with WCF (Han, 2017; Han & Xu, 2019; Storch & Wigglesworth, 2010).

Most studies have suggested that engagement is a crucial mediating variable that explains how students make use of feedback. Key findings indicated that extensive engagement with feedback led to high levels of uptake (Zhang & Hyland, 2018; Zhang, 2017) and lack of engagement with feedback may be attributed to individual factors including both linguistic and affective factors. For example, Zhang and Hyland (2018) found that highly engaged learners tended to spend more time working with feedback, show more positive attitudes toward it, and employ more revision strategies, whereas moderately engaged learners were less motivated and showed less willingness to use the feedback.

Other studies (Yu, Zhang, et al., 2019; Zheng & Yu, 2018) showed the complex relationship between the three dimensions of student engagement. Zheng and Yu (2018) found that students' lower English proficiency cause imbalances among the three dimensions of engagement. Despite the students' positive affective engagement with teacher WCF, their behavioural and cognitive engagement was not extensive, resulting in unsuccessful revisions and scant awareness at the level of understanding WCF. These findings were further supported by Yu et al. (2019) who found that although students affectively engaged with peer feedback in a paradoxical way, they failed to use cognitive and meta-cognitive operations in the revision process, which reflected their low behavioural and cognitive engagement.

Given that student engagement with WCF is jointly mediated by learner factors, research has uncovered individual variations, ranging from learners' beliefs, experiences about WCF and L2 writing, their L2 learning goals and feedback literacy, to the interactional context in which WCF was received and processed (Han, 2017; Han & Hyland, 2015; Han & Xu, 2019). Han (2017), for instance, found a non-linear and reciprocal relationship between learner beliefs and learner engagement with WCF. Particularly, different types of beliefs (e.g., person-related beliefs, task-related beliefs, and strategy-related beliefs) exerted direct and indirect influences on students' engagement with WCF. In addition, Han and Xu (2019) stressed the importance of understanding student feedback literacy to foster deeper engagement with WCF. Taking a case-study approach, the authors found that unbalanced development of students' feedback literacy tended to limit their engagement with WCF.

In summary, findings from previous studies demonstrate the importance of students' engagement with WCF in promoting the efficacy of WCF and how the three dimensions of student engagement are interrelated in a dynamic and complex manner. Also, L2 writing researchers are aware that student engagement is jointly mediated by multiple learner variations and other contextual factors (Cheng & Liu, 2022; Pearson, 2022b; Zhang, 2020).

2.5.3 Research on behavioural engagement with teacher and automated feedback

Although researching all three dimensions allows for a rich characterization of students' responses to feedback, Fredricks et al. (2004) argued that the richness of involving the three components leads to the challenge of conceptualizing and studying each and their combination in nuanced ways. Accordingly, most studies have investigated the impact of a single type of engagement (e.g., behaviour) and a single outcome of interest (e.g., correlation between behavioural engagement and achievement). In studying student engagement with written

feedback, L2 writing scholars have reached a consensus that behavioural engagement with feedback concerned learners' uptake and revisions elicited by feedback and can be evaluated mainly through revision operations which can further be classified into correct revisions, incorrect revisions, deletion, etc. (Koltovskaia, 2020; Zhang & Hyland, 2018).

Research on behavioural engagement with feedback has investigated the nature of student engagement with feedback provided on their writing through analysing their feedback uptake (Han, 2017; Han & Hyland, 2015; Han & Xu, 2019; Tian & Zhou, 2020; Yu et al., 2019; Zhang & Hyland, 2018; Zheng & Yu, 2018). These studies adopted a naturalistic case study approach and involved students in EFL writing courses. A review of previous studies indicated relationships between extensive engagement with feedback and high levels of feedback uptake (Han, 2017; Han & Hyland, 2015; Zheng & Yu, 2018). Han and Hyland (2015) suggested that engagement is a crucial mediating variable in how students use feedback and how it impacts their writing development. Moreover, the authors underlined the importance of individual differences (e.g., beliefs and L2 learning goals) in understanding their engagement with feedback.

Studies examining the uptake of surface- and meaning-level feedback revealed that students seemed to utilize surface-level feedback more frequently (Dressler et al., 2019; Ene & Upton, 2014). Dressler et al. (2019) found that students addressed surface-level feedback focusing on writing mechanics more frequently (89.44%) than meaning-level feedback focused on argumentation, flow, and content (79.08%). Along similar lines, Ene and Upton (2014) posited that the highest rate of successful uptake was found in grammar (75%) compared to content, organization, and vocabulary. These findings shed light on students' successful incorporation of surface-level feedback into revisions and difficulties in understanding and addressing meaning-level feedback which requires more reworking to be integrated.

Given the growing popularity of automated writing evaluation tools in L2 writing classrooms, recent studies have considered how students engaged with automated feedback (e.g., Koltovskaia, 2020; Ranalli, 2021; Zhang, 2020) and compared how students engaged with teacher and automated feedback. Regarding studies that examined students' engagement with automated feedback, Zhang (2020) found variations in the levels of behavioural engagement: students adopted different revision operations in their new texts and exhibited varying patterns of engagement with automated feedback. Moreover, as with teacher feedback studies, findings revealed that it was the student engagement with automated feedback rather than the feedback itself that contributed to student learning.

Other studies contrasted the uptake of teacher and automated feedback and found that students' uptake rate of teacher feedback was higher than that of automated feedback (Shi, 2021; Tian & Zhou, 2020; Zhang & Hyland, 2018). Zhang and Hyland (2018) demonstrated that student engagement with two feedback sources differed in revision operations. Also, Tian and Zhou (2020) found that Pigai (http://www.pigai.org/) provided the highest amount of feedback, but resulted in the lowest in uptake rate (24.2%), whereas the teacher offered the least amount of feedback, but achieved the highest uptake rate (85.3%). Findings were similar in another study (Shi, 2021): the quantity and incorporation of feedback rate across the two feedback sources differed across genres. Another line of research compared the two feedback sources in terms of their feedback areas (Dikli & Bleyle, 2014; Thi & Nikolov, 2021b) and found the complementary nature of these feedback sources. Therefore, these studies provided recommendations for employing automated feedback to reduce teacher feedback burden and offer more room for the teacher to target higher-level writing concerns.

2.6 Conclusions and the way forward

This chapter provided a theoretical overview of previous research regarding the role of written feedback in L2 writing and a critical synthesis of the four key variables: research designs, feedback treatments, writing tasks, and accuracy measures which impact the efficacy of WCF in developing learners' written accuracy. My goal was to indicate how these four aspects should be considered methodologically and pedagogically to guide future studies in this burgeoning field of inquiry. The remaining sections (see Sections 2.3, 2.4, and 2.5) provide further insights into WCF research, attempting to connect the missing dots in a broader picture. I surveyed the current state of the field pertaining to the three research strands: (i) the effect of feedback from multiple sources on EFL students' writing; (ii) the effect of feedback on syntactic complexity; and (iii) student engagement with feedback. A review of the prior research led to identification of research gaps, which provided the impetus for this study.

Studies examining the efficacy of teacher and automated feedback provided evidence that the provision of feedback is conducive to L2 writing development. However, being cognizant of the limitations of teacher feedback resulting from contextual constraints, recent studies in WCF research (Dikli & Bleyle, 2014; Koltovskaia, 2020; O'Neill & Russell, 2019a, 2019b; Ranalli, 2018) have demonstrated the value of automated feedback in reducing the workload of teachers. For example, it is possible that automated feedback can alleviate lower-order concerns so that teachers can concentrate on higher-order concerns (e.g., content, organization)

instead. However, studies comparing the nature of teacher and automated feedback and how automated feedback could be used as a precursor to teacher feedback tend to be limited. To this end, I attempt to examine the nature of teacher and automated feedback (Grammarly), students' revision operations following feedback from different sources, and their attitudes towards the usefulness of feedback in their EFL course (Study I; Chapter 4).

Although studies on written feedback have confirmed the effectiveness of multiple sources of feedback in promoting learners' accuracy, much remains to be discovered about its impact on other aspects of language development. Concerns were raised regarding the possible unfavourable impact of feedback on the complexity of students' writing which resulted from their attention to producing accurate texts. Also in L2 writing studies, Polio and Shea (2014) investigated the relationships between accuracy and complexity and their findings suggested negative associations between these two constructs. Therefore, more research is needed to consider matters of complexity to investigate whether students' attention to producing accurate texts leads them to produce less structurally complex texts (Study II, Chapter 5).

Providing feedback on students' writing is an aspect of pedagogical practices that teachers do constantly, but how students react to it and benefit from it is sometimes less clear-cut. Differential success in learners' gaining from feedback has largely depended on their engagement with the feedback rather than the feedback itself (Zheng & Yu, 2018). Moreover, Zhang (2020) also argued that compared to a narrow focus on accuracy improvement in L2 student writing, it is more meaningful to examine how learners engage with feedback from different sources to enhance the possible benefits. Echoing their claims about the importance of studying feedback engagement, recent studies have examined how students engage with teacher, peer, and automated feedback by using a naturalistic case study approach. Major findings revealed a positive relationship between the degree of engagement with feedback and students' feedback uptake. There is however a need to examine students' engagement with different sources of feedback in a single study, as students nowadays have access to multiple sources of feedback (e.g., Zhang & Hyland, 2018). Furthermore, most research attends to how students engage with form-focused feedback and neglect their engagement with meaning-focused feedback. However, feedback in a writing course typically covers both linguistic and rhetorical aspects of texts. Therefore, researchers are urged to consider students' engagement with feedback on both local and global issues (Cheng & Liu, 2022). To address this research void, I investigated how EFL students engaged with form-and meaning- focused feedback from teacher and Grammarly (Study III, Chapter 6).

L2 writing research has used syntactic complexity measures to evaluate the impact of pedagogical interventions on the development of writing ability, or to assess the differences in L2 texts written by learners across proficiency levels and over time (Ortega, 2003). Following this theoretical understanding, further investigations were undertaken into syntactic features of students with different L2 writing proficiency. This in turn helped to understand how differently Myanmar and Hungarian students responded to teacher and automated feedback in their EFL writing courses. Moreover, I examined their language-related errors through epistemic network analysis to examine differences between these two groups (Study IV, Chapter 7).

CHAPTER 3. METHODOLOGY OF EMPIRICAL STUDIES

The purpose of this chapter is to present the research design and the methodology employed in my studies. The chapter begins by explaining why an experimental research design was adopted. This is followed by an overview of the research instruments as well as the feedback treatments offered by the teacher and Grammarly. The last section describes the procedure for analysing different types of data.

3.1 Pre-test and post-test experimental design

Informed by the research which explored the efficacy of written feedback on students' writing (Benson & DeKeyser, 2018; Ferris, 2006; Karim & Nassaji, 2018; Mirzaii & Aliabadi, 2013), I adopted the pretest-posttest experimental design in my investigations. Moreover, the selection of the research design was also informed by the rationale for the research project. Particularly, I explored how students benefit from teacher and automated feedback and their engagement with these feedback sources in EFL classes. To fulfil these research aims, I designed four classroom-based studies in two higher educational institutions. Here, it should be stressed that although causal inference research in education provides generalisable conclusions, classroom-based studies conducted in naturalistic classroom conditions also provide valuable insights, as they "produce knowledge that is context-specific, with potentially powerful, practical impact that can enhance teaching and learning" (Lee, 2022, p. 552).

Though many studies in the WCF research included a comparison group (Bonilla Lopez et al., 2018; Kurzer, 2018) , or a no-feedback group (Hamano-bunce, 2022; Lachner et al., 2018; Zhang & Cheng, 2021), I made no attempts to include a control group in the current inquiries due to ethical reasons. Particularly, I found it unethical to withhold feedback from students that they would typically receive in their EFL courses. Additionally, the rationale for the study is to examine how students incorporated teacher and Grammarly feedback into revisions and to track writing improvement over time rather than saying whether WCF is more effective than no feedback. A few studies in the WCF research (Ferris, 2006; Riazantseva, 2012; Vyatkina, 2010) provided justifications for the absence of control groups. Vyatkina (2010), for instance, noted that the study did not use a control group to avoid ethical concerns associated with asking students for a second draft if they had not given them feedback. As a result, the author acknowledges the limitation that no claims could be made as to whether WCF improves writing over time more than no feedback.

As for the methodological considerations, previous studies were markedly different in terms of the duration of intervention programmes. Though some studies (e.g., Ashwell, 2000; Ellis et al., 2008) failed to report the duration of the intervention, the phases in the majority of studies (e.g., Evans et al., 2010a; Riazantseva, 2012; Truscott & Hsu, 2008) lasted between 3 and 15 weeks. Responding to the call for longitudinal evidence on the efficacy of WCF on students' writing, the present study used longitudinal data and a pre- and post-test design to examine the effect of feedback from three sources in their revisions and to track their progress during the semester.

The next section provides information about the research procedures, and instruments including sources of data, methods of data collection and data analysis. However, it should be noted that due to the differences in the rationale for each inquiry, the data sources and how I analysed them tend to be different. Therefore, for the sake of simplicity and clarity, this chapter focuses on an introductory overview of the data collection timeline, data sources, instruments, and data analysis (Table 3.1) and further explanations regarding the methods of three studies are included in the subsequent chapters.

Table 3.1

An Overview of Three Experimental Studies

Study	Timeline	Research aims	Data source(s)	Instruments	Data analysis
I	August to October 2020 (13-week semester)	o To examine the effectiveness of feedback on students' writing	o Students' texts o Teacher feedback on students' texts o Grammarly feedback on students' texts o Students' questionnaire	o Writing tasks o Assessment rating scale o Self-assessment questionnaire	o Written feedback analysis o Revision analysis o Paired-samples *t*-tests o Qualitative analysis
II	August to October 2020 (13-week semester)	o To investigate the impact of feedback on syntactic complexity in students' writing	o Students' texts	o Witing tasks o Pre-and post-tests	o Syntactic complexity analysis o Paired-samples *t*-tests
III	February to May 2021 (14-week semester)	o To scrutinise students' behavioural engagement with feedback	o Students' writing o Teacher feedback on students' texts o Grammarly feedback on students' texts	o Writing tasks	o Written feedback analysis o Revision analysis

3.2 Instruments

3.2.1 Writing tasks

I used the writing tasks (including pre- and post-tests) which were adapted from the curriculum prescribed by the Ministry of Education in Myanmar. As these tasks were based on the themes introduced in each unit of the curriculum, I reasoned that students were familiar with these topics and that they had fewer difficulties in generating ideas when completing the task. Each writing task entails four guiding prompts which elicit students' responses (e.g., giving a brief narrative account of personal experiences or sharing views on a proposed statement with justifications). Particularly, these tasks required students to compose a four-paragraph guided essay (length between 300 and 400 words) without separate introduction and conclusion paragraphs. The rationale for providing sub-topics was to help students generate ideas as well as to allow for an entire essay to stay on topic. Although the writing topics are different in terms of topic, they are supposed to elicit free-constructed responses (Norris & Ortega, 2000). These responses were found to be valid measures in examining the efficacy of WCF on students' writing (Ellis, 2010; Li, 2010) as they enable students to produce the target language with few constraints and with meaningful communication.

3.2.2 Writing assessment rating scale

Study I (see Chapter 4) aims to investigate the impact of feedback on students' writing performance. To do so, I made use of an analytical rating scale to assess different aspects of writing. Before deciding what to include in the scale to measure overall quality of L2 texts, communicative competencies relating to linguistic, sociolinguistic, and pragmatic aspects that L2 writers at B1 levels possess were first examined in the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2018). The scales suggest that L2 writers at B1 levels "can write straightforward connected texts on a range of familiar subjects within his/ her field of interest, by linking a series of shorter discrete elements into a linear sequence" (p. 75).

Specifically, in terms of essay writing, they "can write short, simple essays on topics of interest, or a text on a topical subject of personal interest, using simple language to list advantages and disadvantages or give and justify his/ her opinion or summarize" (p. 77). In line with the CEFR illustrative descriptor scales, high-stakes writing assessment organizations such as Euro Exam (http://www.euroexam.org/en/euroexam-assessment-criteria), Cambridge

English Language Assessment (https://www.cambridgeenglish.org/images/231794-cambridge-english-assessing-writing-performance-at-level-b1.pdf), and international foreign language writing exams like TOEFL (https://www.ets.org/s/toefl/pdf/pd-toefl-ibt.pdf), and Cambridge English language exams like IELTS (https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf) develop analytical rating scales to assess writing proficiency (Table 3.2).

Table 3.2

Analytical Rating Criteria Used for Assessing Students' Writing Ability

Writing assessment organizations	Analytical rating criteria
Euro Exam (B1)	Task achievement, Coherence & cohesion, Grammatical range & accuracy, and Lexical range and accuracy
Cambridge English Language Assessment (B1)	Content, Communicative achievement, Organization (coherence), and Language (grammatical and lexical range and accuracy)
TOEFL	Content, Organization (coherence), and Language (syntactic variety and range of vocabulary)
IELTS	Task achievement, Coherence and cohesion, Lexical resource, and Grammatical range and accuracy

All these scales concern common criteria reflecting different aspects of the construct of writing: content/task achievement, organization/coherence and cohesion, and language (also known as grammatical range and accuracy, and lexical range and accuracy). Based on these criteria, I adapted the B1 analytical rating scale of Euro Exam (Euroexam International, 2019) to assess students' writing performance. Euroexam International offers language proficiency tests in general, business, and academic English and German at levels A1 through C1. The writing assessment scale comprises four criteria: task achievement, coherence and cohesion, grammatical range and accuracy, and lexical range and accuracy. Appendix C presents a description of the assessment criteria, along with definitions.

In the process of evaluating the applicability of the assessment scale, the six essays including two very good essays, two not- so-good essays, and two weak essays were initially scored by the researcher and an external marker who has had considerable experience in assessing L2 texts for many years. The first round of scoring indicated a number of issues in the assessment criteria. The first issue concerned a broad coverage of sub-categories under each criterion: task-achievement, coherence & cohesion, and grammatical range & accuracy. For instance, "Task

achievement" comprised five sub-categories: intention, instruction, effect, outcome, and content. Due to the high number of sub-categories under each criterion, the assessment was difficult in terms of duration of time used for scoring a text and the two raters found it hard to see the relevance of all the sub- categories in the actual tasks. Therefore, some necessary modifications were made based on the task requirements and a few sub-categories under the criterion were omitted. In redesigning the current scale, the scales used in assessing the writing proficiency (B1 level) of Hungarian 6th and 8th graders in English and German dual-language schools were adapted (Nikolov & Szabo, 2015). Likewise, similar issues in other criteria were also modified to ensure that the adapted scale particularly focused on the targeted genres and could discriminate the targeted writing performances. Another modification was made regarding the distribution of scores in the original scale, as the distribution of scales ranged from 0-5 under each criterion and the writing performance at the scales 2 and 4 shared features of scales 1-3 and 3-4, respectively, which might create discrepancies between raters. In other words, the distinctions between the different scores were not clear-cut in the original scale. Therefore, the original distribution of scales 0-5 was changed to 0-3 scales with individual descriptors for each scale. Moreover, the "grammatical range & accuracy" criterion was also modified, as it failed to consider the range of grammatical structures in the original version. After redesigning the scale, the two raters assessed the six essays and compared the scores based on the modified rating scale to reach an agreement. Inter-rater reliability coefficients (Pearson r) between the two scorers were .92 for the pre-test and .94 for the post-test on the overall assessment scale.

3.2.3 Language background questionnaire

This questionnaire aims to collect students' background information (e.g., age and gender) and their previous learning experiences regarding English writing (see Appendix D). For example, the students were asked to rate on a five-point intensity-scale from usually to never for items including "How often do you write in English in academic subjects at the university?", "How often do you engage in writing English as real-life activities, not for academic purposes?", and "How often do you receive feedback on your writing from your teachers in your English classes at the university?". They were also asked to rate their English writing proficiency level and explain their strengths and weaknesses in English writing.

3.2.4 Self-assessment questionnaire

A self-assessment questionnaire was developed to probe into students' emic perspectives on the effectiveness of feedback from three sources (see Appendix E). It includes closed items, eliciting information on the usefulness of feedback, and open-ended questions asking students to comment on how useful feedback is.

3.3 Feedback treatments

The next section discusses feedback treatments students received during the intervention period. Particularly, students received feedback from their teacher and Grammarly in different essays during their course. Therefore, a brief description of how the teacher and Grammarly gave feedback on students' texts and general characteristics of each feedback source are provided.

3.3.1 Teacher feedback

The provision of feedback took place either in Microsoft Word by using the "Track Changes" function (Study I & II) or in Google Docs (Study III) where the students submitted their work and the teacher provided written feedback on different aspects of their texts. In order to keep the feedback process as natural as possible, the instructors were not asked to change their normal practice or to limit their feedback to language- or content-related issues. The teacher offered error feedback and commentary feedback which corresponded to four assessment criteria of the grading rubrics: (i) task completion, (ii) vocabulary, (iii) structures, and (iv) coherence. In particular, teacher feedback focuses on language-and content-related issues associated with idea development, supporting details, task achievement, coherence and cohesion, grammatical range and accuracy, and lexical range and accuracy.

3.3.2 Grammarly feedback

Though all my investigations included Grammarly as a feedback provider, it should be noted that I used a free version in Study I and II, whereas Grammarly Premium was utilized in Study III. The rationale behind selecting this automated feedback tool is due to its reliability in checking grammatical accuracy (Vojak et al., 2011) and the fact that many studies used this software for identifying linguistic inaccuracies of written texts (e.g., Yi & Ni, 2015).

In both versions, Grammarly offers instant feedback for improvement once a paper is uploaded online, but the feedback scope differs depending on the version being used. For example, feedback in Grammarly free version is limited to spelling, grammar, punctuation, and conventions, such as spacing, capitalization, and dialect-specific spelling. With the premium version, writers receive feedback on four broad areas of writing issues: accuracy (grammatical and mechanical errors), clarity (writing issues that impact conciseness), delivery (issues relating to tone detection, politeness, formality, and inclusive language), and engagement (issues relating to word choice and sentence variety). Additionally, in the paid version, a feedback report can be downloaded in which writing issues in agreement with the criteria are indicated jointly with possible error correction.

3.4 Data analysis

This section overviews different types of data analyses conducted in the studies. In line with the rationale for each study, different analyses are performed. For example, Study I examines the possibility of using Grammarly as a supplementary feedback tool to support teacher feedback. To do so, feedback strategies and feedback scope of each feedback type are examined through written feedback analysis. Below is a summary of how written feedback is analysed, followed by a revision analysis, a syntactic complexity analysis, and a qualitative analysis of students' self-assessment questionnaire.

3.4.1 Written feedback analysis

As guided by Lee (2009), a written feedback analysis was performed to investigate the scope of teacher and Grammarly feedback. It involved error identification, categorization, and counting of feedback points: "an error corrected/ underlined, or a written comment that constitutes a meaningful unit" (p. 14). Feedback points marked on the students' first drafts were initially classified into language- and content-related issues and coded for analysis. For language-related issues, linguistic errors in the students' drafts were identified and categorized based on Ferris's (2006) taxonomy with adaptations.

For content-related issues, in-text and end-of-text comments were classified into four categories: (i) giving information, (ii) asking for information, (iii) praise, and (iv) suggestion according to the aim or intent of the comment suggested by Ferris et al. (1997). Particularly, drawing on the taxonomy of Searle's (1976) speech acts, the authors made a distinction between suggestion and giving information. The main goal of giving a suggestion was to explicitly

request specific information which should be added when revising a text. In contrast, the goal of the other commentary (i.e., giving information) was to provide information to the student. Unlike the first category, the teacher does not directly tell the student what to do with the information, but it is certainly implied that the student takes action on the information, which is known as illocution (see Searle, 1976). It should be noted that praise (e.g., *Overall, your essay is coherent and illustrated your personal connection to technology*.) was handled separately as they cannot be acted on directly and were thus excluded from the later analysis of students' revisions. I followed the same categorization scheme to code Grammarly feedback. Due to the different nature of teacher and automated feedback (Dikli & Bleyle, 2014), no Grammarly feedback focused on content-related issues.

3.4.2 Revision analysis

Following Koltovskaia (2020) and Tian and Zhou (2020), I scrutinized students' revision operations. The revisions made by the students in response to teacher and Grammarly feedback were identified as revision operations and feedback uptake (i.e., revisions made in response to feedback). Each feedback point was cross-linked to students' revisions in the revised texts and changes were analysed based on the revision operations of how the students responded to feedback. Specifically, the revision analysis of form-focused feedback adapted the coding schemes of Ferris (2006) and Han and Hyland (2015) and classified the students' revision patterns into four categories: (i) correct revision, (ii) incorrect revision, (iii) no revision, and (iv) deletion.

As for the meaning-focused feedback, I employed a coding scheme adapted from Faigley and Witte (1981) and analysed the students' revisions in three categories: (i) minimal changes (i.e., changes that paraphrase the concepts including simple adjustments or elaboration of an existing idea, maintaining the original meaning), (ii) substantive changes (i.e., changes that result in sweeping alterations in original meaning), and (iii) no changes (i.e., no attempts are made to consider minimal or substantive changes).

3.4.3 Syntactic complexity analysis

Two software packages (Coh-Metrix and L2SCA) were used to extract 17 features to cover the multidimensional nature of the syntactic complexity construct. I included three indices of syntactic complexity from Coh-Metrix (Crossley & McNamara, 2014): (i) syntax similarity,

(ii) left embeddedness (number of words before main verb), and (iii) number of modifiers per noun clause (Table 3.3).

Table 3.3

Coh-Metrix Indices Used in the Study

Syntactic features	Code	Measures/Indices
Syntactic variety	SYNSTRUTt	Sentence syntax similarity (across paragraphs)
Phrase-level complexity	SYNLE	Left embeddedness (words before main verb)
Phrase-level complexity	SYNNP	Number of modifiers per noun phrase

Another widely used computational system for syntactic complexity analysis is L2SCA (Lu, 2010): it entails fourteen ratio-measures selected from a large set of measures reviewed in Wolfe-Quintero et al. (1998) and Ortega (2003). These indices can be categorized into five sets which are detailed in Table 3.4.

Table 3.4

A Description of Syntactic Complexity Measures

Construct	Measure	Code	Formula
Length of production unit	Mean length of clause	MLC	# of words / # of clauses
	Mean length of sentence	MLS	# of words / # of sentences
	Mean length of T-unit	MLT	# of words / # of T-units
Sentence complexity	Sentence complexity ratio	C/S	# of clauses / # of sentences
Subordination	T-unit complexity ratio	C/T	# of clauses/ # of T-units
	Complex T-unit ratio	CT/T	# of complex T-units / # of T-units
	Dependent clause ratio	DC/C	# of dependent clauses / # of clauses
	Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Coordination	Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
	Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
	Sentence coordination ratio	T/S	# of T-units / # of sentences
Particular structures	Complex nominals per clause	CN/C	# of complex nominals / # of clauses
	Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
	Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

The selection of these indices was not only informed by how these measures incorporated the early L2 syntactic complexity measures reviewed by Wolfe-Quintero et al. (1998) but also advised by the findings of previous studies that examined syntactic complexity in L2 writing (e.g., Crossley et al. 2010; Lu 2010; Maamuujav et al. 2021; Yoon and Polio 2017). To conduct the syntactic complexity analysis, I exported students' essays into Coh-Metrix and L2SCA to automate the syntactic complexity measures. Then, I conducted independent samples t-tests on 17 measures of syntactic complexity to examine the differences in the essays.

3.4.4 Qualitative analysis of students' self-assessment questionnaire

The self-assessment questionnaires comprised closed and open-ended questions (see Appendix E). For closed items, the frequencies of responses were calculated and then, the students' perceived areas of improvements were reported. For open-ended items, a qualitative analysis was conducted on their responses commenting on how useful the feedback was. To this end, their explanations were summarized according to emerging common themes.

In summary, the chapter began with a description of the research design and the rationale for adopting a pre-test and post-test experimental design. Before detailing the instruments and feedback treatments, an overview of the studies with the timeline of data collection procedures and data sources was presented. The chapter ended with an explanation of how different types of analyses were conducted. The next four chapters (Chapters 4, 5, 6, and 7) will present the results of each experimental study.

CHAPTER 4. THE EFFECTS OF TEACHER, AUTOMATED, AND COMBINED FEEDBACK ON EFL STUDENTS' WRITING PERFORMANCE

In this chapter, the findings of the first study examining how automated feedback complements traditional teacher feedback in writing instruction are presented. First, the chapter outlines contextual issues Myanmar teachers encounter when providing feedback on students' writing, as well as how automated tools can be used to address surface-level writing problems. Following the statement of problem, research questions, procedures, and data analyses are introduced. Finally, the chapter discusses the findings that shed light on the possibility of integrating teacher and automated feedback in EFL courses.

4.1 Introduction

Writing in English is not only an essential component of learners' literacy development in school curricula, but it is also a passport to personal and academic advancement. Providing feedback on students' written texts is a common teaching practice for improving their writing skills. In line with these instructional practices, investigating the effectiveness of written feedback on writing performance has been a burgeoning field of inquiry and many researchers (e.g., Ferris, 2004, 2007; Karim & Nassaji, 2018; Lee, 2009) have stressed its importance. For example, Ferris (2004) suggested that feedback helps bridge the gaps between students' present knowledge, which indicates areas of potential improvement, and the target language that they need to acquire.

Providing feedback on students' writing requires a large amount of time and efforts on part of writing teachers (Zhang, 2017). Contextual issues including time constraints, excess workload, and large classes further increase teacher feedback burden. To reduce teacher burnout, AWE tools have come into play and complement teacher feedback in writing classes. The integration of automated feedback into writing instruction stems not only from the employment of automated essay scoring assessment to analyse and assess text features at lexical, syntactic, semantic, and discourse levels in high-stakes tests including Graduate Record Examinations (GRE) and Graduate Management Admission Test (GMAT) but also from the use of educational technology in today's classrooms. In line with favourable evidence of the reliability of AWE feedback (Li et al., 2015), L2 writing researchers (e.g., Koltovskaia, 2020; Ranalli, 2018) recommended integrating automated feedback into writing instruction to increase the efficacy of teacher feedback through freeing up teachers' time to focus less on

lower-order concerns (e.g., grammar, mechanics) and more on higher-order concerns (e.g., content, organization).

It is, therefore, of ultimate importance to investigate how automated feedback can be utilized as a support tool in instructional practices. Hence, this study investigated the potential to integrate Grammarly into writing instruction with the prospect of supporting traditional teacher feedback. Findings from the present study are expected to resolve teacher feedback burnout, addressing a potential of using Grammarly as a pedagogical tool and to provide insights into students' acceptance of automated feedback as a feedback provider in their EFL course.

4.2 Problem statement and research questions

Due to the country's political and educational situation, ELT research in Myanmar, especially classroom-based research is scant (Tin, 2014). Given the scarcity of publications in the Myanmar, the present study is a naturalistic classroom-based inquiry in a general English class at a major university in Myanmar. The course aims to improve students' English language skills and the prescribed syllabus intends to enhance their communicative language skills. Though developing students' English writing ability is one of the foci, teachers have limited time for providing adequate feedback on students' writing due to their heavy workload and other contextual constraints including large classes of mixed-ability students.

The study aimed to explore the potential of integrating automated feedback into writing instruction. To this end, I examined feedback strategies and the scope of teacher and Grammarly feedback to students' writing. I explored how students exploited feedback from multiple sources (i.e., teacher, Grammarly, and combined) in their revisions. I further scrutinized the general impact of feedback provision on students' writing performance over an academic semester. To triangulate research with students' perceptions of feedback, I probed into their emic perspectives regarding the usefulness of feedback from different sources through self-assessment questionnaires. This study was guided by four research questions:

RQ1. What are the feedback strategies and the scope of teacher and Grammarly feedback when responding to language- and content-related issues in Myanmar EFL students' writing?

RQ2. To what extent do the students make use of the feedback in three conditions (i.e., teacher, Grammarly, and combined) in their revisions?

RQ3. To what extent does the provision of feedback lead to improvement in their writing performance as assessed on a pre- and post-test over a 13-week semester?

RQ4. What are the students' views on the usefulness of feedback from different sources in their EFL course?

4.3 Methodology

4.3.1 Context

This study took place in a general English class at a major university in Myanmar in the second semester of 2020-2021 Academic Year. To meet students' language skills requirements, the English Department offers undergraduate and postgraduate courses to both English majors and non-English majors. The present study was conducted in an English-major class, in the communicative skills course (Eng. 1103). In classroom settings, students are engaged in communicative activities to develop their language skills under the guidelines of the prescribed curriculum, New Language Leader (Cotton et al., 2013), which is an integrated language course. The Academic Calendar consists of two semesters and each semester lasts fourteen weeks of instruction and two weeks of examinations. Lecture classes last 3 hours and tutorial classes last 2 hours on a weekly basis.

4.3.2 Participants

The sample comprised an intact class of 30 first-year English majors. Though the students' English proficiency level was supposed to be at B1 level, their L2 proficiency varied in their command of English grammar, familiarity with structures and vocabulary used in different writing tasks, and in previous formal EFL instruction. All were native speakers of Burmese and started learning EFL in schools around age five. Eleven were male and 19 were female, and all were of typical university age, between 17 to 18 years old. All students participated in the study on a voluntary basis and were informed of their rights to withdraw from the research at any time during data collection. Three students failed to complete one of the writing tasks, thus their data were excluded. The class teacher had an MA degree in Teaching English to Speakers of Other Languages (TESOL) and had over nine years of experience teaching the course in higher education institutions in Myanmar.

4.3.3 Instruments

Three types of instruments were used for data collection in this study: writing tasks, an assessment scale to assess improvement in students' writing performance, and self-assessment questionnaires (see Section 3.2.4).

Writing tasks: Six writing tasks were developed (including a pre- and post-test) on topics familiar to the students. The tasks were supposed to be ecologically valid, as they were retrieved from the prescribed curriculum. The genres comprised both argumentative and narrative essays as these two genres prevail in the syllabus. As shown in Figure 4.1, four guiding prompts were provided in the writing tasks which were similarly structured to minimize the possible linguistic differences. To reduce the possible bias and difficulties in comprehending the writing prompts, five faculty members who are familiar with the curriculum were initially asked to offer suggestions pertaining to the possible bias that the writing tasks might tap into. As they highlighted concerns related to vocabulary, topic familiarity, and the coverage of essay topics, the tasks were improved in response to their suggestions.

Figure 4.1.

A Sample Writing Task Used in Week 7

Write the essay according to its instructions. Your writing will be evaluated according to the following four criteria:
- the degree to which you have achieved the task;
- the organization and coherence of paragraphs;
- the range and accuracy of grammatical structures;
- the range and accuracy of vocabulary.

For the following essay, four prompts are given. Write about the four points in separate paragraphs. Write about 300-400 altogether.

Extroverts

Write an essay on your ideas about someone who is extroverted. In the text,
- characterize the person;
- describe the context where you met;
- tell a story about him/her and how he/she behaved with you and other people;
- discuss how his/her personality influenced his/her life.

Writing assessment scale: As previously mentioned in Section 3.2.2, the rating scale consists of four criteria ranging from task fulfilment to grammatical and lexical accuracy. To assess the students' essays, two researchers scored all written texts (pre- and post-test) independently and

calculated the mean scores (raters' total scores divided by two). Inter-rater reliability coefficients (Pearson *r*) between the two raters were *.92* for the pre-test and *.94* for the post-test on the assessment scale.

Self-assessment questionnaire: This questionnaire aims to gather students' responses to teacher feedback and Grammarly feedback on their written work. The students were asked to respond to three closed-ended questions (e.g., In which area(s) do you think you improved following the teacher feedback?) and five open-ended questions (e.g., Why do you like teacher feedback? List 3 reasons).

4.3.4 Research procedures

Considering that Studies I and II used the same dataset, which was gathered at a Myanmar higher education institution, I attempted to summarise the specifics of the data collection in one go. Data were collected over 13 weeks from August to October 2020: students completed six writing tasks including pre-and post-tests (Table 4.1). In the first week, the research project was introduced, then, they took the pre-test in the second week. The course followed a weekly syllabus: students were given a writing task and received feedback from the teacher, Grammarly, or both sources on the subsequent week after the completion of the initial writing task. There were four treatment sessions in the whole programme, and the students revised their texts following the feedback and sent the revised texts to the teacher via email the same week. The process continued until Week 10 when they completed the revised version of the fourth writing task. In Week 13, students completed the post-test and the self-assessment questionnaire.

Table 4.1

Data Collection Timeline

Timeline	Experimental group
Week 1	Introducing the research project
Week 2	Pre-test
Week 3	Essay 1
Week 4	Teacher feedback + Revision
Week 5	Essay 2
Week 6	Grammarly feedback + Revision
Week 7	Essay 3
Week 8	A combination of teacher feedback & Grammarly feedback + Revision
Week 9	Essay 4

Week 10	A combination of teacher feedback & Grammarly feedback + Revision
Week 13	Post-test
	Self-assessment questionnaire

4.3.5 Data analysis

To explore feedback strategies and the scope of teacher and Grammarly feedback, a qualitative written feedback analysis was performed following Lee's (2009) guidelines. Particularly, the analysis involved error identification, categorization, counting of feedback points: "an error corrected/underlined, or a written comment that constitutes a meaningful unit (a written comment comprising one sentence can consist of more than one feedback point, if it deals with more than one issue)" (p.14). I classified the feedback into language-and content-related issues, following the taxonomies of Ferris (2006) and Ferris et al. (1997) (Table 4.2).

Table 4.2

Feedback Categories Adapted from Ferris (2006) and Ferris et al. (1997)

Feedback category	Description
I. Language-related issues	
Word choice	Excluded spelling errors, preposition errors, pronouns, informal and unidiomatic usage
Verb tense	Tense and aspect errors
Verb form	Excluded verb tense errors
Word form	Excluded verb form errors and verb tense errors Excluded adjective and adverb errors
Articles/determiners	The misuse of zero, definite, and indefinite articles/ misuse of demonstratives and quantifiers. Excluded pronouns and possessive determiners.
Singular-plural	Noun ending errors
Pronouns	The misuse of pronouns
Run-on	Included comma splices
Punctuation	Inappropriate choice of punctuation marks. Excluded run-ons.
Sentence structure	Included missing and unnecessary words and phrases and word order problems. Excluded run-ons
Idioms	The misuse of idiomatic expressions
Subject-verb agreement	Excluded other singular-plural or verb form errors
Preposition	Inappropriate choice of prepositions
Conjunction	The misuse of conjunction
Collocation	Inappropriate combination of words and phrases

Omission of objects	The omission of objects in transitive verbs
Adjective	The misuse of adjectives
Adverb	The misuse of adverbs
Miscellaneous	Errors that could not be otherwise classified
II. Content-related issues	
Giving information	Giving the student the information about how the reader/teacher perceives the essay's ideas or organization
Asking for information	Asking the student to provide information unknown to the reader/teacher
Praise	Positive comments to attribute credit to the student for some characteristic, attribute, skill, etc., as compliments
Suggestions	Making a suggestion or request which could appear in either statement or question

Feedback points marked on the students' first drafts were cross-linked to students' revisions, and changes were analysed based on their revision operations. This study partly followed the revision analysis categories of Ferris (2006) and Han and Hyland (2015) and classified revision patterns into three categories: correct, incorrect, and no revision (Table 4.3).

Table 4.3

Students' Revision Analysis Categories Adapted from Ferris (2006) and Han and Hyland

(2015)

Revision operations	Description
Correct revision	The error was corrected as the teacher or *Grammarly* intended.
Incorrect revision	The error was addressed incorrectly.
No revision	No response to the error was apparent.

To investigate the impact of feedback on the students' writing performance, I calculated mean scores and standard deviations for the two data collection points (at the beginning and at the end of the course) and used paired-samples *t*-tests to check for the significance of the differences observed. Effect sizes (Cohen's *d*) were calculated to gauge the strength of the effect. As the sample size in this study was small, and the variables were not normally distributed, 'robust statistics' such as bootstrapping was more appropriate than other parametric analyses (see Plonsky et al., 2014). Specifically, Plonsky et al. (2014) argued that *t* tests and ANOVAs, the statistical tests used in quantitative applied linguistics research, may not be suitable for small samples or nonnormally distributed data commonly encountered in L2 research. Instead, the bootstrap approach is recommended as it includes a procedure that

randomly resamples from an observed data set to produce a simulated but more stable and statistically accurate outcome. Put simply, samples are randomly drawn from the original data and tend to be less skewed and kurtotic than the original sample. For these reasons, the *t*-tests with a bootstrap method in SPSS 22 were estimated for each of the 2,000 drawn samples and 95% confidence intervals were used to investigate the difference between pre- and post-test performance.

4.4 Findings and discussion

4.4.1 Feedback strategies of teacher and Grammarly feedback

The examination of teacher's feedback strategies revealed that the teacher used a range of feedback strategies when providing feedback on students' written texts (Table 4.4). Indirect feedback, the most prevalent feedback strategy (46%), was used to identify language-related issues such as sentence structure, verb form, subject-verb agreement, and noun-endings. Surprisingly, I found that the teacher made end comments (17.3%) and marginal comments (17%) on content-related issues including task fulfilment (e.g., *I think your discussion here is mostly related to your eating habits.*) and organization (e.g., *In my opinion, you should separate the sentences. It is too long, and it can sometimes make the reader misunderstand*). The least employed strategies included direct feedback (14%) and metalinguistic feedback (5.7%). Given the time constraint and other excessive workload of the teacher, it is understandable that metalinguistic feedback ranked the lowest among the feedback strategies.

Table 4.4

Distribution of Teacher Feedback Strategies Used in Students' First Drafts

Type of feedback	No.	Percent	Ranking
Indirect feedback	189	46.0	1
End comment	71	17.3	2
Marginal comment	70	17.0	3
Direct feedback	58	14.0	4
Metalinguistic feedback	23	5.7	5
Total	410	100.0	

As far as the Grammarly feedback strategies were concerned, the programme offers indirect, direct, and metalinguistic feedback, reconciling some contrasting theories in WCF

research (for details, see Karim & Nassaji, 2018; Rummel & Bitchener, 2015). For instance, when a grammatical error is detected, Grammarly underlines it and prompts possible ways to correct an error with the provision of indirect, direct, and metalinguistic feedback (Figure 4.2). For example, a click on an underlined expression takes students to preview possible changes and further explanations about the rules of the grammatical features. Moreover, examples are provided if one clicks on the "*Learn more*" symbol. Based on the feedback, students can either click the change in case they agree with the suggestions or dismiss them if they do not. All in all, Grammarly prompts more detailed feedback as far as language-related issues are concerned. As indicated in Figure 4.2, it provides a metalinguistic explanation of how and why an error should be revised.

Figure 4.2.

Grammarly's Interface Showing Text Editor and Providing Feedback on a Grammar Point
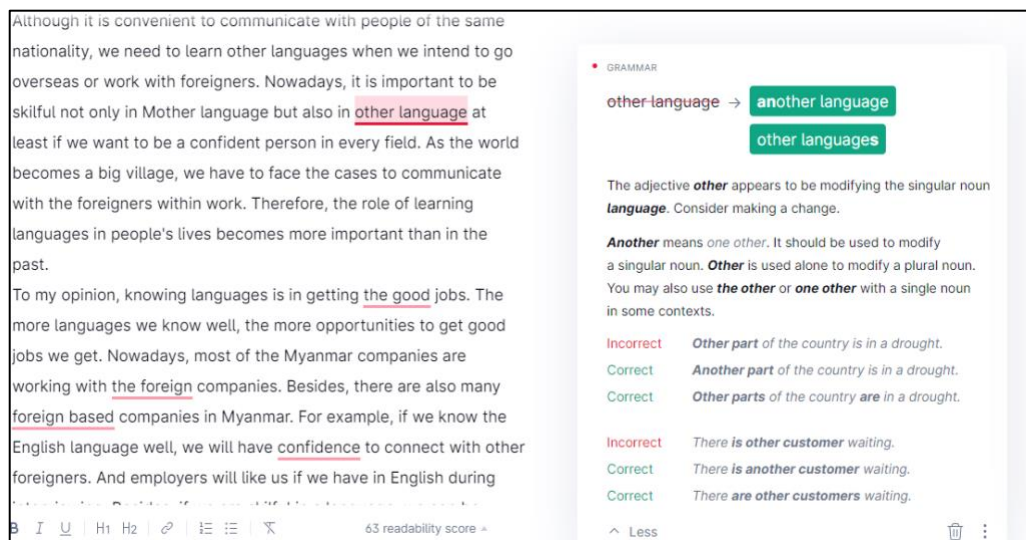
4.4.2 Scope of teacher and Grammarly feedback

Figure 4.3 summarises the scope of teacher feedback in comparison with Grammarly feedback and the percentage of each feedback category marked on the students' first drafts. In general, I found that the teacher focused on a broad coverage of writing issues at word, sentence, and text levels, whereas Grammarly indicated language errors: article/determiner, preposition, and miscellaneous errors including conciseness and wordiness.

The results of feedback analysis revealed that the teacher made 410 feedback points in 27 essays, targeting language errors (68.8%) and higher-level writing issues (31.2%). This result throws light on labour-intensive nature of teacher feedback. A more detailed analysis showed that teacher error feedback mainly concerned conjunction (10%), miscellaneous (9.5%),

punctuation (6.3%), and preposition errors (5.6%). In the teacher's commentary on content, praise got the highest percentage (11.7%), followed by suggestion (7.8%), giving information (6.4%), and asking for information (5.3%). Praise was used in comments on content by means of marginal comments or end comments. The rest of the positive comments included those on organization (e.g., *You can organize the essay systematically and I have found that there are links among the paragraphs*), grammar (e.g., *You can use complex sentences with few errors*) and vocabulary (e.g., *You can use adequate range of vocabulary: choice of word, collocation, etc.*). Given that there were 48 positive comments out of the total 410 feedback points, dealing with language and content issues, positive comments (i.e., praise) accounted for 11.7% of the total feedback, which was rather minimal. The finding that praise accounted for only 11.7% of the total written feedback contradicted those of Hyland and Hyland's (2001) and supported those of Lee's (2009). This might be due to differences in teachers' feedback beliefs about the role of praise in softening criticism when providing feedback on students' writing.

In addition, 7.8% of the content feedback included suggestions on how a text should be revised (e.g., *I think you should provide more information to support your ideas in the third* paragraph.) while discussing the content issues in more details (e.g., *Your second paragraph tends to focus on how you will use your own blog on social media.*). Further results indicated that content feedback was provided by means of directive comments, either by giving information (e.g., *This paragraph does not seem to discuss good and bad things about current eating habits.*), asking for information (e.g., *What can you learn from them?*), and clarification (e.g., *For some teenagers, their favourite food can also be nutritious meals. I don't see what you mean here.*).

Regarding the scope of Grammarly feedback, it targets a wide range of language areas including grammatical, and mechanical errors and other errors relating to usage, style, and conciseness, and offers suggestions to consider revisions to ensure that one's writing is comprehensible with less redundancy. Altogether, it detected 281 errors in 27 essays: the most predominant errors included article/determiner (43%), miscellaneous error (19.5%), and preposition (13.5%). Other less frequently indicated errors included those in conjunction (1%), sentence structure (0.3%), and pronoun (0.3%) (Figure 4.3).

All in all, it appears that Grammarly should probably be used as a learning tool to facilitate teacher feedback. The reason is associated with the scope of each feedback type: teacher feedback covered both language and content issues, whereas Grammarly provided feedback on language-related errors. Though this finding tends to be predictable when seeing Grammarly as a grammar-checking tool, the emphasis of its feedback on language use is its advantage. Its

automatic detection of article and prepositions errors was higher than that of teacher feedback. Thus, utilizing Grammarly effectively for offering feedback on these errors would possibly save time and effort for teachers. Moreover, being aware of errors that Grammarly is unable to handle could inform teachers about the potential areas that need special attention.

Figure 4.3.

Scope of Teacher and Grammarly Feedback
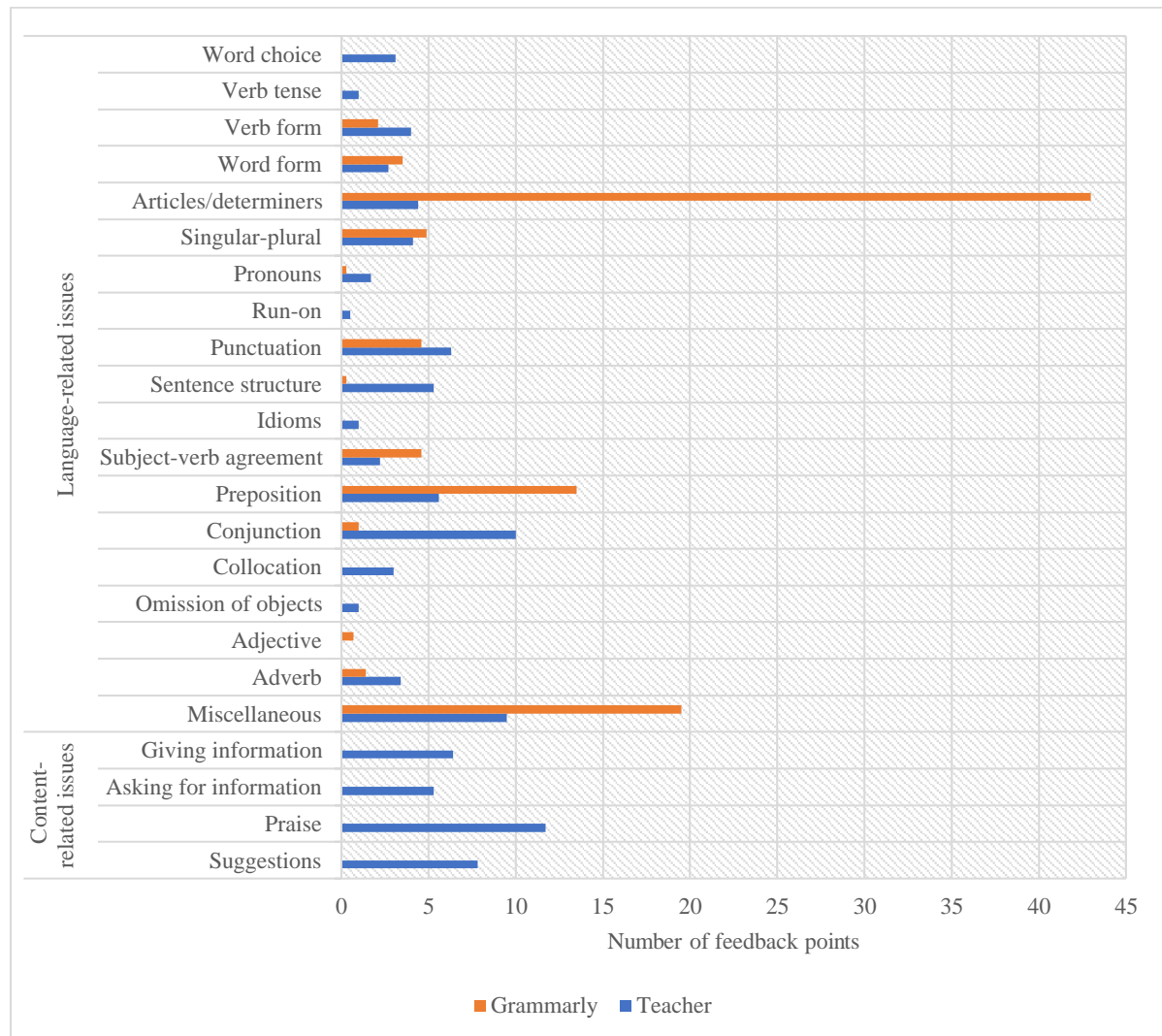
Additionally, it would be fair to say that the use of Grammarly along with the teacher feedback could also enhance the efficacy of teacher feedback. In line with the findings of previous research (e.g., Lee, 2009; Mao & Crosthwaite, 2019), I found that teacher feedback primarily attended to language errors (68.8%). Given time constraints and large classes of mixed-ability students, providing effective and individualized feedback on students' writing is far beyond the capabilities of teachers. In this regard, using automated feedback as an assistant tool should become an outlet for coping with surface errors, lightening teacher feedback

burden: freeing them up to focus on higher-order writing concerns such as content and discourse (Ranalli, 2018).

4.4.3 Impact of teacher, Grammarly, and combined feedback: Successful revision

When examining the influence of feedback on students' revision, the study considered how they acted upon feedback on language-related issues to facilitate comparability across feedback from the three sources. Based on the coded results, a general pattern of students' revision operations led to successful revision regardless of the source of feedback (Figure 4.4), indicating their acceptance of feedback. Some examples of how students acted upon teacher and Grammarly feedback can be seen in Tables 4.5 and 4.6 as follows:

Figure 4.4.

Comparison of Student Revision Operations
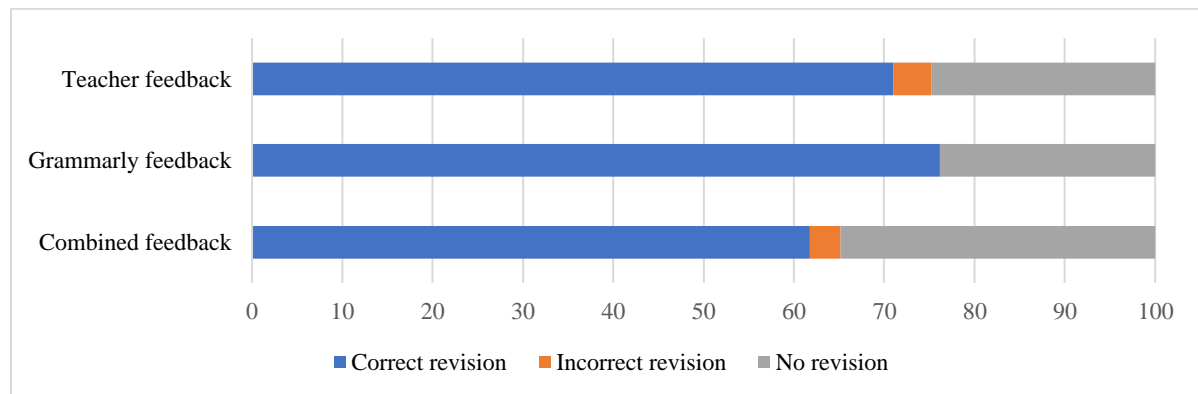
Table 4.5

Examples of Teacher Feedback and Student Revisions

Student original	Teacher feedback	Student revision
What is more, people can be mislead by the fake news and disgusted by the inappropriate pictures for either kids or elders.	What is more, people can be <u>mislead</u> by the fake news and disgusted by the inappropriate pictures for either kids or elders.	What is more, people can be misled by the fake news and disgusted by the inappropriate pictures for either kids or elders. [correct revision]
They are influence on my life.	They <u>are influence</u> on my life.	They influence on my life. [incorrect revision]
We can run a small business using social media is one evident fact.	<u>We can run a small business using social media is one evident fact.</u>	We can run a small business using social media is one evident fact. [no revision]

Table 4.6

Examples of Grammarly Feedback and Student Revisions

Student original	Grammarly feedback	Student revision
Learning language of the country where we want to go is more convenient.	Learning [the] <u>language</u> of the country where we want to go is more convenient. [The noun phrase **language** seems to be missing a determiner before it. Consider adding an article.]	Learning the language of the country where we want to go is more convenient. [correct revision]
If we travel abroad, lacking proficiency in foreign language is really bothersome.	If we travel abroad, lacking proficiency in foreign language is <u>really</u> bothersome. [It appears that **really** may be unnecessary in this sentence. Consider removing it.]	If we travel abroad, lacking proficiency in foreign language is really bothersome. [no revision]

The finding that teacher error feedback led to effective revision is in agreement with the findings of Ferris (2006) and Yang et al. (2006). Moreover, the lowest percentage of unrevised errors reflects their beliefs about and value of the importance of feedback in improving their writing performance. The results were interesting for Grammarly feedback which received the highest rate of correct revision (76.2%). The reason could be because it usually includes a concrete suggestion for revision that students can easily act upon. An example of this is shown in Figure 4.5.

Figure 4.5.

An Example of Grammarly Feedback and Student's Revision Outcome

> *Excerpt 1*
>
> Student original: We need to know at least three languages in order to apply scholarships at foreign universities.
>
> *Grammarly* comment: We need to know at least three languages in order to <u>apply</u> [apply for] scholarships at foreign universities. [The phrase following the intransitive verb **apply** seems to be missing a preposition. Consider adding one.]
>
> Student revision: We need to know at least three languages in order to apply for scholarships at foreign universities. [correct revision]

Further points of discussion concern how students responded to combined feedback. One would assume that combined feedback resulted in more feedback points, compared to other conditions. However, the opposite was true: it prompted fewer feedback points and a lower ratio of correct revision than teacher or Grammarly, the highest ratio of no revision. Possible explanations of lower feedback points may be due to students' increased awareness of teacher

and Grammarly feedback in previous essays or the teacher's reliance on Grammarly feedback, instinctively assuming that it would handle grammar errors.

Although the students could revise their errors successfully, it is worth exploring how well they revised individual error categories (Table 4.7) to further identify the error types that resulted in correct, incorrect, and no revisions. As the overall percentage of successful revisions was high, it was not surprising to note that the percentage of successful revisions in most error categories was also high regardless of the conditions. However, a closer examination of how students utilized feedback revealed thought-provoking new results. In connection with teacher feedback, while feedback on most error categories (e.g., conjunction, article/determiner, singular-plural, adverb, and word choice) was associated with correct revision, some feedback on idioms, pronoun, and sentence structure was left unattended. For example, 40.9% of errors in sentence structure led to no revision. This could be explained by factors such as low number of error identification in these categories and partial understanding of the instruction (Han, 2019). As Goldstein (2004) noted, reasons for unsuccessful or no revision included: lacking the willingness to critically examine one's point of view, feeling that the teacher's feedback is incorrect, lacking the knowledge to do the revision, lacking the time to do the revision, lacking the motivation, being resistant to revision, and many others.

Despite the overall successful revision when acting upon Grammarly (76.2%) and combined feedback (61.8%), the results indicated that the students largely ignored feedback on miscellaneous errors. This finding is probably due to students finding the feedback in this category unhelpful or unnecessary to revise. Figure 4.6 shows a typical example. This underlines how students selectively accept the feedback, filtering suggestions that are incorrect, or unnecessary (Cavaleri & Dianati, 2016).

Figure 4.6.

Example of Grammarly Feedback on a Miscellaneous Error and Student's Revision

Excerpt 2

Student original: If we travel abroad, lacking proficiency in foreign language is really bothersome.

Grammarly comment: If we travel abroad, lacking proficiency in foreign language is really bothersome. [It appears that **really** may be unnecessary in this sentence. Consider removing it.]

Student revision: If we travel abroad, lacking proficiency in foreign language is really bothersome. [no revision]

In addition, the question of whether Grammarly could be integrated into writing instruction could be answered by how the students responded to feedback in their revision. The comparison of revision outcomes in three conditions provides support for the potential of using

Grammarly along with teacher feedback. The reason is associated with high percentage of successful revision in singular-plural (92.9%), subject-verb agreement (92.3%), word form (90%), punctuation (84.6%), article/determiner (84.3%), and preposition (84.2%) following Grammarly feedback. Thus, it seems reasonable to suggest that utilizing Grammarly to handle errors in these categories could be effective and spare time for teachers to focus on other higher-level writing issues. Though the teacher made 22 feedback points in terms of errors in sentence structure, 40.9% of them were left unattended. This partly reflects indirectness or vagueness of teacher feedback which makes it difficult for students to act upon (Tian & Zhou, 2020). What should be stressed is that teachers might be able to pay more attention to these errors if they can efficiently make use of Grammarly to deal with surface-level errors.

Table 4.7

Comparison of Students' Revision Operations by Error Types

	Teacher (no./%)			Grammarly (no./%)			Combined (no./%)		
	CR	IR	NR	CR	IR	NR	CR	IR	NR
Word choice	11/84.6	0	2/15.4	0	0	0	6/85.7	0	1/14.3
Verb tense	2/50.0	1/25.0	1/25.0	0	0	0	7/70	2/20	1/10.0
Verb form	11/68.8	2/12.5	3/18.8	6/100	0	0	3/50.0	0	3/50.0
Word form	9/81.8	0	2/18.2	9/90.0	0	1/10	2/50	0	2/50.0
Articles/determiners	15/83.3	1/5.6	2/11.1	102/84.3	0	19/15.7	21/84	0	4/16.0
Singular/plural	13/76.5	1/5.9	3/17.6	13/92.9	0	1/7.1	3/37.5	0	5/62.5
Pronouns	4/57.1	0	3/42.9	1/100.0	0	0	4/66.7	0	2/33.3
Run-on	1/50.0	1/50.0	0	0	0	0	0	0	0
Punctuation	15/57.7	1/3.8	10/38.5	11/84.6	0	2/15.4	11/68.8	0	5/31.2
Sentence structure	11/50.0	2/9.1	9/40.9	1/100.0	0	0	8/61.5	3/23.1	2/15.4
Idioms	1/25.0	0	3/75.0	0	0	0	1/100	0	0
Subject-verb agreement	7/77.8	0	2/22.2	12/92.3	0	1/7.7	10/90.9	0	1/9.1
Preposition	16/69.6	1/4.3	6/26.1	32/84.2	0	6/15.8	16/69.6	0	7/30.4
Conjunction	33/80.5	1/2.4	7/17.1	2/75.0	0	1/25.0	13/54.2	2/8.3	9/37.5
Collocation	9/75.0	0	3/25.0	0	0	0	5/55.6	0	4/44.4
Omission of objects	4/100.0	0	0	0	0	0	3/100	0	0
Adjective	0	0	0	2/100	0	0	2/75	0	1/25.0
Adverb	10/71.5	1/7.1	3/21.4	4/100	0	0	5/100	0	0
Miscellaneous	28/71.8	0	11/28.2	19/34.5	0	36/65.5	6/20.0	0	24/80.0
Total	200/71.0	12/4.2	70/24.8	214/76.2	0	67/23.8	126/61.8	7/3.4	71/34.8

Note. CR: correct revision, IR: Incorrect revision, NR: no revision. Percentages represent frequencies of revision categories within each error category. For instance, 80.5% of the "conjunction" errors had a correct revision rating.

4.4.4 General impact of written feedback on students' writing performance

After receiving feedback over a semester, the students made improvement in their writing performance, as is shown in the significant increase in their post-test scores across four assessment criteria in both language and content. As presented in Table 4.8, there was substantial improvement in *task achievement* and *coherence and cohesion* in their post-test scores. Similarly, in connection with *grammatical range and accuracy* and *lexical range and accuracy*, the analysis suggested that the students showed notable improvement from the pre- to post-test. The effect sizes for all significant comparisons of learners' writing performance were medium to large (Plonsky & Oswald, 2014). The positive impact of feedback provision on new writing tasks was in line with previous studies (e.g., Karim & Nassaji, 2018; Rummel & Bitchener, 2015). Their improvement was likely attributed to the provision of consistent feedback throughout the course, affording them opportunities to enhance their writing performance.

Table 4.8

Comparison Between Pre-and Post-test Regarding Students' Writing Performance

Assessment Criteria	Pre-test		Post-test		$t(26)$	p	Cohen's d
	Mean	SD	Mean	SD			
Task achievement	2.25	543	2.65	.551	3.82	.003	.71
Coherence & cohesion	2.25	610	2.61	.560	3.90	.002	.75
Grammatical range & accuracy	2.20	559	2.52	.628	2.88	.017	.55
Lexical range & accuracy	2.26	685	2.69	.483	3.55	.003	.68
Overall writing performance	8.98	2.091	10.46	1.965	3.14	.006	.61

4.4.5 Students' views on the usefulness of teacher, Grammarly, and combined feedback

Results of the self-assessment questionnaires revealed that most participants perceived both teacher and Grammarly feedback to be effective and useful for improving their writing ability (Figure 4.7). Even though most responded that Grammarly feedback helped them improve their grammar (88.9%) and vocabulary (77.8%), none reported improvement in content and organization. Conversely, teacher feedback tended to be more valuable in their view as it

facilitated improvement in different aspects of writing and so did combined feedback. Despite students' positive impressions of both teacher and Grammarly feedback, their responses on the areas of improvements following the combined feedback were rated considerably higher across the different aspects of writing. This finding highlights the great potential for integrating Grammarly feedback into writing instruction to complement teacher feedback in line with previous studies by O'Neill and Russell (2019a, 2019b), Ventayen and Orlanda-Ventayen (2018), and Ranalli (2021).

Figure 4.7.

Students' Perceptions of the Usefulness of Teacher, Grammarly, and Combined Feedback
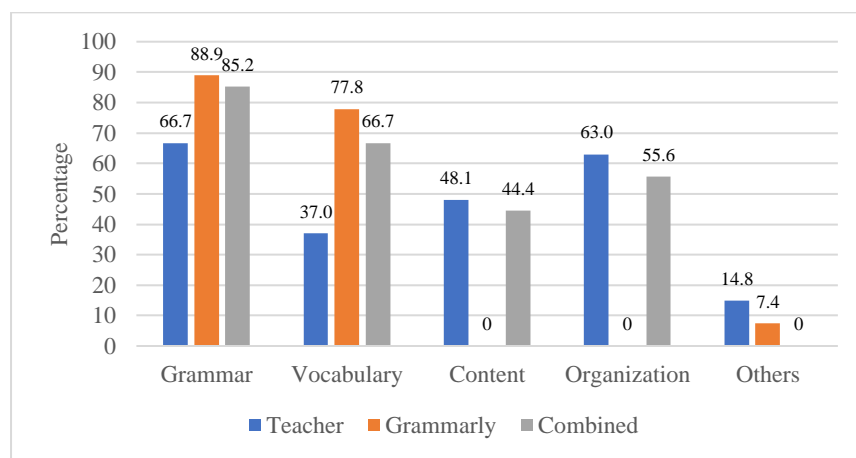
In the second part of the questionnaire, the students verbalized their thoughts as to why they liked the feedback. Almost all students acknowledged the value and effectiveness of teacher feedback. Their comments revealed three emerging themes relating to the nature of the feedback, how it enhances their motivation, and positive perceptions of teacher feedback. As to the nature of feedback, almost all students stressed the value of teacher feedback with positive comments and some included specific points: "*guide me when my writing goes out of context*" (Student 21), "*show me both strengths and weaknesses of my writing*" (Student 2), "*short and clear*" (Student 27), "*detailed feedback*" (Student 5), and "*clear explanation*" (Student 10). Teacher feedback also boosted their writing motivation, as students stated, "*I feel that I improved my writing and thinking*", "*It encourages me to keep on sharpening my advantages*", and "*It makes me confident*". As to how they perceived teacher feedback, students' responses spotlight their positive perceptions: "*Feels like I have a guidance for my writing, more specific than Grammarly, appreciate my effort*", "*I know what my writing is like from someone's perspective*", and "*My essay will be perfect because it encourages me to analyse my essay again and to change it to be a better one*".

Most comments regarding the usefulness of Grammarly feedback concerned its efficiency: "*It is easy to use and available for free*" (Student 3), "*It shows my mistakes clearly, and its explanations are easy to understand*" (Student 6), and "*I could use Grammarly at any time*" (Student 9)*. It's really convenient*" (Student 20). Some students provided more detailed comments on the effectiveness of Grammarly feedback: "*It is very convenient by the fact that Grammarly check all types of errors while texting, the explanations allow me understand better (e.g., collocations) and grammatical errors can be corrected without checking dictionary*". However, a few students identified dissatisfaction with Grammarly feedback: "*To be honest, I don't feel satisfied very much about it*" (Student 15) and "*Honestly, I didn't find Grammarly feedback useful*" (Student 19). Further responses revealed how the combined feedback helped them revise their essays: *"Teacher's feedback tells me my mistakes exactly and Grammarly fix that mistake"* (Student 20)*,* and *"It's a perfect combination"* (Student 25).

4.5 Conclusions

This classroom-based research examined integrating Grammarly into a large class to support teacher feedback. The results showed the pedagogical potential of Grammarly in facilitating teacher feedback due to its effective feedback on surface-level errors and students' general acceptance of automated feedback. The findings also revealed that utilizing Grammarly as an assistant tool will surely lessen teachers' feedback burden and enhance the efficacy of their feedback, as the teacher predominantly attended to language errors in providing feedback. Moreover, it seems that students' successful use of feedback in their revisions and increased performance scores on the post-test provide evidence that they successfully made use of the feedback and that the provision of feedback led to improvement in writing performance. In addition, positive attitudes towards the usefulness of feedback provide further insights into how much they valued feedback from their teacher and Grammarly.

Examining the nature of teacher feedback and Grammarly feedback suggested discrepancies pertaining to feedback strategies and feedback scope. For example, indirect feedback was the most predominant strategy used by the instructor, providing room for the students to engage in deeper cognitive processing (Loewen, 2012) because it requires them to self-correct. These findings ran counter to those offered by Ferris (2006) and Lee (2008): they concluded that the teachers made direct correction in which they provided students with the correct form. These differences may have resulted from contextual factors including large class size and the availability of time the teacher can spend to offer individualized support on students' writing.

In contrast, Grammarly employs three different feedback strategies (i.e., indirect, direct, and metalinguistic feedback) which tend to have a great potential for classroom use as a tool to facilitate instructor feedback, as indicated in previous studies (O'Neill & Russell, 2019b). Concerning feedback scope, both feedback types were unfocused in nature as they targeted a broad range of linguistic features. However, teacher feedback covered other aspects of writing such as task fulfilment and organization. It offered not only positive but also negative feedback. Unlike the instructor feedback, I found that Grammarly primarily targets the errors relating to linguistic features. These findings were in line with those by Ventayen and Orlanda-Ventayen (2018) and Hyland and Hyland (2006) and suggest that computer-generated feedback focuses on most common mistakes in English including subject-verb agreement issues and other grammatical features.

Some limitations should be addressed as I conducted the study in a course at a university. Future research needs to involve more courses at different universities with more participants. The inquiry failed to include a control group because I found it unethical to withhold feedback from students in a writing course. Therefore, comparison between a feedback group and a control group could not be drawn. However, my focus was to examine how students used feedback from three sources in their revisions and to track progress before and after the course. Such investigations offer insights not only into how the drafts changed as a result of feedback, but also into how the students perceived the effectiveness of feedback by triangulating the datasets. It is my hope that the findings from the present study inform teachers about how Grammarly could be used as an effective feedback tool to help relieve their burdensome tasks of responding to surface-level errors in students' texts. Moreover, investigating the nature of teacher and Grammarly feedback in an authentic classroom context further enhances the understandings of how these types of feedback deal with errors and what feedback strategies are employed when responding to issues in language use and other aspects of writing.

CHAPTER 5. THE EFFECTS OF TEACHER, AUTOMATED, AND COMBINED FEEDBACK ON SYNTACTIC COMPLEXITY IN EFL STUDENTS' WRITING

This chapter reports the results of the second study which investigated how feedback influences the complexity of Myanmar EFL students' writing. First, I describe what previous studies have found regarding the impact of feedback on syntactic complexity. Based on the review of related studies, I found that little attention has been paid to how written feedback influences syntactic complexity in students' texts. To address this research void, the following section presents details of complexity measures, data collection, and data analyses. The chapter concludes by reporting the findings and implications for L2 writing and pedagogy.

5.1 Introduction

In studies examining the importance of providing feedback on students' writing, most research aims to examine whether and how different feedback strategies aid the development of students' writing accuracy. Along with these aims, findings from such studies showed that the provision of WCF is beneficial for significant improvements in linguistic accuracy (e.g., Ellis et al., 2008; Karim & Nassaji, 2018; Rummel & Bitchener, 2015; Van Beuningen et al., 2012; Zhang & Cheng, 2021). For example, Ellis et al. (2008) and Van Beuningen et al. (2012) demonstrated that WCF helped the treatment groups improve their writing accuracy during revision and on the post-tests regardless of feedback types (e.g., direct versus indirect feedback). Likewise, studies that employed a dynamic WCF approach (e.g., Evans et al., 2011; Hartshorn & Evans, 2012, 2015; Kurzer, 2018) also found that students from the experimental group demonstrated accuracy gains, compared to those who joined the conventional process writing classes.

While there is consensus that WCF could potentially improve accuracy, little evidence suggests that it could promote syntactic complexity (e.g., Hamano-bunce, 2022; Van Beuningen et al., 2012; Zhang & Cheng, 2021). Indeed, Truscott (1996, 2007) asserted that WCF may probably lead to simplified writing as a result of students' efforts to avoid making mistakes. To respond to his contention, Van Beuningen et al. (2012) investigated the impact of WCF on syntactic complexity in students' writing. Although the authors did not observe any increase in syntactic complexity, WCF did not make students produce less complex writing either. These results were later reinforced by Zhang and Cheng (2021) who found that comprehensive WCF did not enhance syntactic complexity. In contrast, another study by

Hamano-bunce (2022) found that WCF can aid the development of the subordination dimension of syntactic complexity, as students demonstrated a significant increase in the syntactic complexity measured by the subordination.

With these findings in mind, further studies are needed to examine the impact of WCF on syntactic complexity in writing. To fill this research void, the present study attempts to investigate the influence of multiple feedback sources on syntactic complexity of EFL students' writing and to explore the effect of students' levels of proficiency (high-, mid-, and low-performing students) on the changes in their syntactic complexity during the course. The findings of this study are expected to have implications for research on L2 writing and pedagogical practices. Theoretically, the findings, derived from 270 written texts, will contribute to the growing body of WCF research, where few studies devote attention to the impact of WCF on syntactic complexity. From a pedagogical perspective, investigating EFL students' syntactic complexity can help teachers gain a better understanding of which aspects of syntactic complexity could or could not be developed by feedback. Moreover, such awareness can indicate whether feedback on L2 writing leads students to produce structurally less complex writing as a result of attempting to improve their linguistic accuracy.

5.2 Research questions

Given the conflicting results and paucity of studies that have solely focused on the effects of feedback on writing complexity, I examined the influence of feedback from multiple sources on the syntactic complexity of EFL students' texts and the possible effect of students' proficiency levels on the changes in their syntactic complexity. Accordingly, three research questions are addressed:

RQ1. To what extent do teacher, automated, and combined feedback affect EFL students' syntactic complexity in their revisions?

RQ2. To what extent does the provision of feedback impact EFL students' syntactic complexity over a 13-week semester?

RQ3. What is the effect of students' levels of proficiency (high-, mid-, and low-performing students) on the changes in their syntactic complexity during the course?

5.3 Methodology

5.3.1 Participants

I recruited 30 students (11 males and 19 females) who enrolled in an undergraduate English course at a university in Myanmar. They majored in English and registered for a communicative skills module to improve their English language skills. In terms of writing skills, the students were required to write argumentative and narrative essays and revise them based on the feedback as part of the course requirement. All participants were Burmese native speakers and began learning English as a foreign language in school at the age of 5. According to their scores in the National Matriculation Exam and the discussion with the class teacher, the English language proficiency of the participants was considered as low-intermediate which correspond to the B1 level on the CEFR scale (Council of Europe, 2018). Although they were classified as low-intermediate (B1) level of English, their English L2 writing proficiency varied in terms of previous EFL instruction. They were of typical university age, ranging from 17 to 18 years old. Before the study, students confirmed their willingness to participate voluntarily. They were briefed about the objectives of the research and the data to be collected. Moreover, they were informed that their personal information and data would remain anonymous, and that they could withdraw from the study at any stage. I excluded three students from the analysis due to their failure to complete some writing tasks during the intervention.

5.3.2 Syntactic complexity measures

In this study, I used L2SCA (Lu, 2010, 2011), a free automated text analyser that can compute 14 indices of syntactic complexity. I included six measures used in previous studies which looked into the effect of feedback on writing complexity (see Table 2.9 in Section 2.4.3). Two of these measures tap length of production (mean length of T-unit [MLT] and mean length of sentence [MLS]), two measures reflect the degree of phrasal sophistication (complex nominals per clause [CN/C] and complex nominals per T-unit [CN/T], and two measures gauge the amount of subordination (clauses per T-unit [C/T] and subordinate clauses per clause [DC/C]). The selection was informed by Ortega (2003) who reported that MLS, MLT, C/T, and DC/C were the most widely employed syntactic complexity measures across twenty-one studies included in the research synthesis of college-level L2 writing. She also noted three indices (MLT, C/T, and DC/C) were the most satisfactory measures, as they were correlated linearly with programme, school, and holistic rating levels. Moreover, MLT and CN/C indices were

found to be important indicators of English essay quality, as they indicated significant differences in essays written by non-native English students (Lu, 2011; Lu & Ai, 2015). The measure of complex nominals per T-unit (CN/T) was also added because it was supposed to be related to essay quality and complexity in student writing (Eckstein & Bell, 2021). See Figure 5.1 for definitions of the measures of syntactic complexity.

Figure 5.1.

Measurement Variables for Syntactic Complexity

Category	Code	Description
1. Mean length of T-unit	MLT	Number of words per T-unit
2. Mean length of sentence	MLS	Number of words per sentence
3. T-unit complexity ratio	C/T	Number of clauses per T-unit
4. Dependent clause ratio	DC/C	Number of dependent clauses per clause
5. Complex nominals per clause	CN/C	Number of complex nominals per clause
6. Complex nominals per T-unit	CN/T	Number of complex nominals per T-unit

I used an automated approach for assessing linguistic complexity due to its free availability, capability to process files in batches, speed, flexibility, and reliability. Lu (2010) reported correlations between 0.830 and 1.000 for structural unit identification and between .834–1.000 between syntactic complexity scores computed by human annotators and L2SCA. In addition, reliability and validity were later confirmed by Polio and Yoon (2018); they found the majority of measures reliable and valid for investigating variation in writing complexity. All correlations were significant at the .01 level, suggesting that the system achieved a high degree of reliability in the generated syntactic complexity scores.

5.3.3 Data collection

Students were required to complete six writing tasks over the course of a semester (August to October 2020), including their pre-and post-tests as well as the writing tasks during four treatment sessions (see Section 3.2.1). Upon submission of their first drafts, students received either teacher, Grammarly, or combined feedback the same week. They revised their essays based on the feedback and resubmitted them to their teacher the following week, completing one round of feedback. As the students completed four writing tasks and pre-and post-writing assessment during the course, the complete dataset comprised 270 essays including 108 preliminary drafts and their corresponding revised texts.

5.3.4 Data analysis

To examine the revision effects of teacher, automated, and combined feedback on syntactic complexity of EFL students' writing, a comparison was made between students' first drafts and revised essays. For the effects of feedback over the semester, a comparison was made between the pre- and post-tests that they completed on Week 3 and Week 13. Descriptive statistics and paired sample *t*-tests were used to find the effects of feedback. Data analysis for RQ3 was conducted in two stages. Students were initially classified into three groups: high-, mid-, and low-performers according to their scores on the pre-test using a tripartite split (Cardelle & Corno, 1981). The mean scores (the total scores of the two raters divided by two) were calculated at this stage. The inter-rater reliability coefficients (Pearson's *r*) between the two scorers were *.92* for the pre-test and *.94* for the post-test on the assessment scale. Then, I compared the changes in students' writing complexity in four revised texts across the three groups.

5.4 Findings

5.4.1 Effect of teacher, automated, and combined feedback on syntactic complexity of students' revisions

Table 5.1 presents the results of paired sample *t*-tests of syntactic complexity in students' texts between the initial and revised essays following teacher, Grammarly, and combined feedback. Overall, the findings indicated minimal differences between most comparison pairs; this outcome meant no significant effects of feedback from multiple sources on students' writing complexity in their revised texts. This was not the case, however, for some complexity indices in Essays 1 and 4 in which students received teacher and combined feedback. Particularly in Essay 1, students' decline in three T-unit measures (i.e., mean length of T-unit, T-unit complexity ratio, and complex nominals per T-unit) indicates that they applied fewer words, clauses, and complex nominals in T-units in their revised texts compared with their first drafts. For example, complex nominals per T-unit significantly decreased from initial drafts ($M = 1.64$) to revised essays ($M = 1.58$, $t(26) = 2.22$, $p = .04$). This was also true in the case of Essay 4 in which the students showed a significant decrease in dependent clause ratio and complex nominals per T-unit. They produced shorter complex nominals per T-unit when a comparison was made between first drafts ($M = 1.07$) and their revised versions ($M = 1.05$, $t(26) = 2.45$, $p = .02$). No other results indicated significant differences.

Table 5.1

Paired Sample t-tests of Syntactic Complexity Gains between the Initial and Revised essays

Index	Essay 1 (Teacher)			Essay 2 (Grammarly)			Essay 3 (Combined)			Essay 4 (Combined)		
	ΔM	*t*	*p*	ΔM	*t*	*p*	ΔM	*t*	*p*	ΔM	*t*	*p*
MLT	.73	2.86	.00*	.01	0.18	.86	−.03	−0.25	.81	.11	1.31	.20
MLS	.44	1.57	.13	.05	1.02	.32	−.08	−0.64	.53	.06	0.88	.39
C/T	.06	2.06	.04*	.00	0.36	.72	−.02	−1.56	.13	.02	1.78	.09
DC/C	.02	1.72	.09	.00	0.65	.52	−.01	−1.78	.09	.01	2.24	.03*
CN/C	.02	0.80	.43	.00	0.58	.57	−.01	−0.77	.45	.01	1.44	.16
CN/T	.11	2.22	.04*	.01	0.72	.48	−.02	−1.37	.18	.03	2.45	.02*

Note. ΔM: Mean difference between revised essay and its initial draft; *p ≤ .05.

5.4.2 Impact of written corrective feedback on students' syntactic complexity over the semester

To determine the effect of WCF on writing complexity over the course, I conducted a paired sample *t*-test and compared the means of syntactic complexity on the pre-and post-tests. Students' writing complexity showed little variation over a semester of WCF intervention with no significant differences in the complexity measures between the pre- and post-writing assessment (Table 5.2). Specifically, while results demonstrated increases in the means of MLT, MLS, C/T, and CN/T in the post-tests, these complexity gains did not reach statistical significance. Furthermore, the means of subordinate clauses per clause remained unchanged from pre- ($M = 0.36$) to post-tests ($M = 0.36$). In addition to these results, the students showed a reduction in the measure of complex nominals per clause, suggesting that the students produced fewer complex nominals per clause (e.g., adjective + noun, possessives, prepositional phrases) in the post-tests compared to pre-tests. All in all, it is reasonable to suggest that WCF does not show any effects on students' syntactic complexity development.

Table 5.2

Comparisons of Syntactic Complexity Measures in the Pre-and Post-tests

Index	Pre-test		Post-test		Paired sample *t* tests		
	M	SD	M	SD	*t*	df	*p*
MLT	14.63	2.69	14.64	2.61	0.01	26	0.99
MLS	15.84	3.13	16.12	3.16	0.87	26	0.38
C/T	1.62	0.26	1.64	0.25	0.50	26	0.61
DC/C	0.36	0.08	0.36	0.09	−0.38	26	0.70
CN/C	0.85	0.27	0.84	0.25	−0.22	26	0.82

CN/T	1.34	0.37	1.39	0.5	0.64	26	0.52

5.4.3 Effect of students' levels of proficiency on the changes in syntactic complexity

A comparison of students' syntactic complexity in their revised texts showed variations among the high-, mid-, and low-performing students (Figure 5.2). Overall, students from all three groups exhibited progress in the indices of T-unit complexity ratio and dependent clause ratio with certain levels of decline in the other complexity indices. More specifically, there was an upward trend in the T-unit complexity ratio, but the degree of improvement differed among the three groups of students. Although the high achievers displayed a decline from Essays 1 to 2, there was consistent development across Essays 2, 3, and 4. In contrast, the mid- and low-performing students achieved a significant improvement from Essays 1 to 2 with minor fluctuations from Essays 2 to 4. Similarly, the number of dependent clauses per clause increased from Essay 1 to 4 although I noted some variations among the groups. However, a major difference between the highest achievers and the other groups concerned the degree to which they made improvement: a certain level of development was found between Essays 1 and 2 among the mid-and low-performers, whereas slow and steady growth was observed in the dependent clause ratio of the highest-achievers throughout the course.

Unlike these two indices, other syntactic complexity measures (e.g., mean length of T-unit and sentence, complex nominals per clause and T-unit) indicated a decline during the course. Although the mean length of T-unit in students' writing increased from Essays 1 to 2, the results decreased from Essays 2 to 4 regardless of their levels of proficiency. This was not the case with the mean length of sentence: mid-performing students experienced a gradual reduction from Essays 1 to 3 which was followed by a noticeable improvement in Essay 4. As for the high- and low-performers, the mean length of sentence fluctuated dramatically from Essays 1 to 3 which levelled off or stabilized in Essay 4. As for the indices of complex nominals per clause and T-unit, the results showed a significant reduction among all groups throughout the course. The only exception included the high-achieving students, who showed increases in these two indices from Essays 1 to 2.

Figure 5.2.

Differences in Syntactic Complexity among High-, Mid-, and Low-performers

(a) Mean length of T-unit

(b) Mean length of sentence

(c) T-unit complexity ratio

(d) Dependent clause ratio

(e) Complex nominals per clause

(f) Complex nominals per T-unit
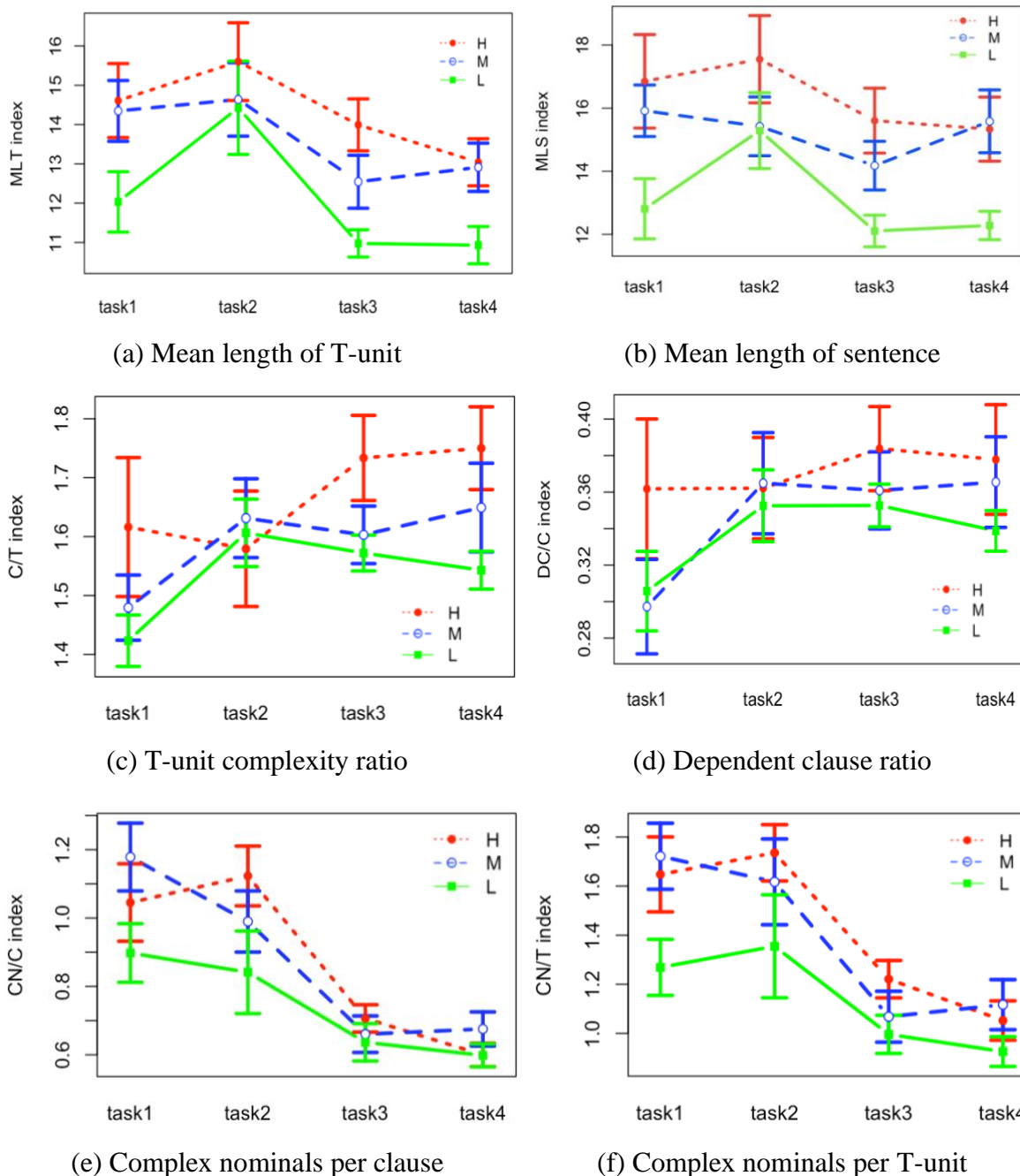
Note. Level of performance was defined as follows: High = scores of 11–12 on the averaged pre-test; Medium = 8–10.5; Low = 5.5–7.5.

5.5 Discussion

The study investigated how teacher, automated, and combined feedback influenced syntactic complexity in EFL students' revisions over the semester. In addition, the potential effect of students' levels of proficiency on the changes in the syntactic complexity in their texts written

during the course was also examined. I discuss my findings in light of previous research that studied the impact of feedback on students' writing complexity. Overall, this study demonstrates that writing complexity was unaffected by feedback, as reflected in the minimal variations between the students' initial drafts and revised texts. Similarly, no significant differences were found in the syntactic complexity of students' writing between pre- and post-tests. The findings concur with those of Evans et al. (2011) and Zhang and Cheng (2021) who found that the provision of WCF did not enhance students' syntactic complexity, and with those of Hartshorn and Evans (2015) who discovered no meaningful differences between the treatment and control groups for the measures of syntactic complexity, and with those of Xu and Zhang (2021) who contended that students' syntactic complexity remained unchanged following automated feedback. In another study, Wind (2022) examined the longitudinal development of syntactic complexity in an advanced writing course. Though students received feedback on their assignments during the semester, no improvements were detected in terms of syntactic complexity. Based on these findings, the author concluded that students might not have received the feedback that is the most appropriate feedback for their needs.

Looking at the T-unit complexity measures (MLT, C/T, and CN/T), all three indices showed a pattern of reduction in the students' revisions after the provision of teacher feedback. This finding corresponds to the previous studies in which the students with dynamic WCF exhibited a decrease in MLT, C/T, and CN/T from pre- to post-tests (Eckstein & Bell, 2021; Hartshorn et al., 2010). An explanation outlined in previous research is that students' attempts to improve accuracy may hinder the development of their syntactic complexity (Eckstein et al., 2020; Hartshorn et al., 2010). Eckstein et al. (2020) argued that L2 writers most probably use linguistically simplified structures to improve their linguistic accuracy. Similarly, Hartshorn et al. (2010) explained that, as students strive to improve their accuracy, complexity may be inhibited slightly by their careful monitoring of their writing.

Given that WCF does not support the development of syntactic complexity, I studied the degree of students' feedback acceptance, as I reasoned that their unsuccessful utilisation of feedback could be an underlying reason for the non-significant impact of feedback on syntactic complexity. However, this was not the case in my study. I found that students utilised feedback effectively in their revisions, resulting in 71.0% (teacher feedback), 76.2% (Grammarly feedback), and 61.8% (combined feedback) of correct revision. Moreover, they made notable improvement in their writing performance when the comparison was made from pre- (M = 8.98, SD = 2.09) to post-writing assessment (M = 10.46, SD = 1.96, $p = 0.006$) (for details, see

Chapter 4). These findings reflect how students utilised feedback in their revision and the general impact of feedback after a semester-long feedback treatment.

Although no significant improvements were found in the syntactic complexity of students' texts following the WCF intervention over a semester, this finding suggested that WCF did not result in structurally less complex writing. This is an important observation in line with Truscott (2007) who reasoned that WCF has a negative influence on syntactic complexity: students avoid complex structures due to the fear of making mistakes. Polio (2012b) also contended that students may ignore complexity in pursuit of accuracy. To put in another way, as teachers provide feedback on L2 writing as a means of improving writing accuracy, students are likely to fasten their attentional focus on how to rectify grammatical errors and produce more accurate revisions, probably resulting in structurally less complex writing. However, the findings from my study contradicted these assertions established in previous studies. My findings are in partial agreement with those previous studies (Hartshorn et al., 2010; Van Beuningen et al., 2012; Zhang & Cheng, 2021) which found that WCF did not lead to participants using fewer complex syntactic structures. As Zhang and Cheng (2021) explicitly stated, the result that WCF does not show any effects on students' syntactic complexity does not support the contention that it negatively affects syntactic complexity, as asserted by Truscott (1996, 2007). Taken together, it is reasonable to conclude that WCF did not negatively affect complexity in students' texts, although it did not result in complexity gains either.

5.6 Implications for L2 writing and pedagogy

The above results have implications for L2 writing and pedagogy. Based on the findings that students did not exhibit any significant improvements in most measures of syntactic complexity, WCF has a negligible effect on complexity in students' writing. Understanding these negligible effects of WCF can inform writing teachers that students' focus on producing accurate texts does not deviate their attention from complexity. This reassures L2 writing teachers that gains in one aspect of writing (i.e., accuracy) do not tend to be conflicted with another aspect of writing development. Thus, it is advisable for teachers to continue their feedback practices in their L2 writing classes, as providing feedback does not make students produce less complex writing.

However, these findings might be mediated by some feedback-related and task-related factors such as feedback sources, topic familiarity, or genres of writing. When students received Grammarly feedback on their writing (Essay 2), there was no difference between their

original and revised writing in the indices of C/T, DC/C, and CN/C. A likely explanation is that the scope of Grammarly feedback is limited to accuracy issues which in turn limits students' attention to the complexity of their writing. Also, non-significant differences in the draft and revised texts provide some indication of the potential influence of feedback sources on the complexity measures. Other points of discussion concern the influence of topic familiarity on syntactic complexity. As Abdi Tabari and Wang (2022) suggested, topic familiarity had a positive effect on syntactic complexity in students' writing. They conclude that L2 learners tend to deploy their subject-matter knowledge quickly when dealing with a familiar writing task and focus more on generating ideas and producing structurally more complex texts. In my study, although the writing tasks were taken from the curriculum, they were somewhat different from one another regarding the degree of topic familiarity. For example, the writing topic "The best teacher who inspired me" would probably be more familiar to the students compared to the topic "The worst teacher who discouraged me", as they had experience in writing about a person they admired at their secondary schools. However, I could not draw any valid conclusions about how topic familiarity supports syntactic complexity, as it is beyond the scope of my study.

Although not central to the purpose of the study, I reasoned that it is important to consider the impact of genre differences on syntactic complexity, as different genres tend to have different communicative and functional requirements which can result in different linguistic features (Lu, 2011; Yoon & Polio, 2017). In my study, students completed argumentative essays (Essays 1 and 2) and narrative essays (Essays 3 and 4) during the treatment sessions. Appendix F provides a visual representation of the syntactic complexity in students' writing across two genres of writing. Overall, the findings show higher complexity measures in argumentative texts than in narrative essays for all measures, except the two subordination measures. My results were similar to those in Yoon and Polio's (2017) publication which found that students' language was more complex in argumentative essays than in narrative essays based on length of production units and phrase-level complexity measures. Interestingly, the study found little genre effect on subordination measures, which is also true in my study.

Research is yet to be conducted on how multiple feedback sources affect syntactic complexity in students' writing, where different groups receive feedback from different sources (e.g., teacher or peer). Such a study may reveal how multiple feedback sources affect syntactic complexity over time. Moreover, future research that investigates the impact of WCF on syntactic complexity, especially a study using multiple drafts, would yield useful insights into

how feedback affects the sub-constructs of syntactic complexity in multiple rounds of feedback.

5.7 Conclusions

This study examined the effects of teacher, automated, and combined feedback on syntactic complexity in EFL students' writing. Overall findings revealed limited changes in the comparison pairs between the initial and revised texts and no significant differences between the pre- and post-tests. Specifically, length of production unit indices (MLT and MLS), C/T, and CN/T increased in the post-tests, but these improvements did not reach statistical significance. I also discussed the variations among the high-, mid-, and low-performing students when the comparison of their syntactic complexity was made in their revised writing throughout the course. The results suggest that the mere exposure to feedback is not sufficient to enhance the complexity of students' writing. Therefore, future research should take a more interventionist approach in which students are exposed to texts with higher syntactic features (e.g., model texts) and explore the impact on syntactic complexity.

Even though the study provided useful insights into the relationships between syntactic complexity measures and feedback, some limitations must also be acknowledged. For example, language development requires a longer observation period compared with that in the study. The findings may indicate a limit to the benefits of WCF on several subcomponents of syntactic complexity. Therefore, the need emerges for longitudinal research that investigates how WCF affects syntactic complexity in the long run. Future research could examine patterns of differences among students with high and low proficiency levels to provide a clearer picture of the impact of WCF on the construct. Moreover, previous research selected a limited range of complexity measures (typically one to four). However, assessing different aspects of complexity would be beneficial for capturing a comprehensive picture of L2 writing development (or lack thereof) due to the multi-dimensional nature of the construct.

Despite the small sample size, the study contributed important findings to the research on L2 writing and the practice of L2 writing pedagogy. The analysis of 270 texts added to the growing body of research on WCF that focuses on syntactic complexity and informed scholars about the importance of considering different aspects of writing in assessing learners' development. Moreover, the current findings provide pedagogical implications that contribute to the understanding of the influence of providing WCF on the writing complexity of students' texts over a semester.

From the writing assessment perspective, the study stressed the importance of automated tools in assessing the writing and development of EFL learners. For L2 writing teachers, a computational system for automatic analysis of syntactic complexity could facilitate the comparison of linguistic complexity of writing samples by assessing changes in the linguistic complexity of texts after a particular pedagogical intervention, or by monitoring students' linguistic development over a certain period. In the same manner, using computational tools can help L2 writing researchers understand the holistic aspects of students' L2 syntactic development at varying proficiency levels and evaluate the effectiveness of pedagogical interventions that aim to promote syntactic complexity development.

CHAPTER 6. HIGHER-PROFICIENCY STUDENTS' ENGAGEMENT WITH AND UPTAKE OF TEACHER AND GRAMMARLY FEEDBACK IN AN EFL WRITING COURSE

In this chapter, I present the results of the third study, which examined how students engaged with feedback from different sources in an EFL writing course in Hungary. A brief introduction to the concept of student engagement with feedback is followed by a discussion of gaps in previous research and the rationale of the present study. This is followed by a short description of the research context and student participants. After setting the research scene, the findings regarding the student engagement with feedback are presented. The final section draws conclusions from the findings which indicated students' moderate to low levels of student engagement with teacher and Grammarly feedback. I conclude by discussing the pedagogical implications of the study.

6.1 Introduction

Feedback provision in L2 writing is considered an important pedagogical practice that teachers use to help improve writing performance. However, the facilitative role of feedback has been widely debated, as it does not always fulfil this potential (Truscott & Hsu, 2008). A great deal of research on written feedback has investigated the relative effects of implicit and explicit corrective feedback on students' writing (Karim & Nassaji, 2018), compared feedback from different sources (Dikli & Bleyle, 2014; Yang et al., 2006), and explored students' perceptions of feedback (Huang & Renandya, 2020). The underlying hypothesis of most studies is that feedback enables learners to notice the mismatches between the target language and their interlanguage system (Van Beuningen, 2010). However, the mere provision of feedback does not always result in improvement, as several factors including the degree of student engagement with the feedback likely impact the benefits. For example, Zheng and Yu (2018) stated that if learners were not fully engaged with feedback, they were less likely to benefit from it. Zhang (2020) also argued that compared to a narrow focus on accuracy improvement in L2 student writing, it is more meaningful to examine how learners engage with feedback from different sources to enhance the possible benefits.

Student engagement with written feedback remains understudied compared with the mainstream feedback studies which investigated the effectiveness of different feedback strategies. Understanding student engagement with feedback, that is how and to what extent students respond to feedback, is an important issue (Ellis, 2010; Zheng & Yu, 2018), as it helps

build links that connect the provision of feedback and its effects on writing development. Students' feedback uptake, a form of engagement (Shen & Chong, 2022), is considered a basic parameter to understand how they incorporate feedback in their revisions. Moreover, authors researching this area reasoned that student engagement with feedback is influenced by contextual and individual factors (e.g., language proficiency). For example, previous studies found that learners at higher language proficiency levels are more likely to be better candidates for utilising feedback, as they tend to have adequate linguistic competence to address errors (Bonilla Lopez et al., 2017; Park et al., 2016). However, as those of higher proficiency wish to write with flexibility and creativity (Chen & Cheng, 2008), it is also likely that they will get frustrated by mechanical feedback, may question its accuracy, or incorporate it selectively (Zheng & Yu, 2018), which will result in reduced willingness and ability to use feedback for error correction. This might be the case with automated feedback which primarily flags low-order issues in student writing. Therefore, more research is needed to examine how high proficiency students respond to form- and meaning-focused feedback from different sources during the revision process. This would inform writing teachers about how they can integrate feedback from different sources into their practice and the areas where students need more assistance to make effective use of feedback. Also, examining how differently these students engaged with teacher and automated feedback could provide a space where Grammarly can be used to target specific error types it detects accurately as a forerunner of teacher feedback. To this end, this study investigated how Hungarian EFL university students engaged with teacher and automated feedback and the extent to which their engagement led to successful feedback uptake.

6.2 Statement of problem

Prior research has examined how and why individual students engage or do not engage with multiple feedback sources; students in such studies received only one type of feedback. However, with the availability of multiple sources of feedback for learners nowadays, more research is needed to examine how students engage with these sources in a single study (Zhang & Hyland, 2018). In this way, L2 writing researchers could better understand how these feedback sources could complement each other effectively if used sequentially. Moreover, little effort has been made to investigate how students at higher-proficiency levels engaged with teacher and automated feedback. This study was designed to fill these gaps: it examined how students engaged with form- and meaning-focused teacher and Grammarly feedback.

Extending the line of previous research, this exploratory study examined how Hungarian university students engaged with teacher and automated feedback, and their feedback uptake. Comparison of teacher and Grammarly feedback was made with the intention to understand feedback scope and how students engaged with two feedback sources. In addition to the influence of form-focused feedback previously examined by Thi and Nikolov (2021b) in the Myanmar EFL context, the effect of meaning-focused feedback was also investigated, as feedback in a writing course typically covers both linguistic and rhetorical aspects of writing (Cheng & Liu, 2022). As the present study was conducted in an English writing skills course, the writing tasks and assessment rubrics were different from those in the previous study. Thus, my research is guided by the three research questions as follows:

RQ1. What feedback do the students receive from their teacher and Grammarly on their L2 writing?

RQ2. In what ways do the students respond to form-focused feedback from their teacher and Grammarly when revising their second drafts?

RQ3. In what ways do the students respond to meaning-focused feedback from their teacher when revising their final drafts?

6.3 Methodology

6.3.1 Context

Students were enrolled in a Reading and Writing Skills II course which aims to develop their reading and writing skills and help them prepare for the corresponding components of the Proficiency Examination (see also Horváth, 2016). The proficiency exam, also known as a PE exam, is a C1 level exam assessing all four language skills that students have to take at the end of the first academic year upon their completion of the three obligatory courses (i.e., Listening and Speaking Skills I-II, Reading and Writing Skills I-II, and English Grammar in Use I-II). It includes three test components: writing, listening, and oral components. The written component consists of a grammar and usage test, a reading test, and a writing test. Upon successful completion of the written component, students are eligible to take a listening comprehension test and an oral exam. In regard to the writing portion of the test, students selected one of two themes and are instructed to discuss in detail each of the four topics required under the chosen theme. As part of the assessment process, students are informed of the four criteria that will be used to evaluate their work: task completion, vocabulary, structures, and coherence (Horváth, 2016).

The Reading and Writing Skills II course was offered in a ninety-minute session on a weekly basis. As part of the course requirements, students were asked to write three PE essays during the course on Weeks 5, 9, and 12. The course provided guidance and practice in a number of areas including formulation, proper usage, structuring, vocabulary extension and working with different genres in reading and writing. Various writing strategies including narrowing down your topic, prewriting, and using a logical structure were introduced. Moreover, writing tasks which correspond to different written genres (e.g., description essay, classification essay, and narrative essay) were practiced throughout the course.

6.3.2 Participants

The participants were 31 first-year undergraduate students (aged between 18 and 26 years) in a BA English Studies programme. They were enrolled in two online English Reading and Writing Skills II courses and their class teacher in the spring semester of the 2020/2021 academic year at a university in Hungary. Because of the COVID-19 pandemic, this course was delivered online via Microsoft Teams over a 14-week semester. The class teacher completed his PhD in Applied Linguistics and has been teaching this course for five years. The students were Hungarian native speakers, and their English proficiency was supposed to be at upper intermediate (B2) level which is the entry level of the programme. According to the levels specified in the *CEFR* scale, those at B2 proficiency levels "can write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesizing and evaluating information and arguments from a number of sources" (Council of Europe, 2018, p. 75). Data from the students' background questionnaires revealed that 75% rated their English writing proficiency levels as upper-intermediate and advanced, whereas 25% estimated it as lower-intermediate. Many noted that they had difficulties with vocabulary, coherence, organization, and advanced structures in English writing.

6.3.3 Instruments

Writing tasks: As part of the course requirements, students wrote three PE type essays in which they gave a narrative account of their personal experiences or shared their views on proposed statements with justifications (Figure 6.1). In each essay, they were presented with four prompts that were similarly structured, and they could select one from two tasks (Task A or Task B). Students made revisions based on the feedback from the teacher, Grammarly, or both

before the final submission of two essays out of three at the end of the course as part of their portfolio.

Figure 6.1.

Sample PE Essay Task used in Week 5

Write a four-paragraph long essay (cover each point in a separate paragraph and do not write introduction and conclusion paragraphs) in about 300 to 400 words using one of the essay options.

Proficiency Exam type essay #1

Task A: Parents just don't understand
Write an essay about your views on parent-child generational differences. In the text,
• introduce a common problem caused by generational differences;
• discuss how the two generations view the same problem;
• explain your attitude towards generational differences;
• tell a story about how you could or could not solve a problem caused by generational differences with your parents.

Task B: 'Tech, I love you'
Write an essay about your views on a technological device essential for you. In the text,
• introduce a device you find difficult to live without;
• discuss how your life has changed since you started using that device;
• explain your attitude towards technology addiction;
• tell a story about an event where your favorite device was or was not helpful.

6.3.4 Research procedure

The study used a naturalistic classroom-based approach to examine the students' behavioural engagement with feedback from different sources in an online EFL writing course. Data collection took place over a 14-week semester from February to May 2021 (Table 6.1): I collected the students' drafts, teacher and Grammarly feedback on the drafts for further analysis, and their revised texts in file format. The students were also given a questionnaire on their language background, especially L2 writing experiences and self-assessed writing proficiency. For the purpose of the study, I analysed the two writing tasks (Weeks 5 and 9) in which they received teacher feedback on the first essay and Grammarly feedback on the second one.

Table 6.1

Timeline of Data Collection

Timeline	Data collection procedures
Week 1-3	Introduction to the Reading and Writing II course
	Academic reading fundamentals
	Essay writing fundamentals
Week 4	Introduction to the research project
	Language background questionnaire
Week 5	PE essay I

Week 6-7	Essay I returned with teacher feedback
Week 9	PE essay II
Week 10-11	Essay II returned with *Grammarly* feedback
Week 12	PE essay III
Week 13	Essay III returned with the combined teacher and *Grammarly* feedback
Week 14	Portfolio submission

6.3.5 Data analysis

Having collected the written data from the students, I first categorised the teacher feedback into form-focused (stating the surface-level errors relating to the choice and appropriacy of grammar and vocabulary) and meaning-focused feedback (expressing multiple-sentence level issues such as task completion, organization, cohesion, and coherence which take the form of marginal or end-of-text comments) (for a detailed discussion, see Section 3.4.1).

As automated feedback tends to be fallible (John & Woll, 2020), I reasoned that this might affect how students engaged with Grammarly feedback. Therefore, I examined the accuracy of Grammarly feedback following the work of Ranalli (2021) and classified it into three categories: (i) accurate, (ii) inaccurate, and (iii) indeterminate feedback (see Appendix G). Accurate feedback refers to a flagged error which represents the label (i.e., error category) applied to it or a legitimate issue that needs correction. Inaccurate feedback includes issues where Grammarly mistakenly flags a correct form as an error. Indeterminate feedback comprises issues relating to conciseness, wordiness, politeness, and clarity which do not require urgent correction or cases where Grammarly provided no specific suggestions for addressing highlighted texts (e.g., clarity issues).

6.4 Findings

In the next section, I report the findings on students' engagement and uptake of teacher and Grammarly feedback. Before detailing the results, it is important to understand the areas of writing issues these feedback sources address. Therefore, I present a summary of teacher and Grammarly feedback on students' writing. Then, I report how students engaged with form-focused and meaning-focused feedback. The closing discussion focuses on the role of students' proficiency levels in understanding their engagement with feedback.

6.4.1 Teacher and Grammarly feedback on students' first drafts

The first research question was addressed by presenting the type of feedback that the students received in their EFL writing course. Noting the differences in the nature of teacher and automated feedback, a distinction between form-focused and meaning-focused feedback was made. Figure 6.2 shows the distribution of form-focused feedback in the students' first drafts. Overall, both teacher and Grammarly feedback addressed a few error categories, ranging from language conventions (e.g., spelling, punctuation, and grammar) to other writing issues related to word choice, conciseness, and formality. The teacher offered 107 feedback points, whereas Grammarly flagged 481 errors in the students' texts ($n = 31$). Most form-focused feedback concerned punctuation, miscellaneous errors, and word choice across the two modes of feedback. Further results suggested that the teacher feedback addressed more error categories than automated feedback, although differences in the two writing tasks may have influenced the number of occurrences of errors and length of texts.

Figure 6.2.

Comparison of Form-focused Feedback Provided by Teacher and Grammarly
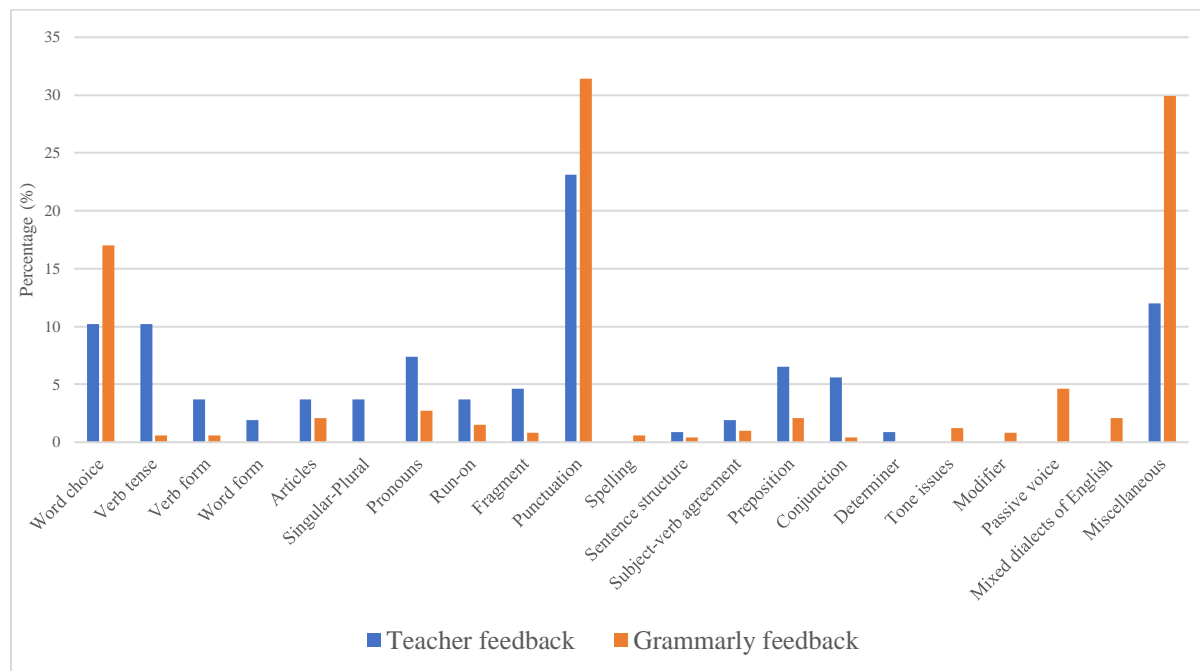
A detailed analysis of the two modes of feedback revealed that the teacher paid uniform attention across language-related issues, whereas Grammarly feedback primarily targeted punctuation, word choice, and miscellaneous errors (i.e., writing issues relating to conciseness, wordiness, and clarity). Although the teacher feedback was superior to automated feedback in terms of feedback scope, Grammarly provided several suggestions on how to address an error,

allowing students to filter suggestions and make informed choices: either to incorporate or reject the feedback (Table 6.2).

Table 6.2

Example of Grammarly Suggestions

Student original	Grammarly suggestions	Student revision
The easiest way to do it is by passing on them to the next generation and encouraging them to follow these customs.	1. Delete "by" (wordy sentence – clarity) 2. Replace "by passing" with "bypassing" (confused word – correctness) 3. Replace "passing on them" with "passing them on" (misplaced word – correctness)	The easiest way to do it is by passing them on to the next generation and encouraging them to follow these customs. (Correct revision)

Other differences were noted in terms of the ways in which the teacher and Grammarly provided feedback on a written text. Whereas teacher feedback took the form of in-text error corrections and marginal comments, Grammarly offered in-text indirect feedback followed by end-of-text direct feedback (Figures 6.3 & 6.4).

Figure 6.3.

An Example of Form-focused Teacher Feedback
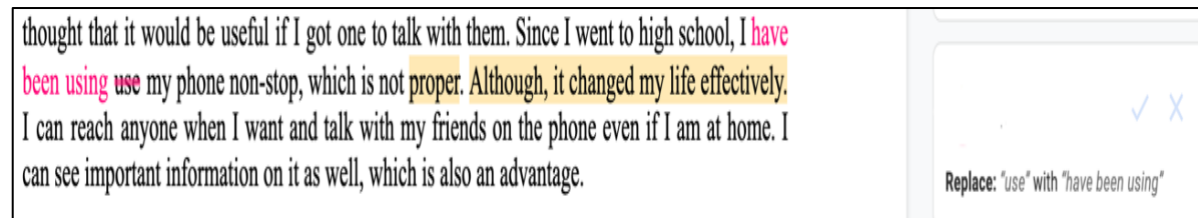
Figure 6.4.

Examples of Form-focused Feedback Generated by Grammarly
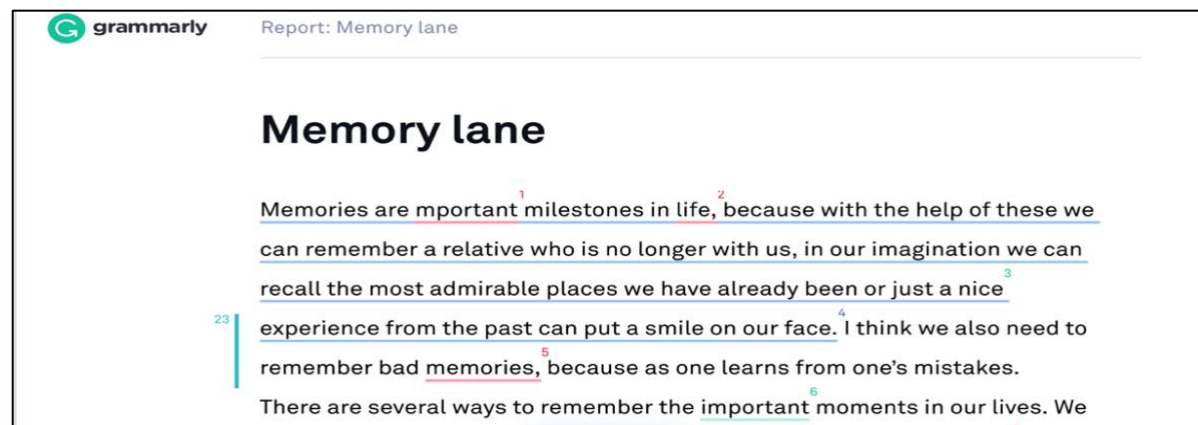
a) An Example of In-text Indirect Feedback

b) An Example of End-of-text Direct Feedback

In terms of accuracy of corrective feedback generated by Grammarly, Table 6.3 shows that it was highly accurate (44.1%) in detecting a wide range of errors including errors in verb tense, verb form, article, pronoun, run-on, sentence structure, preposition, conjunction, and modifier. These findings indicate that Grammarly can be integrated in writing classes to target specific error types it performed well on. Few inaccurate flaggings (5.6%) were attributed to passive voice, fragment, and subject-verb agreement errors. However, indeterminate feedback (50.3%) outnumbered accurate and inaccurate flaggings, indicating that Grammarly flagged clarity, tone, and conciseness issues more frequently (See Appendix H).

Table 6.3

Accuracy of Grammarly Feedback

	Grammarly feedback (n/ %)		
	Accurate Feedback	Inaccurate feedback	Indeterminate feedback
Word choice	0	0	82/100
Verb tense	3/100	0	0
Verb form	3/100	0	0
Word form	-	-	-
Articles	10/100	0	0
Singular-Plural	-	-	-
Pronouns	13/100	-	-
Run-on	7/100	0	0
Fragment	3/75	1/25	0
Punctuation	148/98	3/2	0
Spelling	3/100	0	0
Sentence structure	2/100	0	0

Subject-verb agreement	4/75	1/25	0
Preposition	10/100	0	0
Conjunction	2/100	0	0
Determiner	-	-	-
Tone issues	0	0	6/100
Modifier	4/100	0	0
Passive voice	0	22/100	0
Mixed dialects of English	0	0	10/100
Miscellaneous	0	0	144/100
Total	212/44.1	27/5.6	242/50.3

More important, and central to the findings of the study, is that apart from teacher feedback on language issues, the students also received meaning-focused feedback. This type of feedback took the form of marginal and end-of-text comments and reflected different criteria of the writing assessment scale. The teacher's end-of-text comments followed certain patterns: the first part (or sentence) of the comment diagnosed the issue in the essay and the second part (or sentence) built upon the earlier statement followed by possible impact on the improvement of the essay (Figure 6.5). Among the four categories of the teacher's commentary feedback, giving information (41.5%) and suggestions (39%) ranked the highest (Table 6.4). These types of comments provide information about the aspects of writing students are good at and those they need further improvement. Though positive feedback or praise serves to minimize the force of criticism when responding to student written work, it only accounted for 14.6 per cent of the total content feedback in the present study; this outcome reflects minimal acknowledgement of the students' good performance.

Figure 6.5.

Two Examples of Teacher's Meaning-focused Feedback

a) An Example of Teacher's Marginal Comment

You have the makings of a great essay here but some of your paragraphs are a bit underdeveloped (like paragraph 3) and would benefit from more support. Improving these issues will also increase your task completion and coherence scores.

b) An Example of Teacher's End-of-text Comment

Table 6.4

Types and Ratio of Teacher's Meaning-focused Feedback on L2 Writing

Feedback type	Teacher feedback (N/%)
Giving information	17/41.5
Asking for information	2/4.9
Praise	6/14.6
Suggestion	16/39.0
Total of comments	41/100%

6.4.2 Behavioural engagement with teacher and Grammarly form-focused feedback

To address the second research question examined the behavioural engagement with the form-focused feedback through analysing the students' uptake of feedback. The findings showed that they used four revision operations: correct revision, incorrect revision, no revision, and deletion (Tables 6.5 and 6.6). Figure 6.6 illustrates the comparison of the percentage of overall changes in the revisions. Generally, the results suggested that students engaged with both teacher and Grammarly feedback as reflected in their revision operations and feedback uptake; however, the degree of engagement varied across the two feedback modes. Particularly, of the 107 feedback points from the teacher, 53 (49.5%) were considered for revision regardless of whether they led to correct or incorrect revision outcomes. In the case of Grammarly, students considered 138 (28.7%) feedback points for revision out of 481 flagged errors. Although the comparison of revision ratios suggested that students' uptake of teacher feedback tended to be higher than that of Grammarly feedback, the fact that 129 error flaggings resulted in successful revision indicated how helpful they found Grammarly feedback and how consciously they engaged with it (for a detailed discussion, see Section 6.5).

Table 6.5

Examples of Teacher Feedback and Student Revisions

Student original	Teacher feedback	Student revision
I have been living my life on my phone since the Pandemic started last March. Before that, I ~~had~~ used it less than I'm using it now.	Delete "had"	I have been living my life on my phone since the Pandemic started last March. Before that, I <u>used</u> it less than I am using it now. (Correct revision)
Despite how useful a cell phone can be, it can be <u>so harmful</u> as well.	You can easily use a much better phrase here.	Despite how useful a cell phone can be, it can be <u>disadvantgeous</u> as well. (Incorrect revision)
I use it for storing pictures, important notes, <u>putting on</u> reminders, and watching videos.	"Scheduling" would be a better choice here.	I use it for storing pictures, important notes, <u>putting on</u> reminders, and watching videos. (No revision)
I took pictures of what I liked, I recorded songs that I ~~singed~~, and I also played Snake on it.	Replace "singed" with "sang"	Deleted the sentence in the revision

Table 6.6

Examples of Grammarly Feedback and Student Revisions

Student original	Grammarly feedback	Student revision
If somebody <u>would give</u> me the opportunity to go back in time and relive a memory of mine, I would say yes.	Replace "would give" with "gave"	If somebody <u>gave</u> me the opportunity to go back in time and relive a memory of mine I would say yes. (Correct revision)
I decided to relive that memory of mine because I am no longer friends with them but <u>in spite of</u> that, I was carefree and I was living my best life back in those days.	Replace "in spite of" with "despite"	I decided to relive that memory of mine because I am no longer friends with them but <u>despite of</u> that, I was carefree, and I was living my best life back in those days. (Incorrect revision)
More firms <u>increases</u> the amount of choice and it will reduce our customers.	Replace "increases" with "increase"	More firms <u>increases</u> the amount of choice and it will reduce our customers. (No revision)
I think the most common key to preserve memories is talking about past events, such as festivals, holidays, trips etc. with the ones you have been <u>there with</u>.	Replace "there with" with "therewith"	The most common key to preserve memories is talking about past events, such as festivals, holidays, trips, etc., with the ones you have been. (Deleted "there with")

A closer examination of students' revised essays showed that 44 (41.1%) teacher feedback points and 129 (26.8%) Grammarly error flaggings resulted in correct revisions. The low percentage of students' incorrect revisions (3.7% and 1.5%, respectively) reflected their adequate linguistic competence to do correct revisions once they decided to use the feedback.

However, the fact that most errors were left unrevised (50.5% and 71.3%, respectively) across the two modes of feedback raised critical issues ranging from the extent to which students benefitted from feedback to possible reasons for their disregard of feedback.

Figure 6.6.

Revision Operations of Teacher and Grammarly Form-focused Feedback
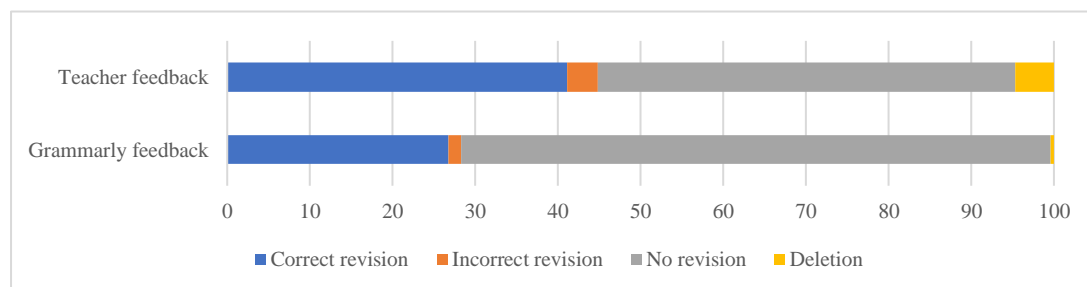
Due to the moderate acceptance of teacher feedback in their subsequent texts, I further analysed how students acted upon different feedback categories (Table 6.7). Overall, the revisions showed that they did not blindly accept feedback but made selective and cautious decisions about feedback uptake. More precisely, I found that students successfully corrected four common error types identified by the teacher: run-on sentences (100%), sentence structure (100%), determiners (100%), and articles (75%). However, most errors relating to conjunction (83.3%), punctuation (72%), and verb tense (54.5%) were left unattended. Teacher feedback on verb form, word form, singular-plural, and subject-verb agreement errors resulted in 50% correct and 50% incorrect revisions. These findings shed light on the complexity of students' revision patterns and indicate how selectively they responded to teacher feedback.

Table 6.7

Comparison of Students' Revision Operations by Error Type

	Teacher feedback (N/%)				Grammarly feedback (N/%)			
	CR	IR	NR	D	CR	IR	NR	D
Word choice	6/54.5	1/9.1	4/36.4	0	29/35.4	0	53/64.6	0
Verb tense	5/45.5	0	6/54.5	0	2/66.7	0	1/33.3	0
Verb form	2/50	0	1/25	1/25	0	0	3/100	0
Word form	1/50	0	1/50	0	-	-	-	-
Articles	3/75	0	1/25	0	4/40	0	6/60	0
Singular-Plural	2/50	0	2/50	0	-	-	-	-
Pronouns	3/37.5	0	3/37.5	2/25	2/15.4	0	11/84.6	0
Run-on	4/100	0	0	0	1/14.3	0	6/85.7	0
Fragment	1/20	0	2/40	2/40	0	0	4/100	0

Punctuation	4/16	3/12	18/72	0	41/27.2	2/1.3	107/70.9	1/0.7
Spelling	0	0	0	0	2/66.7	0	1/33.3	0
Sentence structure	1/100	0	0	0	0	0	2/100	0
Subject-verb agreement	1/50	0	1/50	0	1/20	0	4/80	0
Preposition	4/57.1	0	3/42.9	0	4/40	0	6/60	0
Conjunction	1/16.7	0	5/83.3	0	1/50	0	1/50	0
Determiner	1/100	0	0	0	-	-	-	-
Tone issues	-	-	-	-	1/16.7	1/16.7	4/66.7	0
Modifier	-	-	-	-	0	0	4/100	0
Passive voice	-	-	-	-	3/13.6	2/9.1	17/77.3	0
Mixed dialects of English	-	-	-	-	0	0	10/100	0
Miscellaneous	5/41.7	0	7/58.3	0	38/26.4	2/1.4	103/71.5	1/0.7
Total	44/41.1	4/3.7	54/50.5	5/4.7	129/26.8	7/1.5	343/71.3	2/0.4

Note. CR: correct revision, IR: Incorrect revision, NR: no revision, D: Deletion. Percentages represent frequencies of revision categories within each error category. For instance, 54.5% of the word choice errors had a correct revision rating.

As for Grammarly feedback, the students engaged with it far less: they mostly rejected the feedback regardless of what the linguistic errors were. Apart from higher successful revision rates on verb tense (66.7%), spelling (66.7%), and conjunction errors (50%), most errors flagged by Grammarly (71.3%) led to no revision. For example, students refused to act upon all feedback on verb form (100%), modifier (100%), run-on sentence (85.7%), pronouns (84.6%), subject-verb agreement (80%), passive voice (77.3%), miscellaneous errors (71.5%), punctuation (70.9%), and many others. Despite getting specific suggestions on how to revise errors (i.e., mechanical revision without deep engagement), students seemed unwilling to act upon the feedback they got. Within the four areas of writing issues suggested by Grammarly (see Section 3.3.2), the uptake rates were the highest in engagement (31.31%) which included issues in word choice and sentence variety, whereas the opposite was true in delivery (13.4%) which considered tone issues, the use of sensitive language, and inappropriate colloquialisms. Cross-referencing their revision operations with the accuracy of Grammarly feedback suggested that students were selective in their utilization of automated feedback. Particularly, the finding that Grammarly flagged errors relating to conciseness, wordiness, politeness, and clarity (50.3%) led to a lower ratio of feedback acceptance. Moreover, all passive voice errors were mistakenly flagged which resulted in reduced willingness to integrate the feedback into revisions. Figure 6.7 illustrates an example of how Grammarly flags a passive voice error in a student's writing.

Figure 6.7.

Passive Voice Error Flagged by Grammarly
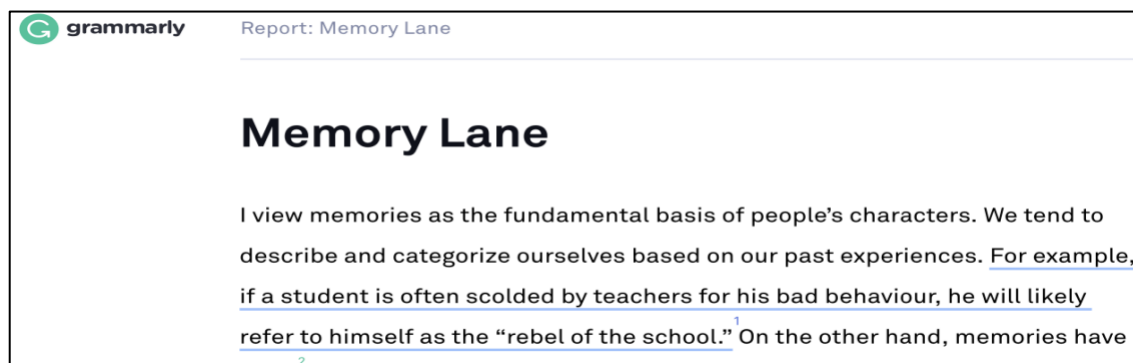
6.4.3 Behavioural engagement with meaning-focused teacher feedback

In addition to form-focused feedback, I analysed the students' behavioural engagement with the teacher's feedback on meaning (Table 6.8). Figure 6.8 illustrates the students' revision patterns. Overall, a high ratio of teacher's comments (64%) was not considered in the students' revised essays. These unattended comments might be attributed to students' low engagement with meaning-level feedback or their partial understanding of the teacher commentary feedback. For example, the teacher's comments on students' texts tended to be vague or obscure (e.g., *Your essay has valid ideas, but some paragraphs need to be revised. Take a look at how you could connect the ideas found in your first two paragraphs.*). This may partly explain why they failed to integrate the feedback into their revisions. Furthermore, the fewest attempts were made for substantive revision on their essays (4.0%), although 24% of the comments were considered for minimal revision.

Table 6.8

Examples of Student's Minimal and Substantive Revision Changes

Student original	Teacher feedback	Student revision
Last year I had to go to orientation day because I have never been to Pécs before, [] I decided to go sightseeing.	Add "however,"	Last year when I first came to the city, I decided to go sightseeing by myself. (Minimal revision)
I was nine years old when I got my first flip phone and I shared it with my twin sister. She used it more because she went to music classes in the nearby town and the phone was just a device to feel safe. At that time, I simply used it to play Snake on is and listen to	In this paragraph, you should focus on one device only.	I was around twelve years old when I got my first smartphone which changed my life completely. It made me feel more adult-like because in my mind only adults could use such devices. With my phone, I got free access to the Internet and it caused some arguments with my parents, however, those could have been prevented if they had talked about their reasons behind their

music while I was doing something. I changed when I got my first own phone. It was still not a smartphone, but I was happy to have my phone. I took pictures of what I liked, I recorded songs that I singed, and I also played Snake on it. Maybe I used my phone mostly in useless ways as any young teenager would have done it but it changed my life because I felt more mature.

actions. The problems lasted for a few years in our family until I learnt that "with great power comes great responsibility". Not every memory of my phone are positive ones and still, I love it and cannot imagine my life without a smartphone. (Substantive revision)

Figure 6.8.

Revision Operations of Meaning-focused Teacher Feedback
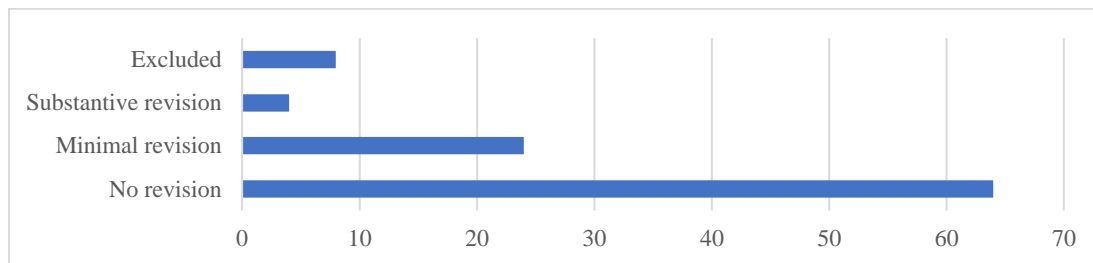
6.5 Discussion

This study examined the focus of teacher and automated feedback and how Hungarian English majors incorporated feedback into their revisions over the semester. Overall, I found that teacher feedback was directed at a variety of targets including task completion, vocabulary, structures, and coherence. Grammarly feedback focused on language-related issues, offering multiple suggestions on how an error should be revised. This encouraged students to rely on their prior linguistic knowledge and make informed decisions.

As for the engagement with form-focused feedback from the teacher and Grammarly, an investigation of students' uptake rates revealed that they incorporated more teacher feedback than automated feedback into their revisions, although Grammarly provided many more grammar and mechanics level feedback than the teacher. These findings corroborate results of previous studies showing that students' uptake rates of teacher feedback was much higher than that of automated feedback (Shi, 2021; Tian & Zhou, 2020). A plausible explanation is that the students must have viewed their teacher as an expert with a better understanding of their strengths and weaknesses in L2 writing. Also, they must have integrated the teacher feedback knowing that their tutor would assess their submissions along the line of assessment criteria.

Regarding students' engagement with Grammarly feedback, it should be stressed that factors such as the need to evaluate accuracy and flagged issues related to clarity and conciseness without giving specific suggestions seemed to hinder their feedback uptake. In previous studies (e.g., Koltovskaia, 2020; Ranalli, 2021; Zhang & Hyland, 2018), L2 writing scholars expressed their doubts about the accuracy of automated feedback which can influence students' decisions to reject it. Likewise, frequent flaggings of side issues may contribute to reduced willingness to integrate feedback. With these considerations in mind, it is reasonable to point out that students engaged with Grammarly feedback and found it helpful, as reflected in their successful error corrections. The advice they did not follow must have been due to their distrust arising from the nature and fallibility of Grammarly feedback; thus, they utilized its suggestions selectively.

Overall, the results indicated that students tended to respond to automated feedback in a selective or focused manner. It should also be stressed that the results were likely influenced by their language proficiency level. As suggested by Chen and Cheng (2008), learners at a more advanced language proficiency tend to write with more flexibility and creativity. At the same time, they need feedback focusing on meaning rather than machine-controlled feedback on language. As Grammarly feedback failed to focus on meaning, it is very likely that participants did not find it conducive to scaffolding their advanced writing skills, resulting in low feedback uptake. A second explanation might lie in the different instructional contexts. In the present study, students' learning of L2 writing was driven by the writing proficiency exam targeting C1 level at which emphasis was placed on task completion as well as other assessment criteria (i.e., vocabulary, structures, and coherence). If students earn only one of three points on task completion as the first criterion, other criteria are not assessed further, and they need to retake the exam. These instructional practices might partly explain why the form-focused feedback from Grammarly was not more appreciated.

Although it was unsurprising to find that the teacher paid more attention to language-related issues (72.3%) because writing tasks were primarily language-focused, it is worth noting that such feedback led to moderate levels of successful uptake (41.10%). This finding is not in accord with those of previous studies in which the participants showed successful uptake in their revisions between 80% to 85% of the time when they revised their texts in response to surface-level teacher feedback. Part of the explanation has to do with the instructional practices (e.g., the purpose of revisions and cause-and-effect relationships between students' revisions and their writing score improvement). For example, if students are unaware of the benefits of revising their work, it is unlikely that they would expand much effort on improving their work

aligned with the teacher feedback. In the present study, the prescribed curriculum (Anker, 2010) stresses the importance of revising and editing a text and provides follow-up exercises to practice revising for unity, details, and coherence. Students' beliefs and awareness about the value of revising texts, however, should be explored further. Another explanation could lie in the students' poor knowledge of strategies by which the feedback could be used (Winstone et al., 2017). This is particularly true when responding to meaning-focused feedback suggesting substantial changes in students' essays (e.g., reorganizing a paragraph, justifying an argument, or adding more supporting details). Out of other possible reasons of students' low uptake rates, follow-up discussions with the class teacher revealed that "being fed up with online learning" might be one of the contributing factors impacting students' engagement with feedback, as most students spent over three semesters in the virtual learning environments due to Covid 19 (personal communication, Simon, October 11, 2021).

An examination of students' engagement with meaning-focused teacher feedback suggested that they mostly failed to integrate the feedback reflected in their unsuccessful revisions. These findings contradict results of some previous studies (Ferris, 1997; Nurmukhamedov & Kim, 2010) which found that commentary types including imperatives and hedging comments were associated with substantive and effective revisions. Although the reason for this tendency should be investigated in follow-up student interviews, I was not able to do so, as students were overloaded with their schoolwork and courses. Instead, a follow-up interview with the head of the department was feasible. It suggests that individual differences (e.g., student agency, beliefs, and other motivational aspects), contextual factors including students' increased workload, and difficulties in understanding how to respond to meaning-level feedback are possible explanations impacting their levels of engagement with feedback (personal communication, Lugossy, October 12, 2021). Moreover, I found that the students incorporated more form-focused feedback than meaning-focused feedback in their revisions. This outcome is consistent with another study: Dressler et al. (2019) also found that the students addressed surface-level feedback focusing on writing mechanics rather than meaning-level feedback on argumentation, flow, and content.

Considering Grammarly's high accuracy in detecting and correcting errors in verb tense, articles, prepositions, and conjunctions, as well as students' high incorporation rates from such feedback (40-66.7%), it is reasonable to suggest that Grammarly should be used in a selective fashion to target specific error types it flags reliably. As an alternative, a prior review with Grammarly feedback before teacher feedback may offer teachers more opportunities to deal with higher-level concerns. Used this way, Grammarly can reduce teacher feedback burden,

functioning as a precursor to teacher feedback, and the focus of teacher feedback can be shifted from a concern over mechanical accuracy to coherence and content development through a recursive process of writing and rewriting (Yu et al., 2020).

An additional finding, although not central to the focus of the present study, indicated that the students made self-initiated revisions beyond the teacher feedback (Table 6.9). Also, there are many examples in which the students made correct revisions without following the feedback identified in their work. These changes reflect their autonomous learning, confidence in their prior linguistic knowledge, and their in-depth decision-making processes on whether to follow or reject feedback. Therefore, whether the students showed high or low levels of engagement with the feedback seems to be determined not only by the correct revision rates but also by considering the ways in which they revised their errors. In other words, the kind of mechanical revisions (i.e., correcting errors that are explicitly identified and corrected by teacher and Grammarly) can be done automatically without deep engagement as opposed to responding to more implicit feedback such as circling, underlining, questioning, or commenting.

Table 6.9

Examples of Student's Self-initiated Revision Changes

Student original	Student revision
It's small, easy to carry around but has <u>everything that I would ever need</u> from technology.	It's small, easy to carry around but has <u>what I need</u> from technology.
I don't remember exactly how my first phone changed my life but I'm sure it <u>made some things easier and some harder</u>.	I don't remember exactly how my first phone changed my life but I'm sure it <u>made things easier or harder at times</u>.
Calling my parents or my friends was done with two buttons, <u>I could listen to music anywhere</u> and if I didn't know something, I simply searched it up.	Calling my parents or my friends was done with two buttons, [] and if I didn't know something, I simply searched it up.

6.6 Conclusions

This study explored first-year undergraduate students' behavioural engagement with feedback from teacher and Grammarly in the Hungarian university context. Adopting a naturalistic classroom-based approach, it analysed data obtained from 31 undergraduates and their writing teacher to study how these students responded to feedback provided by the teacher and Grammarly on their L2 texts. Moderate level of behavioural engagement was found in their

revised drafts that were observable outcomes of their engagement. It should also be noted that selective and self-initiated revisions beyond the feedback indicated their adequate linguistic knowledge on how to revise errors correctly.

The findings offer implications for better utilization of feedback from different sources. First, the study reminds us that despite the students' intermediate (B2) level proficiency in English, it is likely that facilitation is needed to ensure their active engagement with feedback, especially concerning meaning which requires reworking of students' entire texts. Therefore, identifying revision strategies and engaging them in learning-oriented activities that facilitate their understanding of meaning-level feedback would help them clarify feedback information they receive and increase awareness about addressing teachers' marginal comments on coherence and content development. Second, students' awareness of how engagement with feedback informs learning should be promoted so that they remember that feedback is provided to their writing to scaffold their noticing and improve their L2 writing autonomously. Finally, the study indicates that the integration of Grammarly into advanced writing classrooms would benefit students and teachers in meaningful ways: Grammarly is highly accurate in flagging specific error types common in L2 student writing and can afford students with partial autonomy by allowing them to filter suggestions and make informed choices. Also, writing teachers can rely on Grammarly in a selective manner and shift their feedback focus to higher-level writing concerns that Grammarly does not do. Teachers should remind students to be cognizant of the fallibility of automated feedback and false alarms that may arise from the system. Taking these caveats into account, future studies should consider investigating the fallibility of automated feedback and factors underlying students' disregard for automated feedback, including distrust arising from inaccurate feedback.

Although the study has meaningful outcomes, its limitations are obvious. I recruited a small number of students at similar L2 proficiency level, so findings cannot be generalized to a student population at different proficiency levels, but the emerging issues go beyond this study. Thus, future research should consider involving a more varied range of learners and compare their high/low engagement with feedback to examine the possible relationships between learners' proficiency and engagement with feedback. It was not feasible to explore the reasons underlying students' low uptake of feedback, rendering it difficult to understand how students decide to engage or not to engage with the feedback. Thus, learners' factors, instructional contexts, and other barriers to feedback engagement and uptake that hinder students' use of feedback should be examined through self-reports in future research. To provide better profiles of L2 writers with varying levels of engagement, qualitative research with multiple sources of

data including in-depth interviews could be conducted. Such additions would improve the understanding of L2 writers' individual differences. Though my study targeted a single dimension of engagement with feedback, further research should scrutinize all-inclusive perspectives of engagement so that relationships among these dimensions can also be observed.

CHAPTER 7. INVESTIGATING SYNTACTIC COMPLEXITY AND LANGUAGE-RELATED ERROR PATTERNS IN EFL STUDENTS' WRITING

This chapter presents the results of the fourth study: it examined syntactic complexity and language-related errors in EFL students' writing. Particularly, I investigated how automated measurement of syntactic complexity helped distinguish writing proficiency of students from two higher education institutions in Myanmar and Hungary. Moreover, I examined language-related errors in students' writing to indicate the differences in the error patterns in the two groups. In consideration of these purposes, I reviewed previous research on the importance of studying learner corpora and language-related errors in L2 writing (for a review on syntactic complexity and writing proficiency, see Section 2.4.2). Following this, the research instruments and data analysis used in this study are presented. A discussion and a summary of the findings about differences in syntactic complexity and language-related errors in students' texts conclude the chapter.

7.1 Introduction

Across 21st century competency frameworks around the world, most researchers, employers, and policy makers stress the need for competencies in communication, collaboration, ICT-related domains, and social and/or cultural skills (Voogt & Roblin, 2012). As a result of globalization, it is necessary to acquire multicultural social interaction and communication skills, in which competency in a foreign language has been emphasized and proficiency in academic writing plays a key role (Binkley et al., 2012). Furthermore, as English is the global lingua franca, enhancing L2 learners' English writing skills is one of the core objectives in foreign language courses around the world.

Learners' L2 writing performance is influenced by several individual and contextual factors including their L2 proficiency level, first language, previous writing experiences, how L2 writing is learnt, and other socio-cultural norms and expectations. For example, the cultural background was supposed to be a key factor influencing writing in a second language (Atkinson, 1999; Atkinson & Ramanathan, 1995; Myles, 2002; Uysal, 2008) and studies investigating whether writers from shared cultural backgrounds display typical writing patterns in their texts indicated shared preferences in rhetorical patterns in terms of organization, macro-level rhetorical patterns, coherence, and use of transition devices. Moreover, as writing is consciously learnt in a certain way in a particular context, the cultural-specific nature of

schemata is very likely to influence when students write in a second language. Myles (2002) noted that during writing under pressure, L2 writers rely on rhetorical norms from their first language for synthesis of meaning. However, this does not necessarily mean that those who know how to write a summary in their first language will also be able to summarize well in English, as they have varying proficiency levels and writing experience.

In promoting learners' academic writing performance, it is of great importance to understand their background proficiency levels. In the case of learning academic writing in English, the use of a variety of sentence structures with embedded clauses and specialized vocabulary are the main demands in developing learners' proficiency. Biber et al. (2011) and Schleppegrell (2001) stated that syntactic constructions in academic writing are more complex and highly specialized compared to those in everyday language. Therefore, assessing the range and varied use of grammatical structures (i.e., syntactic complexity) has been included as a major assessment criterion in evaluating learners' writing proficiency.

As an alternative to holistic and analytical rating assessment, measuring syntactic complexity of student writing through automated tools has become a promising way to assess learners' writing proficiency. To this end, this study investigated the syntactic complexity in the writing of first-year undergraduate EFL students in two higher education institutions in Hungary and Myanmar. Moreover, I examined the language-related error patterns (grammar, usage, vocabulary, and mechanics) in students' writing to better understand the role of errors in L2 writing proficiency.

7.2 Importance of studying learner corpora in L2 writing

The study of students' corpora has become a major field of research in corpus linguistics over the past twenty years, as it provides useful information regarding different aspects of students' writing including syntactic and lexical features, and text organization patterns of written learner discourse. This information can be used by L2 writing researchers and teachers to identify areas in which students need to improve and to address them in writing instruction (Biber, Gray, et al., 2011; Ghaboosi & Horváth, 2015; Horváth, 2001, 2016; Lan & Sun, 2019; Rokoszewska, 2022). Lan and Sun (2019), for instance, studied noun phrase complexity in writings of first-year students and concluded that they would benefit from targeted instruction on advanced noun modifiers during academic writing courses. In connection with lexical features, Horváth (2016) established a lexical profile of Hungarian university students' exam corpus and stated that the development of such learner corpora contributed to valid and useful information for

future development of writing skills. Unlike these studies, Rokoszewska (2022) established a learner developmental corpus over a three-year period and investigated monthly growth rates of CAF measures in English L2 writing at a secondary school in Poland. The results showed that the CAF measures developed at significantly different average monthly growth rates and that the growth rates of CAF measures were non-linear and fluctuant. Overall, these findings shed light on the importance of studying learner corpora to inform writing instruction and to understand the development of students' writing abilities.

7.3 Studies on language-related errors in L2 writing

Despite the relative difficulty in assessing students' academic writing proficiency, the ability to use grammar accurately is regarded as a crucial element in the assessment criteria applied by L2 writing instructors (Romano, 2019). Also, linguistic accuracy contributes to clarity and idea development in writing which in turn helps students attain high scores in task fulfilment. Therefore, writing accuracy in L2 writing is considered essential to the evaluation of students' academic writing proficiency and learning success, especially in tertiary education (Biber et al., 2011; Mazgutova & Kormos, 2015).

For the purpose of improving writing instruction, previous studies analysed students' errors in L2 writing, targeting a wide variety of errors such as errors in word forms, verb forms, subject-verb agreement, articles, word choice, noun plurals, and sentence structure (e.g., Dahlmeier et al., 2013; Olsen, 1999; Phuket & Othman, 2015; Zheng & Park, 2013). Olsen (1999), for example, studied the written texts of Norwegian EFL learners and found a relationship between students' language proficiency and their errors in writing: less proficient learners had a higher number of grammatical, orthographical, and syntactical errors. Moreover, other studies (e.g., Dahlmeier et al. 2013; Phuket & Othman 2015) targeted the errors found in university undergraduate students' writing. For example, Dahlmeier et al. (2013) took a corpus-based approach and analysed the undergraduate university students' errors in the NUS Corpus of Learner English (NUCLE). The authors found that wrong collocation/ idiom/ preposition, local redundancies, article or determiner, noun number, and mechanics were the top five error categories. Similarly, another study by Phuket and Othman (2015) analysed the narrative essays composed by Thai university students and found the most frequent types of errors including translated words from Thai, word choice, verb tense, preposition, and punctuation. What emerges from these studies is that accuracy is one of the main criteria in evaluating the

performance of L2 writers and that understanding their errors in writing inform writing instruction and materials development in EFL academic settings.

7.4 Research contexts and research questions

Although differences are present in the two educational contexts (Myanmar and Hungary) relating to educational, cultural, and sociolinguistic characteristics, it is undeniable that L2 learners across formal learning contexts (e.g., learning English at the university level) have a common goal of studying English for academic purposes. Comprehension and production of English for academic purposes require not only communicative competence but also pragmatic competence, and the culturally appropriate use of English. It is against this backdrop that the current study aims to investigate the syntactic complexity of EFL undergraduate students in two countries and to examine their language-related errors in writing. The following three research questions guided my study:

RQ1. What textual characteristics do the Myanmar and Hungarian EFL students' essays exhibit?

RQ2. To what extent do syntactic complexity measures differentiate the essays produced by the students at a university in the two countries?

RQ3. To what extent are there differences in the language-related error patterns in the essays produced by the students from the two contexts?

7.5 Methods

7.5.1 Participants

Two intact classes of first-year undergraduate English majors: 30 students (19 females and 11 males) from a Myanmar university and 28 students (15 females and 13 males) from a Hungarian university were recruited in the first and second semester of 2020/2021 Academic Year. Here, it should be noted that they were the same students who participated in previous chapters (see Chapters 4 and 6). The Myanmar students were native speakers of Burmese and those from the other sample were Hungarian native speakers. They were of typical university age, ranging from 17 to 18 years (Myanmar) and from 18 to 26 years (Hungary). The students' English proficiency level was defined by the programme level. At the time of the study, both groups were enrolled in university undergraduate studies which required them to have intermediate (B1–B2) levels on the CEFR scale. The students voluntarily participated and were informed that their written texts were anonymously analysed for research purposes.

7.5.2 Instruments

As shown in Figure 7.1, I used the following proficiency essay practice writing tasks (i.e., four-paragraph guided essays without separate introduction and conclusion paragraphs). The essays entailed four guided prompts to elicit responses (giving a narrative account of personal experiences or sharing views on a proposed statement) (see Section 3.2.1). Though the writing topics were different in terms of topic, they were supposed to elicit free constructed responses (Norris & Ortega, 2000) and allowed students to produce language with relatively few constraints and with meaningful communication as the goal for L2 production. When responding to the writing tasks, they were required to write a text in 300-400 words using one of the essay options.

7.5.3 Research procedure

As stated earlier in Section 7.5.1, I used the data from the students who participated in Studies I and III. In this study, I examined a selection of 58 students' texts written by Myanmar and Hungarian students. The selection was made on the basis of the writing topics, as topic tends to have a significant effect of syntactic complexity features of writing (see Yang et al., 2015). Though the writing topics were not identical (Figure 7.1), they were supposed to have similar themes. For example, Tasks A in Hungary and Myanmar elicited students' views on an essential technological device and social media. Both topics were related to modern modes of communication resulting from technological advancements. They completed these essays in Week 3 (Myanmar) and Week 5 (Hungary) as the first writing tasks before receiving any feedback treatment from the teacher or Grammarly (see Tables 4.1 and 6.1).

Figure 7.1.

Topics of the Writing Tasks

Task A: 'Tech, I love you'
Write an essay about your views on a technological device essential for you. In the text,
• introduce a device you find difficult to live without;
• discuss how your life has changed since you started using that device;
• explain your attitude towards technology addiction;
• tell a story about an event where your favorite device was or was not helpful.

Task B: Parents just don't understand
Write an essay about your views on parent-child generational differences. In the text,
• introduce a common problem caused by generational differences;
• discuss how the two generations view the same problem;
• explain your attitude towards generational differences;
• tell a story about how you could or could not solve a problem caused by generational differences with your parents.

a) Writing tasks in Hungary

Task A: Social media
Write an essay on your ideas about social media. In the text,
• discuss the role of social media nowadays;
• describe an important event where social media matters;
• explain why you think social media can have positive or negative influence on people; give examples;
• discuss how social media influences your life.

Task B: The worst teacher who discouraged me
Write an essay on your ideas about the worst teacher who discouraged you. In the text,
• characterize the teacher;
• describe the context where you met;
• tell a story about him/her and how he/she discouraged you;
• explain how he/she influenced your life.

b) Writing tasks in Myanmar

7.5.4 Data analysis

Two software packages (Coh-Metrix and L2SCA) were used to extract 17 features to cover the multidimensional nature of the syntactic complexity construct (for details, see Section 3.4.3). Statistical analyses were carried out, using R statistical programme. To answer RQ1, descriptive statistics of textual features of the students' essays (e.g., essay length, paragraph length, and sentence length) were compared. Also, independent samples t-tests were used to compute for the difference between the sentence length of the essays produced by the two groups. As for RQ2, I first exported the essays into Coh-Metrix and L2SCA to automate the syntactic complexity measures. I then conducted independent samples t-tests on 17 measures of syntactic complexity to examine the differences in the essays produced by the two groups.

To address RQ3 regarding the language-related error patterns in the texts, I probed into the discourse networks of each country using an Epistemic Network Analysis (ENA). The ENA approach (https://www.epistemicnetwork.org/) is used to describe individual (group) cognitive framework patterns through quantitative analysis of discourse data and identify meaningful and quantifiable patterns in discourse or reasoning (Zhang et al., 2022). ENA moves beyond the traditional frequency-based assessments by examining the structure of the co-occurrence or connections in coded data. It can also be used to compare units of analysis in terms of their plotted point positions, individual networks, mean plotted point positions, and mean networks, which average the connection weights across individual networks. Networks may be compared by using network difference graphs and these graphs are calculated by subtracting the weight of each connection in one network from the corresponding connections in another (for details, see Shaffer et al., 2016).

For performing the ENA analysis, these steps were taken. I initially collected corrective feedback points (i.e., errors) on the students' essays marked by the class teachers. Then, I separated and coded each sentence into a record line and coded the frequency of errors. Finally, adapting the error typology of Dikli and Bleyle (2014), I classified the coded errors into four categories: grammar (subject-verb agreement, pronoun, verb form, verb tense, run-ons, sentence structure, and omission of object), usage (article, noun ending, preposition, word form, idiom, adverb, and conjunction), lexical (word choice and collocation), and mechanical (punctuation, spelling, and capitalization) errors (see Appendix I). Table 7.1 provides a sample excerpt from an actual coded file serving for the ENA analysis.

Table 7.1

Excerpt of a Coded File Containing Language-related Errors in the Myanmar Students'

Essays

METADATA COLUMNS			CODE COLUMNS							
	STANZA COLUMN	UNIT COLLUMN								
Line	student	Country	G.verbForm	G.structure	U.Adverb	U.conjunction	M.punctuation	M.spelling	L.wordChoice	L.collocation
15	E30	Myanmar	0	0	0	1	1	0	0	1
22	E30	Myanmar	0	1	0	0	0	1	1	0
10	E31	Myanmar	0	0	0	0	1	1	1	0
4	E34	Myanmar	0	0	0	1	1	1	0	0
13	E14	Myanmar	0	0	1	1	0	0	0	0
2	E20	Myanmar	0	0	0	1	1	0	0	0
6	E20	Myanmar	0	1	0	1	0	0	0	0
13	E22	Myanmar	0	0	0	0	1	1	0	0
17	E22	Myanmar	1	1	0	0	0	0	0	0
12	E26	Myanmar	0	1	0	0	0	0	1	0

7.6 Findings

7.6.1 Textual characteristics of the Myanmar and the Hungarian EFL students' essays

The first research question concerned the textual characteristics of the essays written by the groups in Myanmar and Hungary. Descriptive statistics indicated that the texts had varying lengths with simple and complex sentences. Table 7.2 presents the average numbers of paragraphs, sentences, and words on first drafts of their essays. Particularly at the paragraph level, the Myanmar group tended to produce longer paragraphs (M = 101.77, SD = 24.38) in comparison to those written by the Hungarian counterparts (M = 97.61, SD = 13.08). Further differences were found at the sentence level: the students in Myanmar attempted to generate more than 27 sentences, whereas the Hungarian students wrote around 20 sentences in their essays. Also at the word level, the Myanmar group tended to produce more words in

comparison to their Hungarian peers, although the writing tasks required both groups to write about 300 to 400 words.

Table 7.2

Descriptive Statistics: Paragraph, Sentence, and Word Counts

Country	Paragraph length		Sentence count		Sentence length		Word count	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Myanmar	101.77	24.38	27.20	7.37	15.34	7.10	417.27	97.53
Hungary	97.61	13.08	20.54	3.94	19.14	7.79	397.39	52.09

Moreover, I found a statistically significant difference in the sentence length in the essays produced by Myanmar (M = 15.34, SD = 7.10) and Hungarian students (M =19.15, SD = 7.97), $t = 9.31$, df = 1154.6, $p < .001$. Figure 7.2 demonstrates the distribution of sentence length in the two groups. Overall, the Hungarian students seemed to generate both shorter and longer sentences, ranging from 8 to 24 words in a sentence, whereas their Myanmar peers used a smaller range of words (8 to 16) in their sentences. Students in the Hungarian cohort applied more diverse sentences: 69.22% complex sentences and 30.78% simple ones, compared to those in the other group who used fewer complex sentences (43.50%) and more simple sentences (56.50%) in their essays.

Figure 7.2.

Distribution of Sentence Length (Words) in Students' Essays
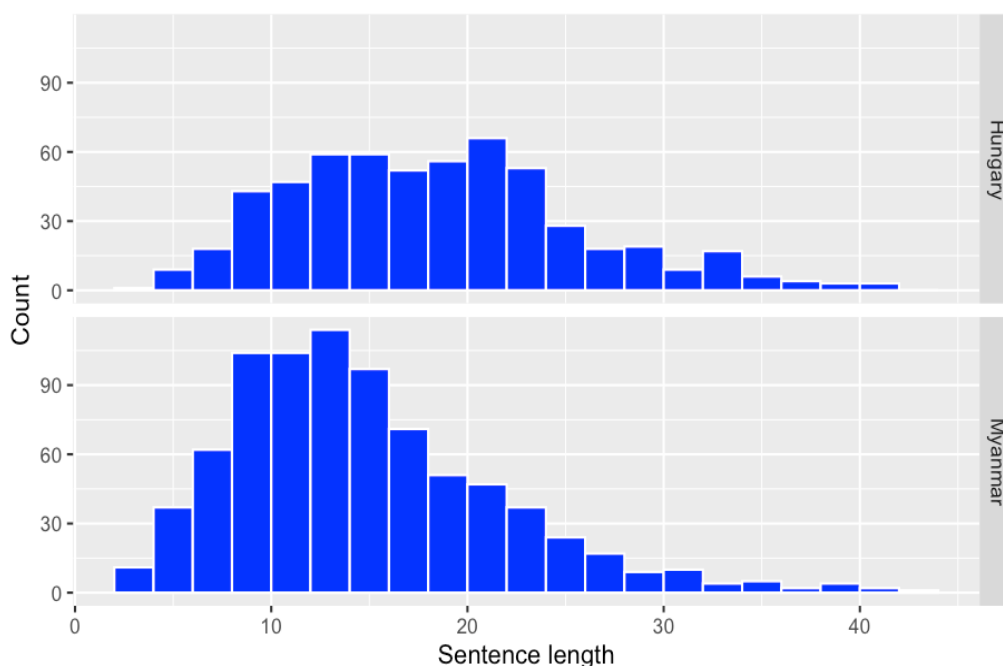
7.6.2 Differences in syntactic complexity in students' texts

Detailed information regarding the syntactic complexity measures identified in the essays is summarized in Tables 7.3 and 7.4. Overall, the findings from the two syntactic complexity analysers consistently demonstrated that most complexity indices were found to distinguish the essays produced by the two groups, indicating that the essays produced by the Hungarian students had greater syntactic complexity in comparison to those of the Myanmar cohort.

Table 7.3

Results of Independent Samples t-tests of Three Syntactic Complexity Measures Computed by

Coh-Metrix

Syntactic complexity measures	Index code	Myanmar		Hungary		Independent samples *t*-test	
		Mean	SD	Mean	SD	*t*	*p*
Sentence syntax similarity	SYNSTRUTt	0.12	0.03	0.08	0.02	-5.16	**<.001**
Left embeddedness, words before main verb	SYNLE	3.43	1.01	4.15	0.99	2.78	**.007**
Number of modifiers per noun phrase	SYNNP	0.67	0.16	0.63	0.09	-1.17	.245

Note. Index code is a typical code presented in Coh-Metrix programme.

As for the indices calculated by Coh-Metrix (Table 7.3), significant differences were found in the two groups in two indices: sentence syntax similarity and left embeddedness (number of words before main verb), but not in the number of modifiers per noun phrase. Particularly, the mean scores of sentence similarity across paragraphs (SYNSTRUTt) in the Myanmar group (M = 0.12, SD = 0.03) were higher than those in the Hungarian cohort (M = 0.08, SD = 0.02). In other words, the sentences in the Myanmar students' essays revealed more uniform syntactic constructions with less complex syntax that is easier for the reader to process, whereas the sentences in the Hungarian students' essays were less similar in terms of syntactic structures, resulting in structurally diverse sentences.

To tap the phrase-level complexity, two indices (SYNLE and SYNNP) from Coh-Metrix were employed. The SYNLE index calculates the mean number of words before main verb with the understanding that more words before the main verb leads to more complex syntactic structure (Crossley & McNamara, 2012). The other index considered the mean number of modifiers per noun phrase with the understanding that noun phrases with more modifiers were

supposed to be more complex syntactic constructions. Based on the independent samples *t*-test, the mean scores of SYNLE index in the Myanmar students' essays (M = 3.43, SD = 1.01) were significantly lower than those of the Hungarian students' essays (M = 4.15, SD = 0.99). In other words, the essays written by the Hungarian students depicted more complex syntactic structures in comparison to those produced by their Myanmar counterparts.

In connection with the 14 indices computed by L2SCA, the independent samples *t*-tests indicated significant differences in the two groups (Table 7.4). Particularly, the two measures of the length of production units: MLS and MLT differentiated the two groups, as the mean scores were significantly higher in the essays of the Hungarian students, compared to the essays of the other group. However, MLC did not separate the proficiency levels, resulting in no statistically significant differences between the two groups.

Table 7.4

Results of Independent Samples t-tests of 14 Syntactic Complexity Measures Computed by

L2SCA

Syntactic complexity measures	Index code	Myanmar		Hungary		Independent samples *t*-test	
		Mean	SD	Mean	SD	*t*	*p*
Length of production unit	MLC	8.67	1.87	8.38	1.07	-0.72	.470
	MLS	15.17	3.28	19.72	3.26	5.20	**<.001**
	MLT	13.69	2.74	15.14	2.62	2.02	**.040**
Sentence complexity	C/S	1.77	0.34	2.37	0.37	6.19	**<.001**
Coordination	CP/C	0.23	0.12	0.17	0.09	-2.19	.030
	CP/T	0.36	0.17	0.31	0.16	-1.22	.230
	T/S	1.11	0.13	1.31	0.17	5.01	**<.001**
Subordination	C/T	1.59	0.19	1.81	0.23	3.80	**<.001**
	CT/T	0.43	0.11	0.57	0.11	4.67	**<.001**
	DC/C	0.35	0.08	0.40	0.07	2.69	**.009**
	DC/T	0.57	0.19	0.74	0.19	3.22	**.002**
Particular structures	CN/C	0.93	0.41	0.75	0.23	-2.06	.040
	CN/T	1.45	0.59	1.36	0.49	-0.62	.530
	VP/T	2.00	0.29	2.32	0.38	3.42	**.001**

Note. Index code is a typical code presented in L2SCA programme.

A similar trend was found in the sentence complexity index; independent samples *t*-tests revealed significant differences across the two groups, favouring the texts of the Hungarian students. However, no significant differences were found in the two groups in terms of the

coordination indices except for T/S. All subordination indices related to syntactic complexity indicated significant differences between the two groups, and thus these indices were supposed to be the clearest separators in differentiating the students' written texts. For example, when the means of the clauses per T-unit in the two groups were compared, the Hungarian students' texts had relatively more clauses per T-units (M = 1.81, SD = 0.23) than the other group (M = 1.59, SD = 0.19). Similar patterns were found in the other subordination indices, favouring the texts written by the Hungarian students.

Clearly, there were differences in the amount of subordination at the paragraph level when the comparison was made between the two sample paragraphs of the student writing (Extracts 1 and 2) and in the amount of subordination indices of each paragraph analysed by L2SCA (Figure 7.3).

Extract 1: A sample paragraph from a Hungarian student's essay

> *There is one thing however that I always try to make sure of, which is not spending too much time on my phone. Many entertaining apps make us addicted to our phones and other technological devices. While they seem to be fun, sometimes they do more harm than good. Social media is a big part of the problem. A lot of people are obsessed with the idea of the "influencer lifestyle" which seems to be so perfect but is not real. This is something that could affect the mental health of the youth. I think it is important to make sure we take some time away from our screens.*

Extract 2: A sample paragraph from a Myanmar student's essay

> *As I mentioned above, the social media plays quite important roles for everyone especially when we aren't allowed to go anywhere. The news channels on Facebook have been providing the news of the coronavirus. Moreover, the social media have brought the national leaders nearer to the people which makes the citizens more familiar to the leaders. However, despite the advantages, the social media have some drawbacks. As people have become so obsessed with the social media, they become less active in reality. Young people spend more time on Facebook and other platforms all the times* recklessly. This is very hideous that is youths wasting their precious times* on the social media. What is more, people can be mislead** by the fake news and disgusted by the inappropriate pictures for either kids or elders.*

* The correct word is 'time'.
** The correct word is 'misled'.

Figure 7.3.

Comparison of the Amount of Subordination Indices of the Two Paragraphs (Extracts 1 and 2) Produced by Hungarian and Myanmar Students
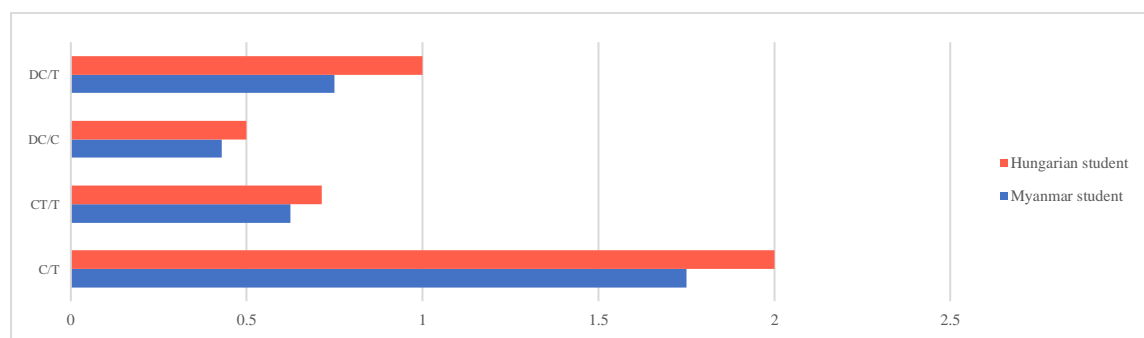
Compared to other measures of syntactic complexity, two features of phrasal sophistication (complex nominals per clause and complex nominals per T-unit) did not differentiate the two groups; these indices discriminated poorly between the two groups of students. In contrast, significant differences were found in verb phrases per T-unit: the Hungarian students tended to include more verb phrases per T-unit (M = 2.32, SD = 0.38) than the Myanmar group (M = 2.00, SD = 0.29).

7.6.3 Differences in language-related errors in students' texts

Figure 7.4 demonstrates the frequencies of language-related errors in the essays produced by the two groups. The overall results suggested that errors were more prevalent in the Myanmar cohort than in the other group regardless of error types. The most frequent errors found in both groups included those in punctuation, word choice, and noun endings, although these errors doubled in the texts of Myanmar students. Punctuation errors were found to be common in both groups, but the Myanmar students had higher error frequency.

When examining each error category, the results revealed that the Myanmar students had more grammar errors relating to sentence structure and verb form, whereas these errors did not occur frequently in the texts composed by the Hungarian students. Likewise, usage errors were more dominant in the Myanmar students' texts (article, conjunction, and preposition errors) as well as lexical and mechanical errors except for capitalization errors.

Figure 7.4.

Frequency of Language-related Errors in Essays Produced by Myanmar and Hungarian Students
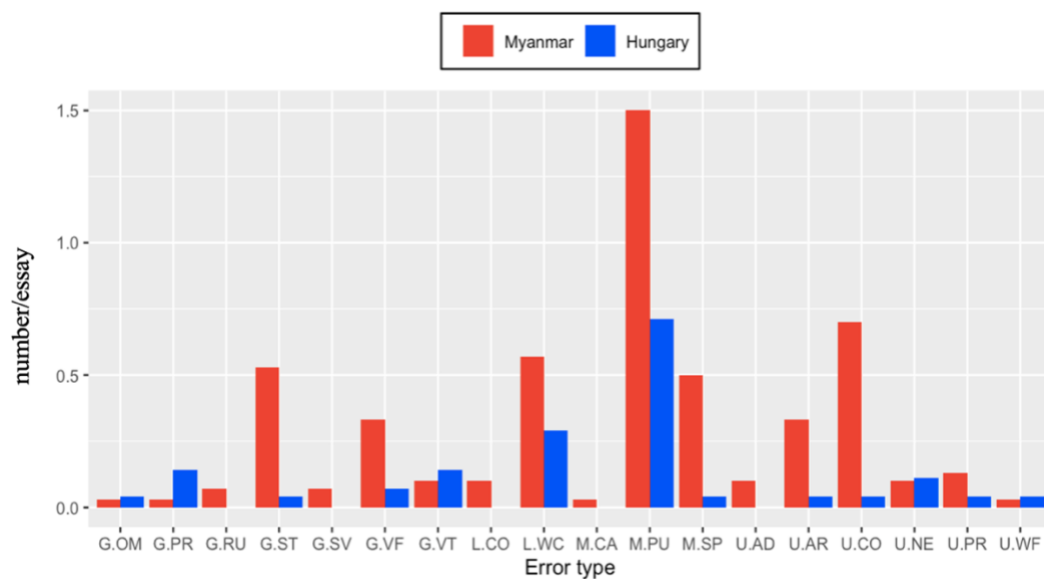
To further investigate the error patterns in the written texts of the two groups, I conducted the ENA analysis with definition of the units of analysis as all lines of data associated with a single value of country subsetted by a student. One unit consisted of all the lines associated with each student. The ENA model included the following codes: G.svAgreement, G.pronoun, G.verbForm, G.verbTense, G.runOn, G.structure, G.OmissedObject, U.articles, U.nounEnding, U.preposition, U.wordForm, U.Idiom, U.Adverb, U.conjunction, M.punctuation, M.spelling, M.capitalization, L.wordChoice and L.collocation. I defined conversations (or stanzas, Table 7.1) as all lines of data associated with a student's text in each country (Shaffer et al., 2016). The resulting networks were aggregated for all lines for each unit of analysis in the model. I aggregated networks using a binary summation, in which the networks were visualized using network graphs where nodes correspond to the codes, and edges reflect the relative frequency of co-occurrence or connection between two codes. The result was two coordinated representations for each unit of analysis: (i) a plotted point, which represents the location of that unit's network in the low-dimensional projected space and (ii) a weighted network graph.

Figure 7.5 shows a sample pattern in the network model presenting the frequency of error occurrence and connections among language-related errors in a student's essay. Particularly, a network is represented via a single point as a centroid in the space (like the centre of mass of an object). Such centroid of the student's network is presented as a red dot in the model. Darker dots and thicker lines indicate the frequency of error occurrence and closer connections among errors. For example, the following model displays several connections among the mechanical errors (punctuation and spelling), usage errors (conjunction), grammar errors (verb tense), and lexical errors (collocation) in a student's essay.

Figure 7.5.

Language-related Error Patterns in Student E48's Essay Simulated in ENA Model
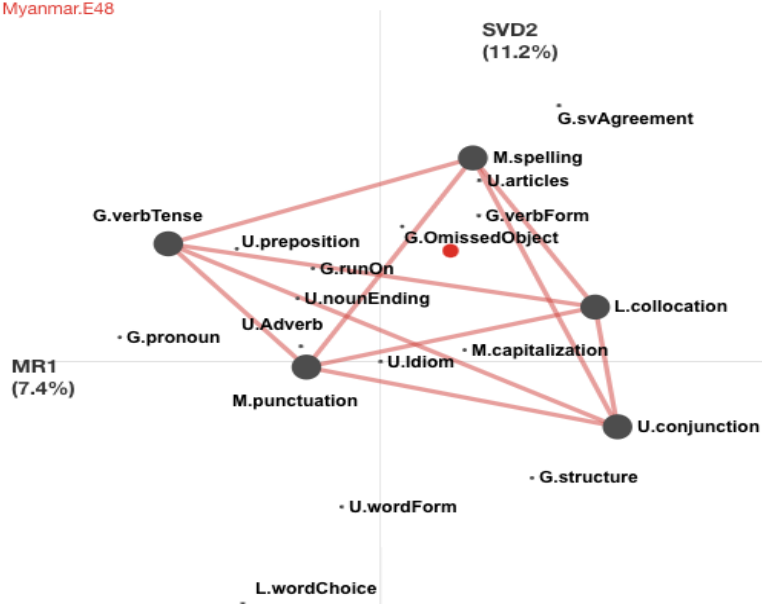
The model had co-registration correlations of 0.94 (Pearson) and 0.93 (Spearman) for the first dimension and co-registration correlations of 0.82 (Pearson) and 0.83 (Spearman) for the second dimension. Figure 7.6 shows the subtraction networks of error patterns in the students' essays. The squares represent the centroid (i.e., mean) for the two student groups and the dots represent the error type in the written texts.

The network structures can be characterized as follows: along the X axis, towards the left, there are verb tense and pronoun errors; towards the right are spelling, structure, and collocation errors. Along the Y axis, towards the top is omission of object error, and towards the bottom are conjunction and word choice error. The highlighted lines in the subtraction networks show the differences between the two groups' epistemic networks. As depicted in the subtraction network of Hungarian (blue) and Myanmar (red) students' essays, the blue lines indicate that the Hungarian students' essays had more associations between pronoun and punctuation errors, and word choice errors. In contrast, the essays produced by Myanmar students revealed more connections between word choice and sentence structure errors, and spelling and conjunction errors.

Furthermore, I used independent samples *t*-tests to examine differences, assuming unequal variance to the location of points in the projected ENA space for units in the essays written by the Myanmar and Hungary students. Along the X axis, the results of the *t*-test revealed that the Myanmar cohort's texts (M = 0.34, SD = 0.40, N = 30) were statistically significantly different from those generated by the Hungarian students (M = −0.36, SD = 0.53, N = 28; *t* (50.63) =

5.69, $p = 0.00$, Cohen's $d = 1.51$). They indicated that different patterns regarding verb tense, pronoun, spelling, structure, and collocation errors were observed in the two groups' texts. Nevertheless, there were no significant differences between the two groups regarding the errors along the Y axis ($p > 0.05$), suggesting that the errors related to omission of object, conjunction, and word choice occurred in the texts of both groups of students.

Figure 7.6.

Comparison of Language-related Error Patterns in Students' Essays in ENA Analysis
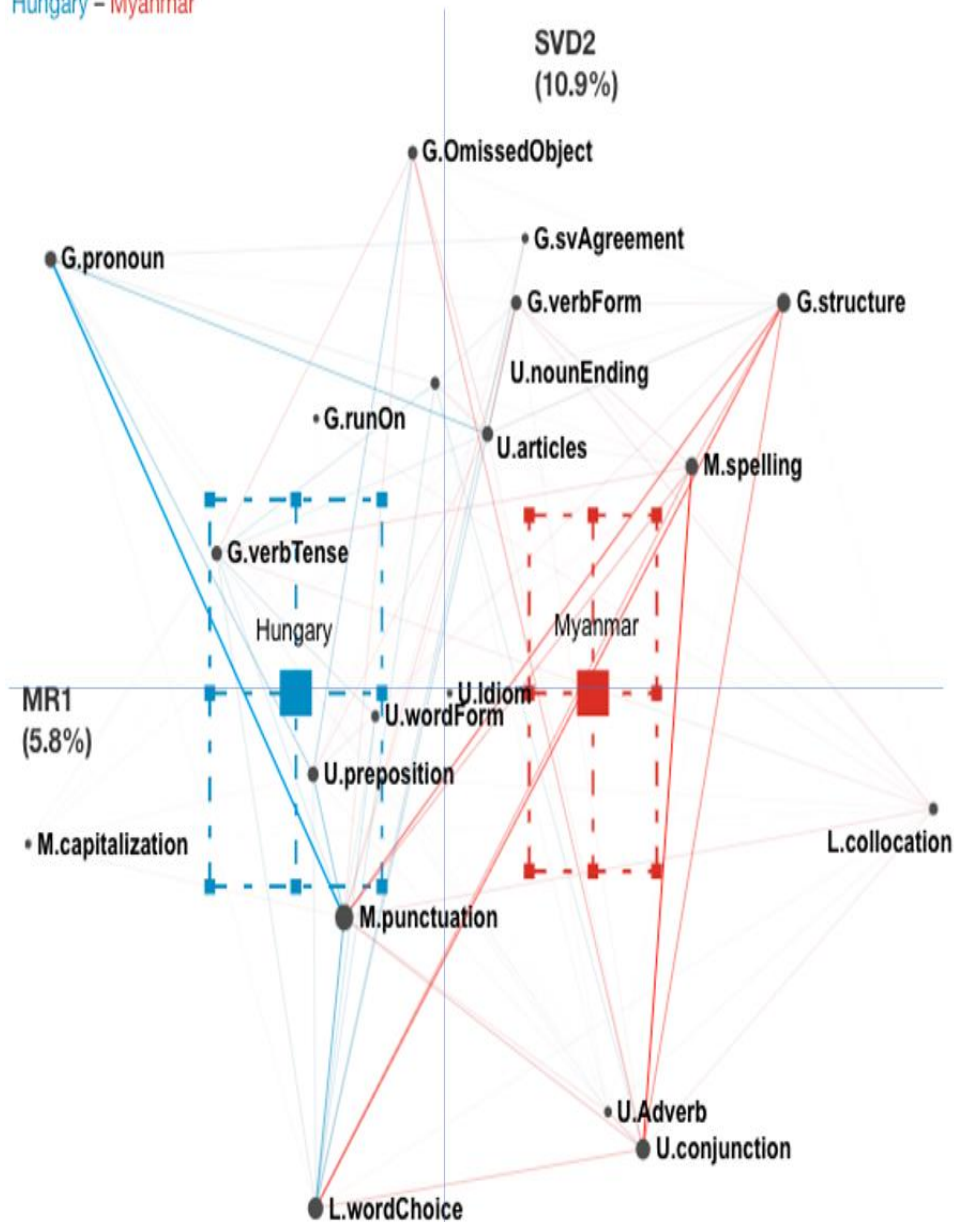
Note. The squares represent the mean (centroid) for each cohort and the black dots represent codes.

7.7 Discussion and conclusions

This study investigated the syntactic complexity and language-related error patterns in the essays of undergraduate students from Myanmar and Hungary. An examination of various dimensions of syntactic complexity revealed that most indices, two indices of Coh-Metrix (SYNSTRUTt and SYNLE), all subordination indices (C/T, CT/T, DC/C, and DC/T) and two indices of length of production unit (MLS and MLT) as well as sentence complexity index (C/S) were found to differentiate the students' texts. Findings from the analysis of the students' language-related errors indicated significant differences in their error patterns. I next discuss the findings with reference to previous research on syntactic complexity and language-related errors in L2 writing.

According to Ortega (2003) and Wolfe-Quintero et al. (1998), more proficient L2 learners tend to produce longer and more varied sentences in writing. In my study, through the analyses of the textual characteristics of the students' essays, I found that the Hungarian students made use of longer sentences with a variety of sentence structures, whereas the Myanmar students used simpler and shorter sentences. Therefore, it is very likely that the Hungarian students are more proficient than the Myanmar peers. To better understand their varying proficiency levels, I also compared students' writing scores assessed by their class teachers. The results of independent samples t-test indicated no significant differences in the writing scores in the essays produced by Myanmar (M = 8.36, SD = 2.01) and Hungarian students (M = 8.93, SD = 1.6), t =1.17, df = 53.13, p = .245. However, it should be stressed that the mean scores of the Hungarian students were somewhat higher than the comparison group, indicating their higher proficiency levels. A possible explanation for these conflicting results between the findings from the teachers' writing assessment and automated evaluations might be due to the difference between the nature of classroom writing assessment and computational analyses. Particularly, the teachers used the rating scale which comprises four criteria: task achievement, coherence and cohesion, grammatical range and accuracy, and lexical range and accuracy to assess the students' essays. However, the computational tools analysed the syntactic complexity of written texts without considering other aspects of writing.

One surprising finding is that although the writing tasks required the students to write in about 300-400 words, the Myanmar students wrote longer essays (M = 417) than their Hungarian peers (M = 397). These variations could partly be attributed to attitudes towards L2 writing, influence of the writing model in their L1 (Myles, 2002), or other socio-cultural norms in the instructional contexts. For example, Hungarian university students are taught to use

longer and more complex sentences, and to avoid writing longer than necessary for the sake of clarity and conciseness (personal communication, Simon, August 17, 2021). In contrast, in Myanmar, students are seemingly encouraged to express more ideas in writing, and they also think they might get higher scores if they write longer texts. These socio-cultural norms pertinent to the study contexts might be implicitly reflected in students' essays.

In a research synthesis conducted by Ortega (2003), indices gauging length of production units (mean length of sentence, mean length of T-unit, and mean length of clause) were the most frequently used syntactic complexity measures and reported to be reliable indicators of proficiency level differences for L2 writers (Wolfe-Quintero et al., 1998). In line with these findings, my study found that most indices of length of production units (MLS and MLT) were able to distinguish the two student groups, except for MLC. These findings were consistent with those of Khushik and Huhta (2019), who found that MLS and MLT best distinguished the three CFER levels (A1, A2, and B1), but not MLC. Also, Lu (2011) found that all three indices of length of production units progressed linearly across proficiency levels.

More importantly, all subordination indices best differentiated the essays produced by the Myanmar and Hungarian students. As noted by Wolfe-Quintero et al. (1998), the complexity measure that gauges the amount of subordination (which are computed by counting all clauses and dividing them over a given production unit of choice) correlated best with writing development. Specifically, in their review of syntactic complexity measures in thirty-nine primary studies, the authors posited that mean length of T-unit (MLT), clauses per T-unit (C/T), and dependent clauses per clause (DC/C) were the most satisfactory measures which were associated linearly and consistently with programme levels. Norris and Ortega (2009) also acknowledged that amount of subordination (e.g., mean number of clauses per T-unit) could be a useful and powerful index of complexification at intermediate and upper-intermediate levels compared to the amount of coordination which might be potentially more sensitive than subordination measures (Bardovi-Harlig, 1992). Thus, it is fair to say that the essays produced by the Hungarian cohort tend to represent their higher proficiency levels in L2 writing compared to the other group.

In connection with the language-related error patterns in student writing, the results of the frequency analysis and ENA approach are mostly consistent. The latter demonstrates a fuller picture of common error patterns in writing and the connections among the students' errors. The most typical errors found in the two groups include those in punctuation, word choice, and noun endings. My results partially agree with those of Dahlmeier et al. (2013), who found that mechanical errors were one of the top error categories that the students frequently made. One

possible explanation of the frequent occurrence of punctuation errors in both groups might concern their knowledge about the target language, their attention, or the influence of their mother tongue. For example, the punctuation rules in the Hungarian language are strict and such rules are in many ways different from English. As a result, Hungarian students end up using commas in the wrong (unnecessary) places when they write in English. Nonetheless, the findings that article errors were also frequent in the students' writing contradict my results which could be explained by cross-language transfer (Zhu et al., 2021) from the students' L1s to L2. In my study, the use of articles is present in the students' L1s (Burmese and Hungarian) and thus, they might possibly find it easy to apply the rules of article systems in English, although it was not the case in Dahlmeier et al.'s (2013) study.

Furthermore, the ENA analysis indicated significant differences in the errors regarding verb tense, pronoun, spelling, sentence structure, and collocation between the two groups, whereas no significant differences were found in omission of object, conjunction, and word choice errors. Along with the variations in the most frequent errors of the Myanmar and Hungarian groups, my findings partly provide a pattern of the language-related errors of undergraduate students in higher education which could inform discussion of how to address these errors effectively by providing supplementary grammar instruction for L2 writers.

The results of the study contribute to the understanding of the linguistic features of EFL students' texts, including error patterns in writing in two educational contexts. As analysing syntactic features in academic writing have gained importance (Biber et al., 2011; Maamuujav et al., 2021), understanding these features in students' texts could provide crucial information regarding their L2 writing proficiency and development. Additionally, ENA was performed as a potential approach in educational research (Shaffer, 2017); it can be replicated to interpret discourse data in learning L2 in school contexts. The successful application in the present study allows us to understand the common error patterns in students' writing, and it has led to recommendations to apply network analyses in future studies.

Some limitations are acknowledged in the present study. The data was drawn from 58 students from two universities on limited and varied tasks; these facts limit the generalization of my findings. Large-scale data needs to be mined to explore the linguistic characteristics and error patterns in student writing. Moreover, studies using similar writing tasks, first languages, and students from different proficiency levels should be considered to obtain a fuller picture of the relationship between syntactic complexity and writing proficiency. Furthermore, as academic writing in educational contexts is shaped by cultural norms in specific instructional contexts (McIntosh et al., 2017), studies investigating the similarities and differences between

the written discourse of EFL learners from different cultural backgrounds would further pinpoint how syntactic features are considered differently. Therefore, this aspect of syntactic complexity should be revisited and explored in future empirical research.

CHAPTER 8. GENERAL DISCUSSION AND CONCLUSIONS

This final chapter has three sections that provide insights into L2 writing research and writing instruction. The first section is devoted to implications for research and discusses the findings of the literature review. Using the insights from the review, I propose recommendations for future WCF research. The second section presents implications for L2 writing instruction based on the results of the empirical studies (Chapters 4, 5, 6, and 7). The final section ends with limitations of the studies and directions for future research.

8.1 Implications for research: Insights from the literature review

Given that providing feedback is a worthwhile pedagogical practice in teaching L2 writing, exploring the efficacy of feedback on students' writing performance has become a burgeoning area of investigation over the past few decades. With these intense interests in the field of L2 writing research, most studies provided evidence that feedback can facilitate improved grammatical accuracy in revised and new pieces of writing. Also, a handful of reviews shed light on the beneficial role of feedback on the quality of students' writing (Biber et al., 2011; Crosthwaite et al., 2022; Kang & Han, 2015; Li & Vuono, 2019; Liu & Brown, 2015; Russell & Spada, 2006). However, concerns were raised regarding the impact of feedback on students' writing due to Truscott's (1996, 2004, 2007) assertion that the use of feedback is ineffective and harmful, and thus it should be abandoned. Moreover, differences in methodological features (Liu & Brown, 2015), treatment variables such as type of feedback and source of feedback (Russell & Spada, 2006), and learners' individual differences including proficiency levels and language analytical ability (e.g., Park et al., 2016; Shintani & Ellis, 2015) further influence the efficacy of WCF.

   With these important insights and concerns raised in the existing research, I conducted a critical review of research on WCF with an emphasis on research designs, feedback treatments, writing tasks, and measures of linguistic accuracy. Findings suggested that studies in WCF research have been informed by skill-acquisition theories (i.e., extensive practice and explicit instruction). Within this framework, most studies provided opportunities for learners to engage in writing tasks followed by feedback to help them improve accuracy over time. The 42 primary studies included in the literature review demonstrated caveats in relation to combined feedback strategies, feedback focus, and writing tasks with high/low linguistic and cognitive demands. For example, despite the prevalence of direct, indirect, and metalinguistic feedback in WCF research, combining feedback strategies limits the understanding of how particular feedback

strategies work. As for the feedback focus, the English article system was frequently targeted in WCF research, which raised the issues of how WCF functions with other linguistic features. Also, meaningful relationships between writing tasks of varying cognitive demands and learners' linguistic performance were examined in Kuiken and Vedder (2008); however, most studies in WCF research have failed to take them into account, which limits their contributions to the field of L2 writing research (see Section 2.2.1.3). Criticism of writing accuracy measures centred on their distinctive strengths and weaknesses (see Section 2.2.8).

With these caveats in mind, I offer recommendations for future WCF research. First, studies should examine the impact of individual feedback strategies (either explicit or implicit) on developing L2 written accuracy. These studies would be more appealing in terms of pedagogical practices and inform writing teachers about which feedback strategies should be used with their students of lower or higher proficiency levels. Second, feedback focus should consider students' areas of difficulties and target linguistic features in specific teaching and learning contexts. For example, targeting common L2 error types and understanding students' difficulties in content-related writing issues would be beneficial to enhance students' writing abilities. In addition, previous research is devoted to limited language errors when examining the efficacy of feedback on students' writing (see Section 2.2.5). However, feedback practices in EFL writing classes consider different aspects of writing including language-and content-related issues. Therefore, how students respond to teachers' written commentary (e.g., Ferris, 1997b; Goldstein, 2004; Pearson, 2022a) addressing content-related issues should be attended to in future research. Findings from such inquiry would inform teachers about students' understanding of written commentary and their engagement with such feedback.

Third, more investigations should examine how writing tasks with varying cognitive demands influence learners' writing performance. Such studies would contribute to the existing body of research and allow us to understand how features of writing tasks impact linguistic accuracy. As for the writing tasks, Polio and Park (2012) claimed that studies in L2 writing research often failed to control for topics, genres, or writing task conditions (timed/untimed, in-class/home assignments) and these variations made it difficult to determine which language changes were related to development and to task differences. For example, Riazantseva (2012) examined the accuracy of students' written work on three academic writing tasks (in-class essay on an everyday topic, in-class summary on an article, and at-home summary on an academic journal article). These writing tasks differed in terms of the degree of cognitive and linguistic demands; the in-class essay was the least difficult since it was students' responsibility to determine the content of their essays, whereas the take-home summary was the most

demanding task, as they had to read and summarize a journal article. Findings from this study indicated that the accuracy rates of take-home summaries were significantly lower than those of in-class essays and in-class summaries in the pre- and post-tests. In other words, the types of writing tasks used as outcome measures appear to affect accuracy rates observed in L2 writing. Therefore, addressing the role of different writing task conditions in L2 writing research would be beneficial to better understand the influence of cognitive and linguistic demands of writing tasks on students' writing performance, especially written accuracy.

Fourth, noting the affordances and limitations of available accuracy measures (see Section 2.2.8), studies should apply at least two accuracy measures (e.g., error ratio and holistic ratings) and investigate whether similar patterns of development are found on these measures. Polio and Shea (2014), for instance, investigated the reliability and validity of measures of linguistic accuracy in L2 writing research and found that different methods (e.g., holistic measures, error-free units, and number of errors) did not prove to be valid. This led the authors to suggest that L2 researchers should use more than one measure in a study. In addition to accuracy gains, more research is needed to consider matters of linguistic complexity and fluency to find out how attention to accuracy impacts other dimensions of language development in writing. Polio and Shea (2014) investigated the relationships between accuracy and complexity, and they suggested negative associations between these constructs. Similarly, Bruton (2010) questioned the complicated relationship between complexity and accuracy in L2 writing and claimed that "any measures of accuracy would have to be accompanied by a measure of complexity" (p. 496). Therefore, further investigations on how students' attention on improving accuracy impacts their syntactic complexity and fluency in writing should be undertaken.

Another key deliberation that received scant attention in WCF research is related to the factors considered in rating linguistic accuracy. Although the primary variable of interest in measuring linguistic accuracy is writers' ability to produce accurate texts, other secondary facets such as writing topic, writing prompts, and raters also determine some degrees of score variance (Evans et al., 2014). For example, topic familiarity and difficulty levels of writing prompts may cause fluctuations in accuracy scores of L2 texts. Based on these possible variances due to different writing topics and prompts, future research should take a combination of factors (writers' ability levels, degree of topics' difficulty, and task complexity) into account in rating linguistic accuracy.

8.2 Implications for teaching and L2 writing research: Insights from empirical studies

With the knowledge gained from the literature, four empirical studies were designed to examine the role of written feedback while addressing the contextual issues regarding feedback practices in the research contexts in focus and to add to the existing literature by investigating how feedback facilitates students' writing abilities. In what follows, pedagogical implications are summarised with reference to the rationale for each study.

The findings of the first study (Chapter 4) pave the way for the complementation of multiple sources of feedback, especially teacher and automated feedback in EFL students' writing. As Grammarly feedback focuses primarily on accuracy issues in students' writing, my findings suggest that teachers should use it as a supplementary tool in their classes on a regular basis or encourage their students to use it independently. This would help students become autonomous writers which is in line with the crucial value of writing pedagogy (Horváth, 2013). In other words, students' autonomous use of Grammarly feedback can make them more responsible for their writing development. Also, teacher feedback burden could be reduced and challenges regarding time constraints and inadequate attention to individuals in large classes may be resolved to some extent. Students' successful incorporation of teacher and Grammarly feedback into their revisions and their significant improvement along the line of writing assessment criteria suggested their acceptance and utilisation of WCF to improve their L2 writing. Specifically, Grammarly's flaggings on article/determiner and preposition errors and students' successful revisions of these errors reflect their acceptance of Grammarly as a feedback provider in their EFL course. With these findings in mind, teachers are encouraged to exploit affordances of Grammarly to enhance the efficacy of their feedback.

Additionally, writing teachers should be more selective about the areas they provide feedback on, thereby improving students' writing performance and motivation. In my study, the students were unable to revise errors relating to sentence structure and most of such cases were left unrevised. This result indicates that students need more assistance to address issues in sentence structure and that indirect feedback would not be sufficient to help them correct their errors and scaffold their writing development. Specifically, Ferris (2006, 2010) claimed that direct feedback should be provided for untreatable error types such as lexical choice, idioms, and sentence structure not to enhance students' cognitive load. Therefore, it is important for teachers to identify error types that students cannot self-correct and provide direct feedback on these error types. Moreover, because of the overlaps in teacher and Grammarly feedback in some language errors (e.g., punctuation, preposition, article/determiner), writing

teachers can identify the areas on which Grammarly can provide feedback effectively, thus allowing them to focus on higher-level writing skills including idea development, organization, and rhetoric rather than points Grammarly can handle.

Despite mentioning the potential of incorporating Grammarly into L2 writing classes as a feedback assistant tool, I did not intend to promote its use as marketing. Instead, Grammarly was employed as a form of automated feedback and other AI-powered writing tools such as Wordtune (https://www.wordtune.com/), ProWritingAid (https://prowritingaid.com/), and Sapling (https://sapling.ai/grammar-check) are viable alternatives that could also be used in instructional contexts. Before using them, the effectiveness and reliability of these tools should be evaluated in the first place. For example, Zhao (2022) addressed the affordances and limitations of using Wordtune as a digital writing assistant in helping EFL writers. Specifically, Wordtune provides real-time rewriting options for highlighted texts, such as altering sentence structure and replacing words with synonyms without altering the original meaning. As a result of rewriting, writers can identify dissonances in their writing, create meaning, and learn from their revisions. Therefore, Zhao suggested that Wordtune might be suitable for users with varying levels of English proficiency.

The second empirical study (Chapter 5) focused on the impact of WCF on a particular aspect of writing: syntactic complexity. The results suggested no significant influence of WCF on students' syntactic complexity development; WCF does not help students produce structurally complex texts. However, these findings raise implications for WCF research and pedagogical practices. First, the study refutes Truscott's (2007) assertion that WCF may lead to simplified writing; the lack of complexity gains attributed to WCF does not support the claim that WCF simplifies texts. Instead, my findings suggest that WCF does not result in students producing structurally less complex writing. Consequently, teachers can rest assured that gains in accuracy are rarely conflicted with gains in another aspect of writing (e.g., complexity). Therefore, it is advisable for teachers to continue their regular feedback practices in L2 writing classes, as providing feedback does not make students produce less complex texts. This finding leads to the question of how to help students improve their syntactic complexity which could be examined in future research. For example, Hamano-bunce (2022) suggested that enhancing syntactic complexity might require an "interventionist pedagogical approach" (pp 14 –15) to promote a higher quality of noticing and raise students' awareness of syntactic features in L2 writing. Therefore, the influence of enhanced input on the development of syntactic complexity in EFL students' writing should be examined in future research.

Though academic English reading is not the focus of my studies, it is of great importance to stress that better writing does not tend to develop only through extensive practice of writing (Tankó, 2012, 2016, 2019). Instead, the development of writing abilities also results from a lot of reading. The reciprocity of reading-writing relationships has been extensively documented in previous research (e.g., Cumming et al., 2005; Grabe & Zhang, 2016). Grabe and Zhang (2016) stated that reading is a common source of input for writing tasks in academic settings. For example, understanding how reading texts are organized will not only improve students' reading comprehension but also their writing production. Moreover, the authors added that research on synthesis writing, note-taking, content-driven essays, and theses are some major aspects of reading-writing relations. Therefore, these reciprocal relationships between academic reading and writing abilities should be considered in future research and studies should explore the development of academic reading and writing abilities.

The results from the third empirical study underscore the importance of student engagement for better utilisation of feedback from different sources. The finding that students incorporated more form-focused feedback than meaning-focused feedback in their revisions inform writing teachers about students' understanding of how to address form-focused feedback and their low engagement with meaning-focused feedback. This finding may not be surprising, as meaning-focused feedback is not straightforward and requires more reworking to integrate (Dressler et al., 2019); however, it highlights the necessity of introducing revision strategies for addressing meaning-level feedback in EFL writing classes. As in the earlier investigation (Study I), this study also provides evidence that the integration of Grammarly into advanced writing classes would benefit students and teachers in meaningful ways: Grammarly is highly accurate in flagging specific error types common in L2 writing (e.g., verb tense, verb form, and articles) and can afford students with partial autonomy and allow them to filter suggestions and make informed choices. Furthermore, writing teachers can use Grammarly in a selective manner and shift their feedback focus to higher-level concerns that Grammarly cannot address.

Further investigations into Myanmar and Hungarian students' writing (Study IV) revealed significant differences between the two groups in terms of syntactic features and language-related errors. For example, the Hungarian students produced essays with greater syntactic complexity than their Myanmar peers. Moreover, findings also show that errors are more prevalent in Myanmar students' writing than in the other cohort regardless of error types. These findings elucidate the importance of studying syntactic features and language-related errors in students' writing and analysing them to understand writing proficiency. Moreover, understanding differences in their writing proficiency helped us understand how those with

varying levels of writing proficiency acted upon feedback differently. In particular, the Myanmar students were at a lower writing proficiency level compared to the Hungarian cohort. As a result, Myanmar students were likely to accept teacher and Grammarly feedback without much questioning them. In contrast, Hungarian students made selective decisions about feedback uptake. These findings align with those in previous studies (Chen & Cheng, 2008; Cheng & Liu, 2022). For example, Cheng and Liu (2022) found that although high-proficiency students could understand the cause of linguistic errors, low-proficiency students followed their teacher's direct feedback without understanding it. An earlier study by Chen and Cheng (2008) revealed that students with higher proficiency were likely to question the accuracy of automated feedback and incorporate it selectively.

Considering the affordances of automated tools (e.g., Coh-Metrix and L2SCA) in analysing syntactic complexity, writing teachers may want to employ automated tools for assessing changes in syntactic complexity after a pedagogical intervention that aims to facilitate the development of syntactic complexity, or to monitor students' linguistic development over time. Given that ENA enables researchers to identify and quantify connections among elements in coded data and represent them in dynamic network models, this approach has been used in various education research (Sun et al., 2023; Zhang et al., 2022). In the present research, ENA enabled us to see connections among students' language-related errors in their writing, which coding and counting approaches were unable to do.

8.3 Limitations of the studies and directions for future research

While the studies in this dissertation provided intriguing insights into some major aspects of WCF research, some of their limitations need to be acknowledged and addressed in future studies. First, the study recruited students from intact university English classes, where students usually receive written feedback on their writing. Therefore, I did not include a control group without feedback, which limits the ability to make comparisons between those who received feedback and those who did not. Truscott (2007) claimed that in the absence of a control group, it is impossible to determine whether an observed improvement in accuracy resulted from feedback. Therefore, it would be beneficial to include a control group in future research to maximize the comparability of results across studies. Second, my inquiries were conducted in specific educational settings where participants were enrolled in undergraduate English Studies programmes. For example, it is possible that the findings would have been different if students from specializations other than English had been included. Therefore, future research should

examine how English majors and non-English majors respond to feedback, as I assume that students' writing motivation, perceptions of the usefulness of feedback, and the degree of engagement with feedback might vary depending on specialization.

Third, the participants of studies came from one Myanmar university and one Hungarian university; thus, it is unclear to what extent the results can be generalized to students from other higher education institutions in Myanmar and Hungary. Also, even at these two institutions, possible differences arising from classroom-related factors (e.g., how much time is devoted to developing writing, and how teachers usually offer feedback in their EFL classes) and other socio-cultural differences might impact the findings of this study. A further limitation is that my investigations relied on small datasets from intact classes; hence, my findings should be interpreted with caution. Therefore, future research should consider recruiting larger samples of written texts and compiling a larger corpus over the years to make generalizable suggestions on the impact of feedback on EFL students' writing and their engagement with multiple sources of feedback.

Additionally, when exploring how students benefit from written feedback in their EFL courses, individual and contextual factors may have influenced the efficacy of feedback, which was not taken into account in my research. This limits our understanding of the results regarding why students responded to feedback from different sources. Particularly, student-related factors (e.g., beliefs about the role of feedback in helping to improve their writing, their motivation, and writing anxiety), teacher-related factors (e.g., teachers' competence in writing pedagogy, their motivation to teach writing, and their beliefs about how helpful feedback is), and classroom-related factors (e.g., how much time is devoted for developing writing, and how teachers usually offer feedback in EFL classes) are some mediating factors that might mitigate the effectiveness of feedback on students' texts. In light of this, future research should collect richer data about the classroom to gain a deeper understanding of how students benefit from feedback in their EFL classes.

A further limitation relates to the categorisation of written feedback analysis (see Chapters 4 and 6). The taxonomy of feedback categories suggested by Ferris (2006) and Ferris et al. (1997) was strictly adhered to, however, intra-rater reliability should have been calculated to increase coding reliability, which the present research failed to achieve. Particularly, the coding process was primarily carried out by me when categorising feedback points into language-and content-related issues. In order to ensure the reliability of the coding process, future studies should consider coding 10% of students' texts a second time, some months later, by the same coder.

Another limitation concerns the influence of familiarity with writing topics on students' writing performance. In my studies, the writing topics could not be changed because they were based on the course syllabus. For example, the writing topics included in Study I were retrieved from each unit of the prescribed curriculum. The writing tasks were assigned by the class teacher after completing each unit. As the students were familiar with these writing topics during the course, I did not expect topic knowledge to play a decisive role. However, it is also likely that writing topics were somewhat different in terms of the degree of topic familiarity although they were extracted from the curriculum. All these factors may have an impact on the interpretation of the results. Therefore, future research should employ a counterbalanced design and examine how students' familiarity with writing topics facilitates different aspects of writing (e.g., complexity, accuracy, fluency, and content) to obtain a full picture of how different writing prompts impact learners' writing performance.

Another important aspect that should be addressed in future research relates to the impact of writing genres on syntactic and lexical features of students' writing. For example, Tankó (2020) found that guided summaries and argumentative essays elicited significantly different written language among high-and low-achiever writers. The results suggested that the language of the summaries is significantly similar to that of academic prose in both subcorpora. Also, research on the linguistic development of ESL students in two written genres (i.e., narrative, and argumentative essays) revealed interesting findings: students' language was more complex in argumentative essays compared to narrative writing (Lu, 2011; Yoon & Polio, 2017). These findings led the authors to conclude that "different genres have different communicative or functional requirements that result in different language use" (Yoon & Polio, 2017, p. 282). In contrast, other studies (Robinson, 2011) argued that more complex language occurs in argumentative essays, as these types of essays require more reasoning demands from the part of L2 writers. Based on the existing research, more research is needed to examine how genres of writing impact linguistic features of students' writing, as assessed by CAF measures. Findings from such inquiries would also have important implications for pedagogical practices. For example, if the findings suggest the impact of genre differences on linguistic features of students' writing, it is important for writing teachers to introduce different genres of writing to students at different proficiency levels (even students from lower proficiency levels). These differences are not related to cognitive factors such as cognitive load, but to the fact that different written genres require different language to serve different communicative purposes.

This study employed proficiency essays (see Sections 3.2.1 and 6.3.1) which tend to differ from traditional five-paragraph essays. For proficiency essay writing, students are given four

guiding prompts that are thematically linked to the main idea of the writing task. However, these types of essays do not include an introductory or a concluding paragraph, thus raising the question of whether these prompts meet the standard requirements of the essay genre (e.g., form and function). It is therefore important to take into account these differences regarding the genres of writing when interpreting the results of the studies.

The present research aimed to examine how teacher and automated feedback facilitate students' writing in their EFL courses and how students engaged with these feedback types. Keeping these aims in mind, my studies examined feedback strategies and the scope of teacher feedback and Grammarly feedback (free and paid versions), as well as how students responded to feedback in their revisions and how their writing performance improved by the end of the course. Results suggested that the integration of these feedback sources have great potential in reducing teacher feedback burden and enhancing the efficacy of teacher feedback. Moreover, given that students' engagement with meaning level feedback from the teacher was rather minimal, my results provide some indication that introducing revision strategies and engaging students in learning-oriented activities would be beneficial in helping them clarify feedback information and increasing awareness about how to address teachers' commentary feedback. Further investigations into the impact of feedback on syntactic complexity indicated that feedback does not tend to reduce the complexity of students' writing, although it did not scaffold the development of syntactic complexity either. Moreover, a comparison of Myanmar and Hungarian students' writing based on syntactic features and language-related errors revealed significant differences between them. These findings shed light on the importance of assessing different aspects of syntactic complexity and language-related errors to distinguish writing proficiency of students. Taken together, I believe that my studies contribute to a better understanding of the role of academic writing in English as a foreign language in general and that of written feedback in the processes and outcomes in writing pedagogy.

As a final note, with the development of large-scale natural language models with writing and dialogue capabilities, Artificial Intelligence (AI) has played an important role in facilitating instructional practices. For example, in the case of my studies, given that automated tools have the potential to provide accurate feedback on language errors in students' texts, teachers can exploit the potential of these tools effectively. A significant recent step in AI technology development is the release of an advanced chatbot named ChatGPT (https://chat.openai.com). This tool has the potential to execute a variety of tasks including "question answering, reading comprehension, text implication, semantic similarity matching, text summarisation, code generation, story creation, and more" (Zhang & Li, 2021, p. 831). With these developments, I

wonder how generative AI technology could be used to assist student writers in the process of learning to write. Thus, future research should investigate how these tools can be used to improve students' writing skills. The downside to utilising these tools, however, relates to how AI might threaten writing pedagogy. An independent utilisation of these tools may result in issues of students' cheats. This is a different challenge from plagiarising texts, as they can be detected by plagiarism tools such as Turnitin and Grammarly. In fact, authenticity of writing is at risk: students can present their texts as their own, but their texts are generated by AI. This scenario might be the challenge that writing teachers should address through redesigning teaching and assessing writing in light of AI technology. Moreover, as AI has brought about unforeseen challenges, future research should focus on how writing pedagogy communities can respond in ways which harness the potential of AI in facilitating students' writing, protect authenticity of writing, prevent student cheating, and develop coping strategies.

REFERENCES

Abdi Tabari, M., & Wang, Y. (2022). Assessing linguistic complexity features in L2 writing: Understanding effects of topic familiarity and strategic planning within the realm of task readiness. *Assessing Writing*, *52*, 100605. https://doi.org/10.1016/j.asw.2022.100605

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research*, *7*(3), 347–376. https://doi.org/10.1191/1362168803lr127oa

Allott, A. J. (1985). Language policy and language planning in Burma. In D. Bradley (Ed.), *Language policy, language planning and sociolinguistics in South-East Asia* (pp. 131–154). Department of Linguistics, Research School of Pacific Studies, Australian National University.

Aluthman, E. S. (2016). The effects of using automated essay evaluation on ESL undergraduate students' writing skill. *International Journal of English Linguistics*, *6*(5), 54–67. https://doi.org/10.5539/ijel.v6n5p54

Anker, S. (2010). *Real writing with readings: Paragraphs and essays for college, work, and everyday life* (5th Ed.). St. Martin's.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, *9*(3), 227–257. https://doi.org/10.1016/S1060-3743(00)00027-8

Atkinson, D. (1999). TESOL and culture. *TESOL Quarterly*, *33*(4), 625–654. https://doi.org/10.2307/3587880

Atkinson, D., & Ramanathan, V. (1995). Cultures of writing: An ethnographic comparison of L1 and L2 university writing/language programs. *TESOL Quarterly*, *29*(3), 539. https://doi.org/10.2307/3588074

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*(2), 390–395.

Barrot, J., & Gabinete, M. K. (2021). Complexity, accuracy, and fluency in the argumentative writing of ESL and EFL learners. *International Review of Applied Linguistics in Language Teaching*, *59*(2), 209–232. https://doi.org/10.1515/iral-2017-0012

Barrot, J. S., & Agdeppa, J. Y. (2021). Complexity, accuracy, and fluency as indices of college-level L2 writers' proficiency. *Assessing Writing*, *47*, 100510. https://doi.org/10.1016/j.asw.2020.100510

Benson, S., & DeKeyser, R. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research*, *23*(6), 702–726. https://doi.org/10.1177/1362168818770921

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: a meta-analysis. In *ETS Research Report Series*. https://doi.org/10.1002/j.2333-8504.2011.tb02241.x

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In *Assessment and teaching of 21st century skills* (pp. 17–66). Springer. https://doi.org/10.1007/978-94-007-2324-5_2

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., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, *12*(3), 409–431. https://doi.org/10.1177/1362168808089924

Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal*, *63*(3), 204–211. https://doi.org/10.1093/elt/ccn043

Bitchener, J., & Knoch, U. (2010a). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, *19*, 207–217. https://doi.org/10.1016/j.jslw.2010.10.002

Bitchener, J., & Knoch, U. (2010b). 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

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, *14*(3), 191–205. https://doi.org/10.1016/j.jslw.2005.08.001

Bonilla Lopez, M., Van Steendam, E., & Buyse, K. (2017). Comprehensive corrective feedback on low and high proficiency writers: Examining attitudes and preferences. *International Journal of Applied Linguistics*, *168*(1), 91–128. https://doi.org/10.1075/itl.168.1.04bon

Bonilla Lopez, 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

Brown, M., & Hung, N. (2022). Higher education in Myanmar. In L. P. Symaco & M. Hayden (Eds.), *International handbook on education in South East Asia* (pp. 1–25). Springer Singapore.

Bruton, A. (2010). Another reply to Truscott on error correction: Improved situated designs over statistics. *System*, *38*(3), 491–498. https://doi.org/10.1016/j.system.2010.07.001

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

Burstein, J., Leacock, C., & Swartz, R. (2001). *Automated evaluation of essays and short answers*.

Can, G., & Walker, A. (2011). A model for doctoral students' perceptions and attitudes toward written feedback for academic writing. *Research in Higher Education*, *52*(5), 508–536. https://doi.org/10.1007/s11162-010-9204-1

Cardelle, M., & Corno, L. (1981). Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly*, *15*(3), 251–261. https://doi.org/10.2307/3586751

Cavaleri, M., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning*, *10*(1), A223-A236–A236.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, *12*(3), 267–296. https://doi.org/10.1016/S1060-3743(03)00038-9

Chen, C. F. E., & Cheng, W. Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology*, *12*(2), 94–112.

Cheng, X., & Liu, Y. (2022). Student engagement with teacher written feedback: Insights from low-proficiency and high-proficiency L2 learners. *System*, *109*, 102880. https://doi.org/10.1016/j.system.2022.102880

Cotton, D., Falvey, D., & Kent, S. (2013). *New language leader*. Pearson Education.

Cotton, D., Falvey, D., & Kent, S. (2014). *New language leader* (2nd ed.). Pearson Education.

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment*. Council of Europe. www.coe.int/lang-cefr

Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443. https://doi.org/10.17239/JOWR-2020.11.03.01

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., & 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

Crosthwaite, P., Ningrum, S., & Lee, I. (2022). Research trends in L2 written corrective feedback: A bibliometric analysis of three decades of Scopus-indexed research on L2 WCF. *Journal of Second Language Writing*, *58*, 100934. https://doi.org/10.1016/j.jslw.2022.100934

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, *10*(1), 5–43. https://doi.org/10.1016/j.asw.2005.02.001

Dahlmeier, D., Ng, H., & Wu, S. (2013). Building a large annotated corpus of learner English: The NUS corpus of learner English. *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, 22–31. https://aclanthology.org/W13-1703

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7–24. https://doi.org/10.1016/j.asw.2012.10.002

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (Second, pp. 94–112). Routledge.

Dekeyser, R., & Criado, R. (2013). Automatization, skill acquisition, and practice in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0067

Dikli, S. (2011). The nature of automated essay scoring feedback. *CALICO Journal*, *28*(1), 99–134. https://doi.org/10.11139/cj.28.1.99-134

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, *22*, 1–17. https://doi.org/10.1016/j.asw.2014.03.006

Dörnyei, Z., & Csizér, K. (2002). Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics*, *23*(4), 421–462.

Dressler, R., Chu, M. W., Crossman, K., & Hilman, B. (2019). Quantity and quality of uptake: Examining surface and meaning-level feedback provided by peers and an instructor in a graduate research course. *Assessing Writing*, *39*, 14–24. https://doi.org/10.1016/j.asw.2018.11.001

Dunne, E., & Owen, D. (2013). *The student engagement handbook: Practice in higher education* (E. Dunne & D. Owen (eds.)). Emerald Group Publishing Ltd.

Eckstein, G., & Bell, L. (2021). Dynamic written corrective feedback in first-year composition: Accuracy and lexical and syntactic complexity. *RELC Journal*, 1–18. https://doi.org/10.1177/00336882211061624

Eckstein, G., Sims, M., & Rohm, L. (2020). Dynamic written corrective feedback among graduate students: The effects of feedback timing. *TESL Canada Journal*, *37*(2), 78–102. https://doi.org/10.18806/tesl.v37i2.1339

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

Ellis, R. (1998). Teaching and research: Options in grammar teaching. *TESOL Quarterly*, *32*(1), 39. https://doi.org/10.2307/3587901

Ellis, R. (2009a). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97–107. https://doi.org/10.1093/elt/ccn023

Ellis, R. (2009b). Corrective feedback and teacher development. *L2 Journal*, *1*(1), 2–18. https://doi.org/10.5070/l2.v1i1.9054

Ellis, R. (2009c). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Ellis, R. (2010). A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, *32*, 335–349. https://doi.org/10.1017/S0272263109990544

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 303–332. https://doi.org/10.21832/9781847691767-015

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

Elwood, J. A., & Bode, J. (2014). Student preferences vis-à-vis teacher feedback in university EFL writing classes in Japan. *System*, *42*, 333–343. https://doi.org/10.1016/j.system.2013.12.023

Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System*, *46*(1), 80–95. https://doi.org/10.1016/j.system.2014.07.011

Euroexam International. (2019). *Euroexam detailed specifications*. https://www.euroexam.org/hu

Evans, N. W., Hartshorn, J. K., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System*, *39*(2), 229–239. https://doi.org/10.1016/j.system.2011.04.012

Evans, N. W., Hartshorn, K. J., Cox, T. L., & Martin de Jel, T. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing*, *24*(1), 33–50. https://doi.org/10.1016/j.jslw.2014.02.005

Evans, N. W., Hartshorn, K. J., Mccollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research*, *14*(4), 445–463. https://doi.org/10.1177/1362168810375367

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*(4), 400–414.

Fazilatfar, A. M., Fallah, N., Hamavandi, M., & Rostamian, M. (2014). The effect of unfocused written corrective feedback on syntactic and lexical complexity of L2 writing. *Procedia - Social and Behavioral Sciences*, *98*, 482–488. https://doi.org/10.1016/j.sbspro.2014.03.443

Fazio, L. L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing*, *10*(4), 235–249. https://doi.org/10.1016/S1060-3743(01)00042-X

Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, *31*(2), 315. https://doi.org/10.2307/3588049

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. (2003). Error correction. In *Response to student writing: Implications for second language students* (pp. 42–68). Routledge. https://doi.org/10.1201/9781420037593.ch26

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. (2006). 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. 81–104). Cambridge University Press. https://doi.org/10.1201/9781420037593.ch26

Ferris, D. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing*, *16*, 165–193. https://doi.org/10.1016/j.jslw.2007.07.003

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, *32*(2), 181–201.

Ferris, D. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, *19*, 6–23. https://doi.org/10.1016/j.asw.2013.09.004

Ferris, D., & Hedgcock, J. S. (2005). *Teaching ESL composition: Purpose, process and practice* (2nd ed.). Routledge. https://doi.org/https://doi.org/10.4324/9781410611505

Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, *22*(3), 307–329. https://doi.org/10.1016/j.jslw.2012.09.009

Ferris, D., Pezone, S., Tade, C. R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions & implications. *Journal of Second Language Writing*, *6*(2), 155–182. https://doi.org/10.1016/S1060-3743(97)90032-1

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*(3), 161–184. https://doi.org/10.1016/S1060-3743(01)00039-X

Frear, D., & Chiu, Y. H. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System*, *53*, 24–34. https://doi.org/10.1016/j.system.2015.06.006

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

Ghaboosi, A., & Horváth, J. (2015). The happy corpus: A diachronic study of university students' written proficiency in EFL. In M. Lehmann, R. Lugossy, & J. Horváth (Eds.), *Empirical studies in English applied linguistics* (pp. 141–151). Lingua Franca Csoport.

Goldstein, L. (2006). Feedback and revision in second language writing: Contextual, teacher and student variables. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185–205). Cambridge University Press.

Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, *13*, 63–80. https://doi.org/10.1016/j.jslw.2004.04.006

Grabe, W., & Zhang, C. (2016). Reading-writing relationships in first and second language academic literacy development. *Language Teaching*, *49*(3), 339–355. https://doi.org/10.1017/S0261444816000082

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment*, *8*(6), 1–44.

Hamano-Bunce, D. (2022). The effects of direct written corrective feedback and comparator texts on the complexity and accuracy of revisions and new pieces of writing. *Language Teaching Research*, 1–20. https://doi.org/10.1177/13621688221127643

Han, Y. (2017). Mediating and being mediated: Learner beliefs and learner engagement with written corrective feedback. *System*, *69*, 133–142. https://doi.org/10.1016/j.system.2017.07.003

Han, Y. (2019). Written corrective feedback from an ecological perspective: The interaction between the context and individual learners. *System*, *80*, 288–303. https://doi.org/10.1016/j.system.2018.12.009

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*, 31–44.

Han, Y., & Xu, Y. (2019). Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education*, *26*(2), 181–196. https://doi.org/10.1080/13562517.2019.1648410

Hartshorn, K. J., & Evans, N. W. (2012). The differential effects of comprehensive corrective feedback on L2 writing accuracy. *Journal of Linguistics and Language Teaching*, *3*(2), 25–50. https://doi.org/10.1515/iral.1971.9.4.319

Hartshorn, K. J., & Evans, N. W. (2015). The effects of dynamic written corrective feedback:

A 30-week study. *Journal of Response to Writing*, *1*(2), 6–34.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, *44*(1), 84–109. https://doi.org/10.5054/tq.2010.213781

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hearst, M. A. (2000). The debate on automated essay grading. *IEEE Intelligent Systems and Their Applications*, *15*(5), 22–37. https://doi.org/10.1109/5254.889104

Horváth, J. (2001). *Advanced writing in English as a foreign language: A corpus-based study of processes and products*. Lingua Franca Csoport.

Horváth, J. (2013). What are BA and MA students proud of? The case of the international pride corpus. *Argumentum*, *9*, 178–185.

Horváth, J. (2016). Lexical profiling of written learner language in EFL: The case of a Hungarian university exam corpus. *Research Papers on Teaching English as a Additional Language*, 125–136.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1–20). John Benjamins.

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, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, *31*(2), 217–230. https://doi.org/10.1016/S0346-251X(03)00021-6

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, *10*, 185–212.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

John, P., & Woll, N. (2020). Using grammar checkers in an ESL context: An investigation of automatic corrective feedback. *CALICO Journal*, *37*(2), 169–192. https://doi.org/10.1558/cj.36523

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

Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research*, *24*(4), 519–539. https://doi.org/10.1177/1362168818802469

Karim, K., & Nassaji, H. (2019). The effects of written corrective feedback: A critical synthesis of past and present research. *Instructed Second Language Acquisition*, *3*(1), 28–52. https://doi.org/10.1558/isla.37949

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research*, *42*(2), 173–196. https://doi.org/10.2190/EC.42.2.c

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, *75*(3), 305–313. https://doi.org/10.1111/j.1540-4781.1991.tb05359.x

Khushik, G. A., & Huhta, A. (2019). Investigating syntactic complexity in EFL learners' writing across Common European Framework of Reference levels A1, A2, and B1. *Applied Linguistics*, *41*(4), 506–532. https://doi.org/10.1093/applin/amy064

Kim, H. R., & Bowles, M. (2019). How deeply do second language learners process written corrective feedback? Insights gained from think-alouds. *TESOL Quarterly*, *53*(4), 913–938. https://doi.org/10.1002/tesq.522

Kim, Y., Choi, B., Kang, S., Kim, B., & Yun, H. (2020). Comparing the effects of direct and indirect synchronous written corrective feedback: Learning outcomes and students' perceptions. *Foreign Language Annals*, *53*(1), 176–199. https://doi.org/10.1111/flan.12443

Kirkpatrick, R., & Hlaing, H. L. (2013). The Myanmar university entrance examination. *Language Testing in Asia*, *3*(1), 1–15. https://doi.org/10.1186/2229-0443-3-14

Knoblauch, C. H., & Brannon, L. (1981). Teacher commentary on student writing: The state of the art. *Composition Studies*, *10*(2), 1–4. https://doi.org/10.2307/3197945

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, *44*, 100450. https://doi.org/10.1016/j.asw.2020.100450

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, *17*(1), 48–60. https://doi.org/10.1016/j.jslw.2007.08.003

Kurzer, K. (2018). Dynamic written corrective feedback in developmental multilingual

writing classes. *TESOL Quarterly*, *52*(1), 5–33. https://doi.org/10.1002/tesq.366

Lachner, A., Backfisch, I., & Nückles, M. (2018). Does the accuracy matter? Accurate concept map feedback helps students improve the cohesion of their explanations. *Educational Technology Research and Development*, *66*(5), 1051–1067. https://doi.org/10.1007/s11423-018-9571-4

Lan, G., & Sun, Y. (2019). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes*, *38*, 14–24. https://doi.org/10.1016/j.jeap.2018.12.001

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing*, *13*, 285–312. https://doi.org/10.1016/j.jslw.2004.08.001

Lee, I. (2005). Error correction in the L2 writing classroom: What do students think? *TESL Canada Journal*, *22*(2), 1–16.

Lee, I. (2008a). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, *17*(3), 144–164. https://doi.org/10.1016/j.jslw.2007.12.001

Lee, I. (2008b). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing*, *17*(2), 69–85. https://doi.org/10.1016/j.jslw.2007.10.001

Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal*, *63*(1), 13–22. https://doi.org/10.1093/elt/ccn010

Lee, I. (2020). Utility of focused/comprehensive written corrective feedback research for authentic L2 writing classrooms. *Journal of Second Language Writing*, *49*, 100734. https://doi.org/10.1016/j.jslw.2020.100734

Lee, I. (2022). Developments in classroom-based research on L2 writing. *Studies in Second Language Learning and Teaching*, *12*(4), 551–574. https://doi.org/10.14746/ssllt.2022.12.4.2

Lee, I., Luo, N., & Mak, P. (2021). Issues of error selection for focused written correction feedback in authentic classroom contexts. *RELC Journal*. https://doi.org/10.1177/00336882211028425

Lee, J.-W., & Yoon, K.-O. (2020). Effects of written corrective feedback on the use of the English indefinite article in EFL learners' writing. *English Teaching*, *75*(2), 21–40. https://doi.org/10.15858/engtea.75.2.202006.21

Leeman, J. (2010). Feedback in L2 learning: Responding to errors during practice. In R.

DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 111–138). Cambridge University Press. https://doi.org/10.1017/cbo9780511667275.007

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, Shaofeng. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*(2), 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Li, Shaofeng, & Roshan, S. (2019). The associations between working memory and the effects of four different types of written corrective feedback. *Journal of Second Language Writing*, *45*, 1–15. https://doi.org/10.1016/j.jslw.2019.03.003

Li, Shaofeng, & Vuono, A. (2019). Twenty- five years of research on oral and written corrective feedback in System. *System*, *84*, 93–109. https://doi.org/10.1016/j.system.2019.05.006

Li, Su, & Li, P. (2012). Individual differences in written corrective feedback: A multi-case study. *English Language Teaching*, *5*(11), 38–44. https://doi.org/10.5539/elt.v5n11p38

Li, W., Lu, Z., & Liu, Q. (2020). Syntactic complexity development in college students' essay writing based on AWE. In K.-M. Frederiksen, S. Larsen, L. Bradley, & S. Thouesny (Eds.), *CALL for widening participation: Short papers from EUROCALL 2020* (pp. 190–194). Research-publishing.net. https://doi.org/https://doi.org/10.14705/rpnet2020.48.1187

Liao, H. C. (2015). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, *70*(3), 308–319. https://doi.org/10.1093/elt/ccv058

Liao, H. C. (2016). 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

Lim, S. C., & Renandya, W. A. (2020). Efficacy of written corrective feedback in writing instruction: A meta-analysis. *Tesl-Ej*, *24*(3), 1–26.

Liu, M., Li, Y., Xu, W., & Liu, L. (2017). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, *10*(4), 502–513. https://doi.org/10.1109/TLT.2016.2612659

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, *30*, 66–81. https://doi.org/10.1016/j.jslw.2015.08.011

Liu, Y., & Huang, J. (2020). The quality assurance of a national English writing assessment: Policy implications for quality improvement. *Studies in Educational Evaluation*, *67*, 100941. https://doi.org/10.1016/j.stueduc.2020.100941

Loewen, S. (2012). The role of feedback. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 24–40). Routledge. https://doi.org/10.1017/s0272263112000423

Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York Academy of Sciences*, 259–278. https://doi.org/10.1111/j.1749-6632.1981.tb42014.x

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. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, *29*, 16–27. https://doi.org/10.1016/j.jslw.2015.06.003

Luo, Y., & Liu, Y. (2017). Comparison between peer feedback and automated feedback in college English writing: A case study. *Open Journal of Modern Linguistics*, *7*, 197–215. https://doi.org/10.4236/ojml.2017.74015

Maamuujav, U., Booth, C., & Chung, H. (2021). Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of Second Language Writing*, *53*, 100822. https://doi.org/10.1016/j.jslw.2021.100822

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

Mao, Z., & Lee, I. (2020). Feedback scope in written corrective feedback: Analysis of empirical research in L2 contexts. *Assessing Writing*, *45*, 100469. https://doi.org/10.1016/j.asw.2020.100469

Maung, E. P., Kálmán, C., & Gordon Győri, J. (2022). Myanmar EFL pre-service teachers' motivational disposition in English writing: A pilot study. *Journal of Adult Learning, Knowledge and Innovation*, *5*(2), 60–70. https://doi.org/10.1556/2059.2022.00061

Mawlawi Diab, N. (2015). Effectiveness of written corrective feedback: Does type of error

and type of correction matter? *Assessing Writing*, *24*, 16–34.
https://doi.org/10.1016/j.asw.2015.02.001

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive
English for Academic Purposes programme. *Journal of Second Language Writing*, *29*,
3–15. https://doi.org/10.1016/j.jslw.2015.06.004

McIntosh, K., Connor, U., & Gokpinar-Shelton, E. (2017). What intercultural rhetoric can
bring to EAP/ESP writing studies in an English as a lingua franca world. *Journal of
English for Academic Purposes*, *29*, 12–20. https://doi.org/10.1016/j.jeap.2017.09.001

McMartin-Miller, C. (2014). How much feedback is enough?: Instructor practices and
student attitudes toward error treatment in second language writing. *Assessing Writing*,
19, 24–35. https://doi.org/10.1016/j.asw.2013.11.003

McNamara, D. S., Graesser, A., McCarthy, P., & Cai, Z. (2014). *Automated Evaluation and
Discourse with Coh-Metrix*. Cambridge University Press.
https://doi.org/10.1017/CBO9780511894664

Mirzaii, M., & Aliabadi, R. B. (2013). Direct and indirect written corrective feedback in the
context of genre-based instruction on job application letter writing. *Journal of Writing
Research*, *5*(2), 191–213. https://doi.org/10.17239/jowr-2013.05.02.2

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred
reporting items for systematic reviews and meta-analyses: The PRISMA statement.
Physical Therapy, *89*(9), 873–880.

Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions,
teacher self-assessment, and actual teacher performance. *Journal of Second Language
Writing*, *16*, 82–99. https://doi.org/10.1016/j.jslw.2007.04.002

Myles, J. (2002). Second language writing and research: The writing process and error
analysis in student texts. *Tesl-Ej*, *6*(2), 1–20. https://www.tesl-
ej.org/wordpress/issues/volume6/ej22/ej22a1/

Nicolas-Conesa, F., Manchon, R. M., & Cerezo, L. (2019). The effect of unfocused direct
and indirect written corrective feedback on rewritten texts and new texts: Looking into
feedback for accuracy and feedback for acquisition. *The Modern Language Journal*,
103(4), 848–873. https://doi.org/10.1111/modl.12592

Nikolov, M., & Csapó, B. (2010). The relationship between reading skills in early English as
a foreign language and Hungarian as a first language. *International Journal of
Bilingualism*, *14*(3), 315–329. https://doi.org/10.1177/1367006910367854

Nikolov, M., & Szabo, G. (2015). A study on Hungarian 6th and 8th graders' proficiency in

English and German at dual-language schools. In *Inspirations in foreign language teaching: Studies in applied linguistics, language pedagogy and language teaching* (pp. 184–206). Harlow: Pearson Education.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Nurmukhamedov, U., & Kim, S. H. (2010). "Would you perhaps consider...": Hedged comments in ESL writing. *ELT Journal*, *64*(3), 272–282. https://doi.org/10.1093/elt/ccp063

O'Neill, R., & Russell, A. M. T. (2019a). Grammarly: Help or hindrance? Academic learning advisors' perceptions of an online grammar checker. *Journal of Academic Language & Learning*, *13*(1), A88–A107. https://journal.aall.org.au/index.php/jall/article/view/591

O'Neill, R., & Russell, A. M. T. (2019b). 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

Olsen, S. (1999). Errors and compensatory strategies: A study of grammar and vocabulary in texts written by Norwegian learners of English. *System*, *27*(2), 191–205. https://doi.org/10.1016/S0346-251X(99)00016-0

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492–518.

Orts, S., & Salazar, P. (2016). Corrective Feedback: An exploratory study on age and level of proficiency. *The Grove. Working Papers on English Studies*, *23*, 109–129. https://doi.org/10.17561/grove.v23.a8

Page, E. B. (1994). Computer grading of student prose, using modern concepts and software. *Journal of Experimental Education*, *62*(2), 127–142.

Park, E. S., Song, S., & Shin, Y. K. (2016). To what extent do learners benefit from indirect written corrective feedback? A study targeting learners of different proficiency and heritage language status. *Language Teaching Research*, *20*(6), 678–699. https://doi.org/10.1177/1362168815609617

Pearson, W. S. (2022a). Response to written commentary in preparation for high-stakes second language writing assessment. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1). https://doi.org/10.1186/s40862-022-00145-6

Pearson, W. S. (2022b). The mediating effects of student beliefs on engagement with written feedback in preparation for high-stakes English writing assessment. *Assessing Writing*, *52*, 100611. https://doi.org/10.1016/j.asw.2022.100611

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), 148–158. https://doi.org/10.1038/nrn2317

Petchprasert, A. (2021). Utilizing an automated tool analysis to evaluate EFL students' writing performances. *Asian-Pacific Journal of Second and Foreign Language Education*, *6*(1), 1–16. https://doi.org/10.1186/s40862-020-00107-w

Phuket, P. R. N., & Othman, N. B. (2015). Understanding EFL students' errors in writing. *Journal of Education and Practice*, *6*(32), 99–106. https://doi.org/10.31436/ijes.v5i2.195

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, *10*(1), 52–79. https://doi.org/10.1093/applin/10.1.52

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Buros Institute of Mental Measurements. https://digitalcommons.unl.edu/burosmetacognition/3

Plonsky, L., Egbert, J., & Laflair, G. T. (2014). Bootstrapping in applied linguistics: Assessing its potential using shared data. *Applied Linguistics*, *36*(5), 591–610. https://doi.org/10.1093/applin/amu001

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Polio, C. (2012a). How to research second language writing. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 139–157). https://doi.org/10.1002/9781444347340

Polio, C. (2012b). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, *21*(4), 375–389. https://doi.org/10.1016/j.jslw.2012.09.004

Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, *7*(1), 43–68. https://doi.org/10.1016/S1060-3743(98)90005-4

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, *47*(1), 101–143. https://doi.org/10.1111/0023-8333.31997003

Polio, C., & Park, J.-H. (2012). Language development in second language writing. In R. M.

Manchon & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 287–306). De Gruyter.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

Polio, C., & Yoon, H. J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, *28*(1), 165–188. https://doi.org/10.1111/ijal.12200

Qassemzadeh, A., & Soleimani, H. (2016). The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory and Practice in Language Studies*, *6*(9), 1884. https://doi.org/10.17507/tpls.0609.23

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. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, *52*, 100816. https://doi.org/10.1016/j.jslw.2021.100816

Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System*, *40*(3), 421–430. https://doi.org/10.1016/j.system.2012.07.005

Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). John Benjamins.

Rokoszewska, K. J. (2022). The dynamics of monthly growth rates in the emergence of complexity, accuracy, and fluency in L2 English writing at secondary school–a learner corpus analysis. *System*, *106*, 102775. https://doi.org/10.1016/j.system.2022.102775

Romano, F. (2019). Grammatical accuracy in EAP writing. *Journal of English for Academic Purposes*, *41*, 100773. https://doi.org/10.1016/j.jeap.2019.100773

Rummel, S., & Bitchener, J. (2015). The effectiveness of written corrective feedback and the impact of Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics*, *1*, 66–84. https://doi.org/10.1075/aral.38.1.04rum

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.),

Synthesizing research on language learning and teaching (pp. 133–164). John Benjamins.

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

San, C. K., & Soe, T. (2016). *Grammatical difficulties with English morphemes encountered by undergraduate students in the context of Myanmar*. https://meral.edu.mm/record/816/file_preview/Grammatical Difficulties.pdf

Saricaoglu, A. (2019). The impact of automated feedback on L2 learners' written causal explanations. *ReCALL*, *31*(2), 189–203. https://doi.org/10.1017/S095834401800006X

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, *12*(4), 431–459.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chi, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (Vol. 4, pp. 721–737). National University of Singapore, Centre for Language Studies.

Schunk, D. H. (2012). Behaviorism. In *Learning theories: An educational perspective* (6th ed., pp. 71–116). Pearson Education. https://doi.org/10.1007/BF00751323

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, *5*(1), 1–23.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255–283.

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, *37*(4), 556–569. https://doi.org/10.1016/j.system.2009.09.002

Shen, R., & Chong, S. W. (2022). Learner engagement with written corrective feedback in ESL and EFL contexts: A qualitative research synthesis using a perception-based framework. *Assessment & Evaluation in Higher Education*, 1–15. https://doi.org/10.1080/02602938.2022.2072468

Shi, Y. (2021). Exploring learner engagement with multiple sources of feedback on L2 writing across genres. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.758867

Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, *22*(3), 286–306. https://doi.org/10.1016/j.jslw.2013.03.011

Shintani, N., & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System*, *49*, 110–119. https://doi.org/10.1016/j.system.2015.01.006

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*(1), 103–131. https://doi.org/10.1111/lang.12029

Sia, P. F. D., & Cheung, Y. L. (2017). Written corrective feedback in writing instruction: A qualitative synthesis of recent research. *Issues in Language Studies*, *6*(1), 61–80. https://doi.org/10.33736/ils.478.2017

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *Modern Language Journal*, *99*(2), 263–282. https://doi.org/10.1111/modl.12212

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. *Feedback in Second Language Writing*, 125–142. https://doi.org/10.1017/9781108635547.009

Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, *10*(2), 29. https://doi.org/10.6018/ijes/2010/2/119181

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, *32*, 303–334. https://doi.org/10.1017/S0272263109990532

Sun, Z., Xu, R., Deng, L., Jin, F., Song, Z., & Lin, C. (2023). Beyond coding and counting: Exploring teachers' practical knowledge online through epistemic network analysis. *Computers & Education*, *192*, 104647. https://doi.org/10.1016/j.compedu.2022.104647

Suzuki, W., Nassaji, H., & Sato, K. (2019). The effects of feedback explicitness and type of target structure on accuracy in revision and new pieces of writing. *System*, *81*, 135–145. https://doi.org/10.1016/j.system.2018.12.017

Tankó, G. (2012). *Professional writing: The academic context*. ELTE Eötvös Publishing Ltd.

Tankó, G. (2016). Written summarisation for academic writing skills development: A corpus based contrastive investigation of EFL student writing. In N. Dobrić, E.-M. Graf, & A. Onysko (Eds.), *Corpora in applied linguistics: Current approaches* (pp. 53–78). Cambridge Scholars Publishing. http://www.lingviko.net/feng/05-23-final.pdf

Tankó, G. (2019). *Paraphrasing, summarizing and synthesizing skills for academic writers: Theory and practice* (2nd ed.). ELTE Eötvös Publishing Ltd.

Tankó, G. (2020). Eliciting written academic performance: An investigation of task effect on the written production of EFL students. In C. Kálmán (Ed.), *DEAL 2020: A snapshot of diversity in English applied linguistics* (pp. 1–37). Eötvös Loránd University.

Thi, N. K., & Nikolov, M. (2021a). Feedback treatments, writing tasks, and accuracy measures: A critical review of research on written corrective feedback. *Tesl-Ej*, *25*(3), 1–25.

Thi, N. K., & Nikolov, M. (2021b). How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing. *Asia-Pacific Education Researcher*, *31*(6), 767–779. https://doi.org/10.1007/s40299-021-00625-2

Thi, N. K., & Nikolov, M. (2023). Effects of teacher , automated , and combined feedback on syntactic complexity in EFL students ' writing. *Asian-Pacific Journal of Second and Foreign Language Education*, *8*(6), 1–17. https://doi.org/10.1186/s40862-022-00182-1

Thi, N. K., Nikolov, M., & Simon, K. (2022). Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching*, 1–16. https://doi.org/10.1080/17501229.2022.2122476

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, *91*, 102247. https://doi.org/10.1016/j.system.2020.102247

Tin, T. B. (2014). Learning English in the periphery: A view from Myanmar (Burma). *Language Teaching Research*, *18*(1), 95–117. https://doi.org/10.1177/1362168813505378

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, *46*(2), 327–369.

Truscott, J. (2001). Selecting errors for selective error correction. *Concentric: Studies in English Literature and Linguistics*, *27*(2), 93–108.

Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing*, *13*(4), 337–343. https://doi.org/10.1016/j.jslw.2004.05.002

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. (2010). Further thoughts on Anthony Bruton's critique of the correction debate. *System*, *38*(4), 626–633. https://doi.org/10.1016/j.system.2010.10.003

Truscott, J., & Hsu, A. Y. (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

Uysal, H. H. (2008). Tracing the culture behind writing: Rhetorical patterns and bidirectional transfer in L1 and L2 essays of Turkish writers in relation to educational context. *Journal of Second Language Writing*, *17*, 183–207. https://doi.org/10.1016/j.jslw.2007.11.003

Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions. *International Journal of English Studies*, *10*(2), 1. https://doi.org/10.6018/ijes/2010/2/119171

Van Beuningen, C., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL-International Journal of Applied Linguistics*, *156*, 279–296. https://doi.org/10.2143/itl.156.0.2034439

Van Beuningen, C., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Ventayen, R. J. M., & Orlanda-Ventayen, C. C. (2018). Graduate students' perspective on the usability of Grammarly® in one ASEAN state university. *Asian ESP Journal*, *14*(7), 9–30.

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*(2), 97–111. https://doi.org/10.1016/j.compcom.2011.04.004

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, *44*(3), 299–321. https://doi.org/10.1080/00220272.2012.668938

Vyatkina, N. (2010). The effectiveness of written corrective feedback in teaching beginning German. *Foreign Language Annals*, *43*(4), 671–689. https://doi.org/10.1111/j.1944-9720.2010.01108.x

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*. https://doi.org/10.1016/j.asw.2020.100449

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English language arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, 94–109. https://doi.org/10.1016/j.compedu.2016.05.004

Wilson, J., Roscoe, R., & Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing*, *34*, 16–36. https://doi.org/10.1016/j.asw.2017.08.002

Wind, A. M. (2022). The development of syntactic complexity and fluency in an advanced writing course. In *DEAL 2022: Challenges and opportunities in contemporary English applied linguistics* (pp. 97–118). Eötvös University Press. https://doi.org/10.21862/ELTE.DEAL.2022.4.97

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). "It'd be useful, but I wouldn't use it": Barriers to university students' feedback seeking and recipience. *Studies in Higher Education*, *42*(11), 2026–2041. https://doi.org/10.1080/03075079.2015.1130032

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawaii Press.

Xu, J., & Zhang, S. (2021). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. *Asia-Pacific Education Researcher*. https://doi.org/10.1007/s40299-021-00577-7

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, *15*(3), 179–200. https://doi.org/10.1016/j.jslw.2006.09.004

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

Yi, B., & Ni, C. (2015). Planning and working memory effects on L2 performance in Chinese

EFL learners' argumentative writing. *Indonesian Journal of Applied Linguistics*, *5*(1), 43–51. https://doi.org/10.17509/ijal.v5i1.830

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, *51*(2), 275–301. https://doi.org/10.1002/tesq.296

Yu, S., Jiang, L., & Zhou, N. (2020). Investigating what feedback practices contribute to students' writing motivation and engagement in Chinese EFL context: A large scale study. *Assessing Writing*, *44*(19), 100451. https://doi.org/10.1016/j.asw.2020.100451

Yu, S., Zhang, Y., Zheng, Y., Yuan, K., & Zhang, L. (2019). Understanding student engagement with peer feedback on master's theses: A Macau study. *Assessment and Evaluation in Higher Education*, *44*(1), 50–65. https://doi.org/10.1080/02602938.2018.1467879

Yu, S., Zhou, N., Zheng, Y., Zhang, L., Cao, H., & Li, X. (2019). Evaluating student motivation and engagement in the Chinese EFL writing context. *Studies in Educational Evaluation*, *62*(19), 129–141. https://doi.org/10.1016/j.stueduc.2019.06.002

Zacharias, N. T. (2007). Teacher and student attitudes toward teacher feedback. *RELC Journal*, *38*(1), 38–52. https://doi.org/10.1177/0033688206076157

Zhang, L. J., & Cheng, X. (2021). Examining the effects of comprehensive written corrective feedback on L2 EAP students' linguistic performance: A mixed-methods study. *Journal of English for Academic Purposes*, *54*, 101043. https://doi.org/10.1016/j.jeap.2021.101043

Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT technology review 2021. *Fundamental Research*, *1*(6), 831–833. https://doi.org/10.1016/j.fmre.2021.11.011

Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers and Education*, *183*, 104485. https://doi.org/10.1016/j.compedu.2022.104485

Zhang, T. (2021). The effect of highly focused versus mid-focused written corrective feedback on EFL learners' explicit and implicit knowledge development. *System*, *99*, 102493. https://doi.org/10.1016/j.system.2021.102493

Zhang, X. (2017). Reading–writing integrated tasks, comprehensive corrective feedback, and EFL writing development. *Language Teaching Research*, *21*(2), 217–240. https://doi.org/10.1177/1362168815623291

Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2

writing: Student perceptions and revisions. *Assessing Writing*, *43*, 100439. https://doi.org/10.1016/j.asw.2019.100439

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, 100586. https://doi.org/10.1016/j.asw.2021.100586

Zhang, Z. V. (2017). Student engagement with computer-generated feedback: A case study. *ELT Journal*, *71*(3), 317–328. https://doi.org/10.1093/elt/ccw089

Zhao, X. (2022). Leveraging Artificial Intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC Journal*, 1–5. https://doi.org/10.1177/00336882221094089

Zheng, C., & Park, T.-J. (2013). An analysis of errors in English writing made by Chinese and Korean university students. *Theory and Practice in Language Studies*, *3*(8), 1342–1351. https://doi.org/10.4304/tpls.3.8.1328-1334

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

Zhou, Y. (2022). Revisiting praise as feedback in L2 writing: Where are we going ? *RELC Journal*, 1–9. https://doi.org/10.1177/00336882221123100

Zhu, X., Li, G. Y., Cheong, C. M., & Wen, H. (2021). Effects of L1 single-text and multiple-text comprehension on L2 integrated writing. *Assessing Writing*, *49*, 100546. https://doi.org/10.1016/j.asw.2021.100546

APPENDIX A. General Characteristics of the Selected Studies on Teacher Feedback (*n* = 42)

Study	Participants	Age	L2 proficiency	L1	L2	Language context	Educational level
Ashwell (2000)	50 students in a writing class	Adult learners (age not stated)	Not stated	Japanese	English	EFL (Japan)	Not stated
Fazio (2001)	112 students from multicultural classroom	Teen (10 to 13 years of age)	Elementary	Minority languages and Francophone	French	ESL (Canada)	Grade 5
Ferris and Roberts (2001)	72 immigrant students at the Learning Skills Centre	Adult learners (age not stated)	Not stated	Southeast Asian and Chinese	English	ESL (California)	University
Chandler (2003) Sample 1	31 music major students	Adult learners (age not stated)	High intermediate/ Advanced (Between 540 and 575 in TOEFL)	East Asian languages	English	ESL (American conservatory)	University (undergraduate course)
Chandler (2003) Sample 2	36 students	Adult learners (age not stated)	Intermediate	Hispanic, Asian, and East Asian languages	English	ESL (American conservatory)	University (undergraduate course)
Bitchener et al. (2005)	53 migrant students	Adult learners (in their late twenties and early thirties)	Post intermediate	Chinese	English	ESOL (New Zealand)	Tertiary level
Ferris (2006)	92 students from first-semester freshmen course and their three teachers	Adult learners (age not stated)	Not stated	Various Asian languages	English	ESL (California State University)	University (undergraduate course)
Sheen (2007)	91 international and immigrant learners who enrolled in the American language programme	Adult learners (21-56 years of age)	Intermediate	Varied	English	ESL (US)	Community College
Bitchener (2008)	75 international learners from two private language schools	Adult learners (average age: 22.7 years)	Low intermediate	East Asian	English	ESL (New Zealand)	Not stated
Bitchener and Knoch (2008)	144 international and migrant students	Adult learners (mean age of international	Low intermediate	East Asian	English	ESL (New Zealand)	University

		Ss: 22.7; mean age of migrant Ss: 34.1)					
Ellis et al. (2008)	49 students from general English class	Adult learners (between 18-20 years of age)	Intermediate	Japanese	English	EFL (Japan)	University (first year)
Truscott and Hsu (2008)	47 graduate students from elective basic writing seminar	Adult learners (age not stated)	Those scored 30-42 (out of 60) in General English Proficiency Test	Chinese	English	EFL (Taiwan)	University (postgraduate course)
Van Beuningen et al. (2008)	62 students from pre-vocational education	Teen (around 14 years of age)	Not stated	Multilingual: Arabic, Turkish, the Netherlands	Dutch	EFL (Dutch)	Second year of secondary education
Bitchener and Knoch (2009)	52 students in the English Language Department	Adult learners (average age: 31.7 years)	Low intermediate	East Asian	English	ESL (New Zealand)	University
Sheen et al. (2009)	80 students enrolled in a ESL programme	Adult learners (varied greatly in terms of age)	Intermediate	Varied	English	ESL (Washington DC)	College
Bitchener and Knoch (2010a)	63 learners enrolled in an Introductory Composition Course for international students	Adult learners (18-20 years of age)	Advanced	East and South Asian	English	ESL (US)	University
Bitchener and Knoch (2010b)	52 students from the English Language Department	Adult learners (average age: 31.7 years)	Low intermediate	East Asian	English	ESL (New Zealand)	University
Evans et al. (2010)	27 students from Applied Grammar classes	Adult learners (between 18 to 33 years of age)	Advanced low (TOEFL)	Not stated	English	ESL (US)	Not stated
Hartshorn et al. (2010)	47 students	Adult learners (aged 18 to 45 years old)	Advanced-low to Advanced-mid	Varied	English	ESL (US)	University (Brigham Young University's Language Centre)
Vyatkina (2010)	66 students	Adult learners (age ranged from 18 to 23)	Beginners	English	German	ESL (US)	College

Study	Participants	Age	Proficiency	L1	Target Language	Context	Setting
Evans et al. (2011)	30 matriculated students	Adult learners (mean age of experimental group: 24; control group: 21)	Minimum scores 6.5 on IELTS	Varied	English	ESL (US)	University (undergraduate course)
Hartshorn and Evans (2012)	47 students and their four teachers	Adult learners (mean age of experimental group: 24; control group: 25)	Not stated	Varied	English	ESL (US)	Not stated
Riazantseva (2012)	32 international students	Adult learners (in their twenties)	Low TOEFL iBT writing proficiency scores ranging from 18 to 23	South Korean, Japanese, Taiwanese, and Chinese	English	ESL (US)	University (postgraduate course)
Van Beuningen et al. (2012)	268 learners	Teen (ranged from 14 to 15 years old)	Not stated	Varied	Dutch	EFL (Dutch)	Secondary school
Mirzaii and Aliabadi (2013)	80 learners	Adult learners (average age: 25)	Advanced	Iranian	English	EFL (Iran)	Institute
Shintani and Ellis (2013)	49 students from five intact academic writing classes in an intensive English language programme	Adult learners (22.6 years old on average)	Low-intermediate	Arabian, Chinese, and Korean	English	ESL (US)	University
Shintani et al. (2014)	140 students from seven general English classes, majoring in a variety of subjects	Adult learners (aged 18 to 21)	Pre-intermediate	Japanese	English	EFL (Japan)	University (first and second year)
Frear and Chiu (2015)	42 English major students from three intact classes	Adult learners (19 to 20 years of age)	Not stated	Chinese	English	EFL (Taiwan)	University
Hartshorn and Evans (2015)	27 students from an intensive English programme	Adult learners (mean age of experimental group: 25; control group: 24)	Intermediate (47 to 60 on TOEFL, 4 on IELTS)	Varied	English	ESL (US)	Not stated

Mawlawi Diab (2015)	57 students	Adult learners (aged 18 to 22)	Scored 560 on SAT or higher or passed a pre-requisite freshman level course	Arabian	English	ESL (Lebanon)	University (sophomores)	
Rummel and Bitchener (2015)	42 learners at an English language centre in a programme of English for Academic Purposes	Adult learners (between the age range of 23 and 27)	Advanced	Laos	English	EFL (Vientiane)	English language school	
Shintani and Ellis (2015)	118 students from seven intact general English classes, majoring in a variety of subjects	Adult learners (aged 18 to 21)	Not stated	Japanese	English	EFL (Japan)	University (first and second year)	
Stefanou and Révész (2015)	89 Greek students	Adult learners (16 years of age)	Intermediate (Oxford placement test)	Greek	English	EFL (Cyprus)	High school	
Bonilla Lopez et al. (2017)	91 students majoring in English and English teaching	Adult learners (first group's mean age: 24; second group's mean age: 29)	Mixed (Low and High proficiency)	Spanish	English	EFL (Costa Rica)	Public university	
Benson and DeKeyser (2018)	151 learners from English for Academic Purposes (EAP) classes	Adult learners (age not stated)	Low intermediate to Advanced	27 L1s (mainly Spanish, Vietnamese, and Arabian)	English	ESL (US)	College	
Bonilla Lopez et al. (2018)	139 learners	Adult learners (age not stated)	Lower intermediate	Spanish	English	EFL (Costa Rica)	Public university	
Karim and Nassaji (2018)	53 students from General English class	Adult learners (age range: 18 to 40)	Intermediate	East Asian	English	ESL (Canada)	Not stated	
Kurzer (2018)	325 student participants from three different levels of developmental ESL writing classes	Adult learners (age not stated)	Beginner, Intermediate, and Advanced	Varied	English	ESL (US)	Western US research university	
Nicolas-Conesa et al. (2019)	46 English major students enrolled in a semester-long English for	Adult learners (mean age: 18)	Intermediate	Spanish	English	EFL (Spain)	University (undergraduate course)	

184

	academic purposes composition course						
Lee and Yoon (2020)	58 students from three intact classes (two natural science classes and one literal art)	Adult learners (19 to 22 years of age)	Intermediate high	Korean	English	EFL (Korea)	University
Kim et al., (2020)	53 students from elementary Korean language course	Adult learners (average age: 19.78)	High beginners	Varied	Korean	ESL (US)	University (first year)
Zhang (2021)	59 students from three classes of English course	Adult learners (18 to19 years of age)	Low intermediate (IELTS 4.5)	Thai	English	EFL (Thailand)	International University (first year)

APPENDIX B. A Summary of Studies on Automated Feedback

Study	Participants	Age	L2 proficiency	L1	L2	Language context	Educational level
El Ebyary and Windeatt (2010)	31 instructors and their 549 1st, 2nd, 3rd, and 4th year trainee EFL teachers	Adult learners	Not stated	Egyptian	English	EFL (Egypt)	University
Kellogg et al. (2010)	59 students enrolled in three sections of English	Adult learners	Not stated	Spanish	English	EFL (Spain)	University
Dikli (2011)	12 students from an intensive English centre	Adult learners	Not stated	Varied (Spanish, Arabic, Turkish, Swiss, and Korean)	English	ESL (Florida)	University
Dikli and Bleyle (2014)	14 students from pre-college writing course	Adult learners	Advanced	Varied	English	ESL (US)	Institution
Li et al. (2015)	70 students from two academic writing courses	Adult learners	Not stated	Varied (mainly Mandarin)	English	ESL (US)	University
Liao (2016)	66 college sophomore students in English essay writing classes	Adult learners (19 – 21 years old)	Mixed ability (CEFR levels A2 to B2)	Mandarin Chinese	English	EFL (Taiwan)	University
Wilson and Czik (2016)	145 students from four eighth-grade English Language Arts classes	Young learners	Not stated	Varied	English	ESL (US)	Middle school
Liu et al. (2017)	110 English-major students enrolled at the comprehensive English class	Adult learners	Not stated	Chinese	English	EFL (China)	Southwest University in China (second year)
Luo and Liu (2017)	Students from college English writing class	Adult learners (average age - 20)	Lower-intermediate	Chinese	English	EFL (China)	University (Sophomore students)
Wilson et al. (2017)	480 students from grade six and eight	Young learners	Mixed (struggling and proficient writers)	Varied	English	ESL (US)	Secondary school
Saricaoglu (2019)	31 students enrolled in academic writing class	Adult learners (age ranged from 18 to 25)	Advanced-low	Varied (mainly Chinese)	English	ESL	University (first year)

APPENDIX C. Operational Rating Scales for Writing Tasks at B1 Level Adopted from Euroexam International (2019)

B1	Task achievement	Coherence & Cohesion	Grammatical range & accuracy	Lexical range & accuracy
3	**Task achieved at a high level** **Rubrics**: Followed completely in all 4 guiding points **Content**: Enough and relevant discussion and details are included on all 4 guiding points * One mark will be penalized if some irrelevant discussion and details included	**Information:** Well organized into a coherent text **Cohesive devices:** Overall good use of cohesive devices	**Range:** Good range of grammatical structures **Accuracy:** Grammatical structures used accurately with no or very few basic errors	**Range:** Good range of lexis to complete the task **Accuracy:** Lexis used appropriately with no or little misuse
2	**Task achieved with minor gaps** **Rubrics**: Followed in 2 or 3 guiding points **Content**: Enough and relevant discussion and details are included on 2 or 3 guiding points; Little or not relevant information is discussed on 1 or 2 points * One mark will be penalized if some irrelevant discussion and details included	**Information:** Part of the text is well organized **Cohesive devices:** Mostly good use of cohesive devices with minor gaps	**Range:** Sufficient range of grammatical structures **Accuracy:** Grammatical structures used mostly accurately with some errors that do not significantly impede meaning	**Range:** Sufficient range of lexis to complete the task **Accuracy:** Lexis used mostly appropriately with minor gaps
1	**Task achieved with major gaps** **Rubrics**: Followed in 1 or 2 guiding points **Content**: Enough and relevant details are included on 1 or 2 guiding points; Little or not relevant information is included on 2 or 3 points * One mark will be penalized if some irrelevant discussion and details included	**Information:** Text is hard to follow **Cohesive devices:** Major gaps in use of cohesive devices	**Range:** Limited range of grammatical structures **Accuracy:** Grammatical structures used inaccurately interfering with meaning	**Range:** Limited range of lexis to complete the task **Accuracy:** Lexis often used inappropriately causing misunderstanding
0	**Task unachieved** Task unattempted/ partially attempted Not enough language to make an assessment	---	---	---

What each criterion is supposed to assess are as follows:

1. Task achievement concerns how well a candidate has fulfilled the task, addressing the guided prompt with relevant details while aiming at the general target reader, in other words, if he has done what he was supposed to do.

2. Coherence and cohesion focus on how well-organized a text is, following a coherent structure to maintain the organization of the whole text while making good use of cohesive devices.

3. Grammatical range and accuracy focus on the accuracy of grammatical structures that a candidate uses, demonstrating a variety of grammatical structures available to him.

4. Lexical range and accuracy focus on the accuracy and lexical items that a candidate uses, displaying the appropriate choice and variety of words with an adequate range of lexis to complete the task.

APPENDIX D. Language Background Questionnaire

Dear student,

This questionnaire aims to collect your background information and English language learning experiences. It focuses on your experiences related to English writing and how you receive feedback on your written work. The information from this questionnaire will only be used for the present research project and your personal data will be handled separately. Thank you for filling in this questionnaire and participating in the project.

1. Age
2. Gender
3. Number of years studying English
4. How would you rate your English reading proficiency level?
   (a) Elementary   (b) Pre-intermediate
   (c) Pre-intermediate   (d) Intermediate
   (e) Advanced
5. How would you rate your English writing proficiency level?
   (a) Elementary   (b) Pre-intermediate
   (c) Pre-intermediate   (d) Intermediate
   (e) Advanced
6. How would you rate your English speaking proficiency level?
   (a) Elementary   (b) Pre-intermediate
   (c) Pre-intermediate   (d) Intermediate
   (e) Advanced
7. How would you rate your English listening proficiency level?
   (a) Elementary   (b) Pre-intermediate
   (c) Pre-intermediate   (d) Intermediate
   (e) Advanced
8. How often do you read in English in academic subjects at the university?
   (a) Never   (b) Rarely
   (c) Sometimes   (d) Often
   (e) Usually
9. How often do you engage in reading English for your own pleasure, not for academic purposes?
   (a) Never   (b) Rarely
   (c) Sometimes   (d) Often
   (e) Usually
10.   How often do you write in English in academic subjects at the university?
   (a) Never   (b) Rarely
   (c) Sometimes   (d) Often
   (e) Usually
11.   How often do you engage in writing English as real-life activities, not for academic purposes?
   (a) Never   (b) Rarely

(c) Sometimes   (d) Often

(e) Usually

12.   How often do you receive feedback on your writing from your teachers in your English class at the university?
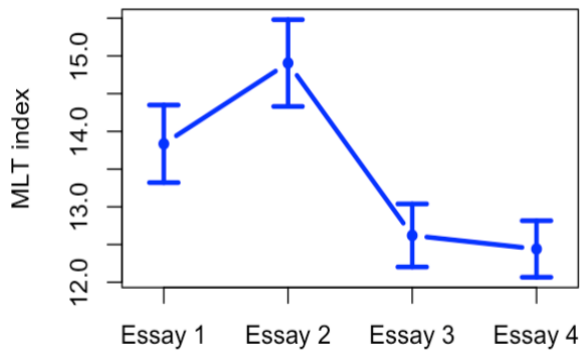
(a) Never   (b) Rarely

(c) Sometimes   (d) Often

(e) Usually

13.   What kind of feedback do you receive on your writing?

14.   What are your strengths in English writing?

15.   What are your weaknesses in English writing?

****

APPENDIX E. Self-assessment Questionnaire

Dear student,

This questionnaire aims to gather your response to teacher feedback and Grammarly feedback on your written work, and your self-assessed language proficiency in L1 (first language) and in English. The information from this questionnaire will only be used for the present research project and your personal data will be handled separately. Thank you for filling in this questionnaire and participating in the project.

1. How useful was the teacher feedback for your revision?
    (a) very useful   (c) not very useful
    (b) useful   (d) useless
2. Please list three examples of useful feedback from the teacher in terms of task achievement, organization, grammar, and vocabulary.
3. How useful was *Grammarly* feedback for your revision?
    (a) very useful   (c) not very useful
    (b) useful   (d) useless
4. Please list three examples of useful feedback from *Grammarly* in terms of task achievement, organization, grammar, and vocabulary.
5. Did you use other sources to help you revise your first draft? If so, which ones?
6. Why do you like the teacher feedback? List 3 reasons?
7. Why do you like *Grammarly* feedback? List 3 reasons.?
8. Why do you like the combination of teacher and *Grammarly* feedback? List 3 reasons.
9. In which area(s) did you improve following the teacher feedback? You may mark more than one.
    (a) Content   (c) Grammar
    (b) Organization   (d) Vocabulary
    (e) Other(s)
10.   In which area(s) did you improve following *Grammarly* feedback? You may mark more than one.
    (a) Content   (c) Grammar
    (b) Organization   (d) Vocabulary
    (e) Other(s)
11. In which area(s) did you improve following the combination of teacher and *Grammarly* feedback? You may mark more than one.
    (a) Content   (c) Grammar
    (b) Organization   (d) Vocabulary
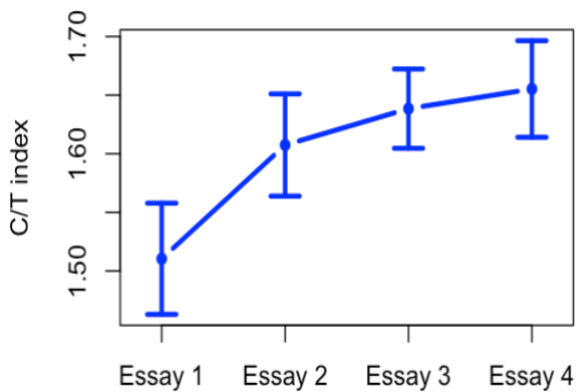   (e) Other(s)

✱✱✱✱✱✱✱✱✱✱✱✱✱✱

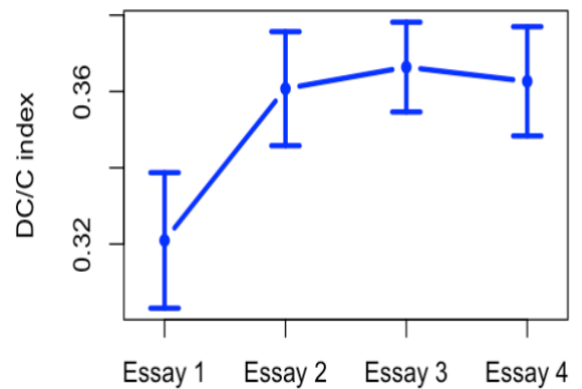# APPENDIX F. Syntactic Complexity across Two Written Genres
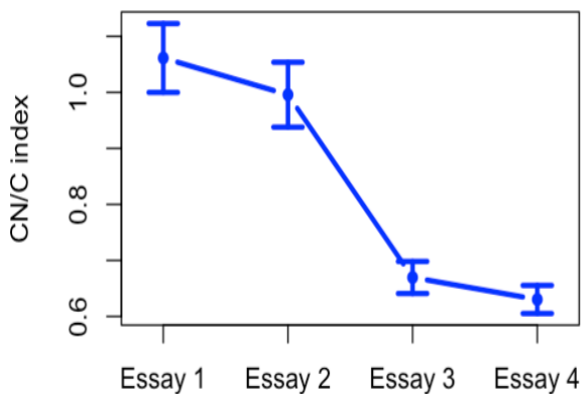

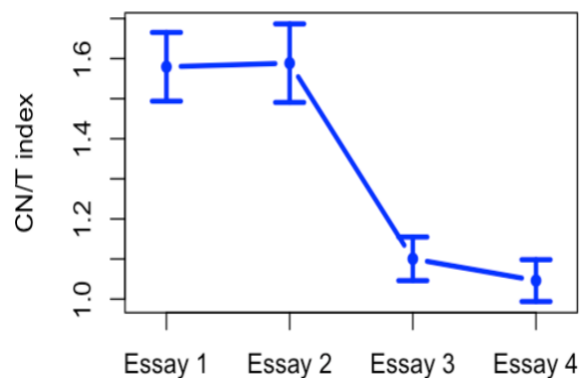(a) Mean length of T-unit
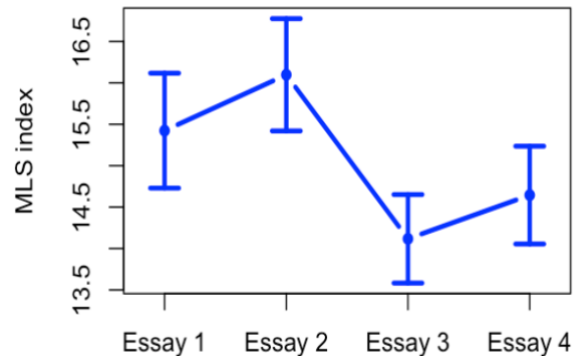
(b) Mean length of sentence

(c) T-unit complexity ratio

(d) Dependent clause ratio

(e) Complex nominals per clause

(f) Complex nominals per T-unit

Note. Students completed argumentative writing in Essays 1 and 2, and narrative writing in Essays 3 and 4.

APPENDIX G. Accuracy of Grammarly's Suggestions

	Grammarly's suggestions	Examples from the data	Accurate feedback	Inaccurate feedback	Indeterminate feedback
Verb tense	would give → gave	If somebody <u>would give</u> me the opportunity to go back in time and relive a memory of mine, I would say yes.	√		
Fragment	No suggestion	* The feeling when you turn the pages in a physical photo album when the memories storm you is priceless.		√	
Miscellaneous	Unclear sentences → clarity (no specific suggestion)	Memories are important milestones in life, because with the help of these we can remember a relative who is no longer with us, in our imagination we can recall the most admirable places we have already been or just a nice experience from the past can put a smile on our face.			√

* *Note.* The accuracy of this sentence is quite debatable. Grammarly flagged it as a fragment, whereas one of the authors (who is a Hungarian) reasoned that this is a typical English as a Lingua Franca (ELF) sentence and that the writer must have been thinking in L1 and coming up with this. On the other hand, a native speaker stated that this sentence is almost unreadable and should be corrected as follows: "One gets a priceless feeling from the memory storm that emerges while turning (the) pages in a physical photo album". Therefore, Grammarly's flagging was coded as *inaccurate* because the sentence does not represent a *fragment* (although it is still regarded as an unreadable sentence).

APPENDIX H. Accuracy of Grammarly's Suggestions and Students' Revisions

	Grammarly's suggestions (n/ %)			Students' revisions (n/ %)			
	Accurate Feedback	Inaccurate feedback	Indeterminate feedback	Correct revision	Incorrect revision	No revision	Deletion
Word choice	0	0	82/100	29/35.4	0	53/64.6	0
Verb tense	3/100	0	0	2/66.7	0	1/33.3	0
Verb form	3/100	0	0	0	0	3/100	0
Word form	-	-	-	-	-	-	-
Articles	10/100	0	0	4/40	0	6/60	0
Singular-Plural	-	-	-	-	-	-	-
Pronouns	13/100	-	-	2/15.4	0	11/84.6	0
Run-on	7/100	0	0	1/14.3	0	6/85.7	0
Fragment	3/75	1/25	0	0	0	4/100	0
Punctuation	148/98	3/2	0	41/27.2	2/1.3	107/70.9	1/0.7
Spelling	3/100	0	0	2/66.7	0	1/33.3	0
Sentence structure	2/100	0	0	0	0	2/100	0
Subject-verb agreement	4/75	1/25	0	1/20	0	4/80	0
Preposition	10/100	0	0	4/40	0	6/60	0
Conjunction	2/100	0	0	1/50	0	1/50	0
Determiner	-	-	-	-	-	-	-
Tone issues	0	0	6/100	1/16.7	1/16.7	4/66.7	0
Modifier	4/100	0	0	0	0	4/100	0
Passive voice	0	22/100	0	3/13.6	2/9.1	17/77.3	0
Mixed dialects of English	0	0	10/100	0	0	10/100	0
Miscellaneous	0	0	144/100	38/26.4	2/1.4	103/71.5	1/0.7
Total	212/44.1	27/5.6	242/50.3	129/26.8	7/1.5	343/71.3	2/0.4

APPENDIX I. Error Categories Used in ENA Analysis

Error category	Error subcategory	Code	Examples from data
Grammar (G)	Omission of object	G.OM	The online criticisms affected [that celebrity] so much that the celebrity was banned from doing any kind of film or music business for one whole year.
	Pronoun	G.PR	We all have encountered a teacher who makes you [us] feel worthless.
	Run-on sentences	G.RU	I try to keep myself on the positive side, often hiding or reporting negative posts as the societies start to handle these with official punishments, I won't spend my time on unnecessary arguments.
	Sentence structure	G.ST	This is very hideous that is youths wasting their precious time on the social media.
	Subject-verb agreement	G.SV	It save [saves] time, energy and also increase [increases] productivity.
	Verb form	G.VF	We all are using it to keep in touch with our friends and also making [make] new friends.
	Verb tense	G.VT	I don't want to imagine what has [would have] happened if my phone wouldn't be there and I couldn't call an ambulance.
Usage (U)	Adverb	U.AD	Last year I had to go to orientation day and because I have never been to Pécs before, [however,] I decided to go sightseeing.
	Article	U.AR	One event where my favorite device was useful was at the [a] field trip.
	Conjunction	U.CO	They have a lot of unsolved conflicts. [and] today one of the most common problems is the lack of respect between the generations.
	Noun ending	U.NE	As I mentioned above, we can share informations [information] and we can earn money through social media.
	Preposition	U.PR	And more, we can also make friends from worldwide [worldwide].
	Word form	U.WF	Therefore, our thoughts can't be same and everything will not be identity [identical].
Lexis (L)	Collocation	L.CO	Teachers teach with videos, live sessions, etc. and students can report [hand in/ submit] their homework and assignments through social media.
	Word choice	L.WC	My life has altered [changed] a lot since I've started using a smartphone, for several reasons.
Mechanics (M)	Capitalization	M.CA	She had to teach us myanmar (Myanmar) and english (English) but most of the time she was not in class.
	Punctuation	M.PU	If social media is properly used in some manner, it can be a boon [,] but if not, it can be a bane.
	Spelling	M.SP	We use Social media for communities, sharing and watching about the imformation [information] around the world.

RELEVANT PUBLICATIONS

Journal articles

1. Thi, N. K., Van, Vo. D., & Nikolov, M. (*in press*). Investigating syntactic complexity and language-related error patterns in EFL students' writing: Corpus-based and epistemic network analyses. *Language Learning in Higher education*.
2. Thi, N.K., Nikolov, M. (2023). Effects of teacher, automated, and combined feedback on syntactic complexity in EFL students' writing. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(6), 1–17. https://doi.org/10.1186/s40862-022-00182-1
3. Thi, N. K., Nikolov, M., & Simon, K. (2022). Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching*, 1–16. https://doi.org/10.1080/17501229.2022.2122476
4. Thi, N. K., & Nikolov, M. (2021a). Feedback treatments, writing tasks, and accuracy measures: A critical review of research on written corrective feedback. *Tesl-Ej*, 25(3), 1–25.
5. Thi, N. K., & Nikolov, M. (2021b). How Teacher and Grammarly Feedback Complement One Another in Myanmar EFL Students' Writing. *Asia-Pacific Education Researcher*, 31(6), 767–779. https://doi.org/10.1007/s40299-021-00625-2

Conference papers

1. Thi, N. K. (2022). Student engagement with teacher and automated feedback in an online EFL writing course. Abstract book: *In SIG 1&4 2022, Exploring research synergies to learn from each other*. Cadiz, Spain: Universidad de Cadiz, Spain, pp. 31–31.
2. Thi, N. K., & Van, Vo. D. (2022). Measuring syntactic complexity in EFL students' writing: A corpus-based analysis. Abstract book: *In SIG 1&4 2022, Exploring research synergies to learn from each other*.Cadiz, Spain: Universidad de Cadiz, Spain, pp. 31–31.
3. Thi, N. K. (2022). Written corrective feedback and its impact on syntactic complexity in EFL students' writing. Abstract book: *Pedagogical Assessment Conference*: XVIII Conference on Educational Assessment. Szeged, Hungary: Doctoral School of Education, University of Szeged, pp. 51–51.
4. Thi, N. K. (2021). An Exploratory Study on How Myanmar EFL Students Benefit from Teacher Feedback on L2 Writing. In: Molnár, Gyöngyvér; Tóth, Edit (ed.): *The answers of education to the challenges of the future*: XXI. ONK. National Educational Science Conference. November 18-20, 2021. Szeged, Hungary: Institute of Education, University of Szeged, pp. 415-415.
5. Thi, N. K. (2021). Assessing Practical Issues in the Efficacy of Written Corrective Feedback: A Review of Empirical Studies. Abtract book: *The 14th Training and Practice International Conference on Educational Science*. Kaposvár, Hungary: Kaposvár University Faculty of Pedagogy, pp. 134-134.
6. Thi, N. K. (2021). An Investigation into Measures of Linguistic Accuracy in Written Corrective Feedback Research. Abstract book: In EARLI 2021: *Education and Citizenship: Learning and Instruction and the Shaping of Futures*. Online, pp. 132-132.

7. Thi, N. K. (2021). An Exploratory Study on How Myanmar EFL Students Benefit from Grammarly Feedback. Abstract book: In EARLI 2021: *Education and Citizenship: Learning and Instruction and the Shaping of Futures*. Online, pp. 14-14.

8. Thi, N. K. (2020). Developing Life Skills Education in Primary Schools in Myanmar. In Hercz, Mária; Lindner, Johannes (ed.) *Developing Social Entrepreneurship in Childhood*: International Conference and Exhibition of the UKids Project. December 5-15, 2020. Budapest, Hungary: Faculty of Primary and Pre-School Education, Eötvös Loránd University, pp. 26-27.

9. Thi, N. K. (2020). Exploring the role of automated feedback in writing classrooms: A review of empirical studies. Abtract book: *The 13th Training and Practice International Conference on Educational Science*. Kaposvár, Hungary: Kaposvár University Faculty of Pedagogy, pp. 90-90.

10. Thi, N. K. (2020). Exploring How Motivation and Corrective Feedback Interact: A Review of Empirical Studies. Abstract book: VI. Ipszilon Konferencia. Budapest, Hungary: ELTE PPK, pp. 46-46.

DECLARATION

I hereby declare that this dissertation, which is submitted for a PhD degree, represents an original report of my research, except where references are clearly stated and acknowledged. This dissertation has not been submitted, in whole or in part, in any previous application for any other degree. Any views expressed in this dissertation are solely those of the author.